On the algebra of Dirac bispinor densities: Factorization and inversion theorems

J. P. Crawford
The Pennsylvania State University, Fayette Campus, P.O. Box 519, Uniontown, Pennsylvania 15401

(Received 12 December 1984; accepted for publication 4 April 1985)

The algebraic system formed by Dirac bispinor densities $\rho_i \equiv \bar{\psi}\Gamma_i\psi$ is discussed. The inverse problem—given a set of 16 real functions ρ_i, which satisfy the bispinor algebra, find the spinor ψ to which they correspond—is solved. An expedient solution to this problem is obtained by introducing a general representation of Dirac spinors. It is shown that this form factorizes into the product of two noncommuting projection operators acting on an arbitrary constant spinor.

I. INTRODUCTION: THE ALGEBRA OF DIRAC BISPINOR DENSITIES

In the context of quantum mechanics all physical observables are represented by real quadratic functionals of the quantum wave function. For a Dirac particle, represented by a four-component spinor wave function ψ, there exist 16 real quadratic forms $\rho_i \equiv \bar{\psi}\Gamma_i\psi$, which do not involve derivatives[1]:

$$\sigma = \bar{\psi}\psi, \quad \pi = \bar{\psi}i\gamma_5\psi, \quad \Sigma_{\mu\nu} = \bar{\psi}\sigma_{\mu\nu}\psi, \qquad (1.1)$$

$$j_\mu = \bar{\psi}\gamma_\mu\psi, \quad k_\mu = \bar{\psi}\gamma_5\gamma_\mu\psi.$$

These 16 bispinor densities are not independent since the spinor wave function, being composed of four independent complex functions, contains only eight independent functions. Furthermore, as the overall phase of the spinor has no effect on the bispinor densities ρ_i we conclude that these 16 functions must satisfy a total of nine algebraic equations. This "Dirac bispinor algebra" was first discussed by Pauli[2] and somewhat later was examined in detail by Fierz.[3] These nine equations are most easily written as[4]

$$j_\mu j^\mu = \sigma^2 + \pi^2, \qquad (1.2a)$$

$$k_\mu k^\mu = -j_\mu j^\mu, \qquad (1.2b)$$

$$j_\mu k^\mu = 0, \qquad (1.2c)$$

$$\Sigma_{\mu\nu} = (\sigma^2 + \pi^2)^{-1}[\sigma\epsilon_{\mu\nu\rho\tau}j^\rho k^\tau - \pi(j_\mu k_\nu - j_\nu k_\mu)]. \qquad (1.2d)$$

Note, however, that Eq. (1.2d) is not valid in the case where both σ and π are zero. An additional overcomplete set of equations is easily derived from Eqs. (1.2) and, as presented, are valid even when both σ and π vanish[5]:

$$\begin{aligned}
&\Sigma_{\mu\nu}j^\nu = \pi k_\mu, \quad \Sigma_{\mu\nu}k^\nu = \pi j_\mu, \\
&\tilde{\Sigma}_{\mu\nu}j^\nu = \sigma k_\mu, \quad \tilde{\Sigma}_{\mu\nu}k^\nu = \sigma j_\mu \\
&\sigma\Sigma_{\mu\nu} - \pi\tilde{\Sigma}_{\mu\nu} = \epsilon_{\mu\nu\rho\tau}j^\rho k^\tau, \\
&\sigma\tilde{\Sigma}_{\mu\nu} + \pi\Sigma_{\mu\nu} = -(j_\mu k_\nu - k_\mu j_\nu), \\
&\tilde{\Sigma}_{\mu\nu}\Sigma^{\mu\nu} = -4\sigma\pi, \quad \Sigma_{\mu\nu}\Sigma^{\mu\nu} = 2(\sigma^2 - \pi^2).
\end{aligned} \qquad (1.3)$$

These equations are given here for the convenience of the reader, as in what follows we will assume that σ and π are not both zero.

An interesting question now arises: Given a set of real functions $\{\rho_i\}$ which satisfy the bispinor algebra, how can we reconstruct the spinor to which they correspond? This problem has most recently been discussed by Takahashi.[6] We present here a new and concise method of solution.

In order to solve this problem in a completely general and very efficient fashion it is convenient to first consider the general structure of an arbitrary spinor.

II. GENERAL SPINOR EXPANSION AND FACTORIZATION THEOREM

We begin this section with an assertion concerning the general form that an arbitrary spinor wave function may take:

$$\begin{aligned}
\psi &= e^{-i\varphi}[\Sigma - \Pi i\gamma_5 + J_\mu\gamma^\mu - K_\mu\gamma_5\gamma^\mu + \tfrac{1}{2}S_{\mu\nu}\sigma^{\mu\nu}]\eta \\
&\equiv e^{-i\varphi}R^i\Gamma_i\eta,
\end{aligned} \qquad (2.1)$$

where the set $\{\varphi, R_i\}$ contains seventeen real functions, and η is an arbitrary constant spinor.[7] Clearly, even for a specific choice of η we will still have great freedom in choosing the set $\{\varphi, R_i\}$, since ψ contains only eight independent functions. To restrict this choice further we make a second assertion: The set of functions $\{R_i\}$ can always be chosen such that they satisfy the bispinor algebra [Eqs. (1.2) and (1.3)]. That this statement is true will be clear when we prove the inversion theorem in the next section and explicitly construct the set of functions $\{R_i\}$. For now we simply note that since the set $\{R_i\}$ satisfies nine independent algebraic equations, there are a total of eight independent functions in the set $\{\varphi, R_i\}$. We now have the first important result of this paper, which we state as a theorem.

Factorization Theorem: When the set of functions $\{R_i\}$ form a bispinor algebra [i.e., satisfy Eqs. (1.2) and (1.3)], the general form for a spinor [Eq. (2.1)] may be factorized in the following manner:

$$\begin{aligned}
\psi = e^{-i\varphi}&[\Sigma - \Pi i\gamma_5 + J_\mu\gamma^\mu] \\
&\times[1 - (\Sigma^2 + \Pi^2)^{-1}(\Sigma + i\gamma_5\Pi)K_\nu\gamma_5\gamma^\nu]\eta.
\end{aligned} \qquad (2.2)$$

Proof: The proof is quite straightforward and, starting from Eq. (2.2), follows:

$$\begin{aligned}
\psi = e^{-i\varphi}&[\Sigma - i\gamma_5\Pi + J_\mu\gamma^\mu - K_\mu\gamma_5\gamma^\mu \\
&- (\Sigma^2 + \Pi^2)^{-1}(\Sigma - i\gamma_5\Pi)J_\mu K_\nu\gamma^\mu\gamma_5\gamma^\nu]\eta,
\end{aligned}$$

where we have used $\gamma_5^2 = 1$ and $\{\gamma_5, \gamma_\mu\} = 0$. Now we make use of the identities

$$\gamma_\mu\gamma_5\gamma_\nu = -\gamma_5(g_{\mu\nu} - i\sigma_{\mu\nu}),$$

$$\gamma_5\sigma_{\mu\nu} = (i/2)\epsilon_{\mu\nu\rho\tau}\sigma^{\rho\tau}$$

 0022-2488/85/071439-03$02.50

to obtain

$$\begin{aligned}\psi &= e^{-i\varphi}\{\Sigma - i\gamma_5\Pi + J_\mu\gamma^\mu - K_\mu\gamma_5\gamma^\mu \\ &\quad + \tfrac{1}{2}(\Sigma^2+\Pi^2)^{-1}[\Sigma\epsilon_{\mu\nu\rho\tau}J^\rho K^\tau \\ &\quad - \Pi(J_\mu K_\nu - J_\nu K_\mu)]\sigma^{\mu\nu}\}\eta \\ &= e^{-i\varphi}[\Sigma - i\gamma_5\Pi + J_\mu\gamma^\mu \\ &\quad - K_\mu\gamma_5\gamma^\mu + \tfrac{1}{2}S_{\mu\nu}\sigma^{\mu\nu}]\eta. \qquad \text{Q.E.D.}\end{aligned}$$

In the last step we have made use of Eq. (1.2d), since the set $\{R_i\}$ satisfies the bispinor algebra by assumption.[8]

To gain insight into the meaning of this factorization consider the special case where

$$\Sigma = m,\quad \Pi = 0,\quad J_\mu = P_\mu,\quad K_\mu = -mS_\mu,$$
$$P^2 = m^2,\quad S^2 = -1,\quad \varphi = P_\mu x^\mu$$

This is the solution to the free massive Dirac equation where the first operator, $(\not{P} + m)$, is the "positive energy projection operator" and the second operator, $(1+\gamma_5\not{S})$, is the "spin projection operator." This observation leads us to the following definitions:

$$\Lambda_\pm \equiv [\Sigma \mp i\gamma_5\Pi \pm J_\mu\gamma^\mu], \tag{2.3a}$$

$$S_\pm \equiv [1 \mp (\Sigma^2+\Pi^2)^{-1}(\Sigma + i\gamma_5\Pi)K_\mu\gamma_5\gamma^\mu]. \tag{2.3b}$$

These operators have the following properties, which are easily verified:

$$\begin{aligned}&\Lambda_\pm^2 = 2\Sigma\Lambda_\pm,\quad \Lambda_+\Lambda_- = 0 = \Lambda_-\Lambda_+,\\ &\Lambda_+ + \Lambda_- = 2\Sigma,\quad S_\pm^2 = 2S_\pm,\\ &S_+S_- = 0 = S_-S_+,\quad S_+ + S_- = 2,\end{aligned} \tag{2.4}$$

and, therefore, satisfy the requirements for projection operators.[9] With these definitions, we can write

$$\psi = e^{-i\varphi}\Lambda_+S_+\eta. \tag{2.5}$$

These definitions prove quite useful in what follows.

III. INVERSION THEOREM

We are now in a position to efficiently solve the problem posed in the first section of this paper. Given a set of real functions $\{\rho_i\}$, which satisfy the bispinor algebra, find the spinor ψ to which they correspond. To begin, we substitute the general factorized expansion of the spinor ψ (to be determined) into the definition of the functions ρ_i:

$$\rho_i \equiv \bar\psi\Gamma_i\psi = \bar\eta\bar S_+\bar\Lambda_+\Gamma_i\Lambda_+S_+\eta, \tag{3.1}$$

where the Dirac adjoint of an operator A is defined in the usual way as $\bar A = \gamma_0A^\dagger\gamma_0$. From the definitions given in Eqs. (2.3) it is quite easy to show that $\bar\Lambda_+ = \Lambda_+$ and $\bar S_+\Lambda_+ = \Lambda_+S_+$ (see Ref. 10) so we now have

$$\rho_i = \bar\eta\bar S_+\Lambda_+\Gamma_i\Lambda_+S_+\eta = \bar\eta\Lambda_+S_+\Gamma_i\bar S_+\Lambda_+\eta. \tag{3.2}$$

In addition, we easily find the useful identities

$$\gamma_5\Lambda_\pm = \Lambda_\mp\gamma_5 \mp 2i\Pi, \tag{3.3a}$$

$$\gamma_\mu\Lambda_\pm = \Lambda_\mp\gamma_\mu \pm 2J_\mu, \tag{3.3b}$$

$$\gamma_5\gamma_\mu\bar S_\pm = S_\mp\gamma_5\gamma_\mu \pm 2K_\mu(\Sigma^2+\Pi^2)^{-1}(\Sigma + i\gamma_5\Pi). \tag{3.3c}$$

It is now straightforward to calculate the ρ_i using Eqs. (2.4), (3.2), and (3.3), and we find

$$\begin{aligned}&\sigma = 4N\Sigma,\quad \pi = 4N\Pi,\quad \Sigma_{\mu\nu} = 4NS_{\mu\nu},\\ &j_\mu = 4NJ_\mu,\quad k_\mu = 4NK_\mu,\end{aligned} \tag{3.4}$$

where we have defined

$$N = \bar\eta\Lambda_+S_+\eta. \tag{3.5}$$

The desired inversion now follows trivially from Eqs. (3.4),

$$R_i = (1/4N)\rho_i, \tag{3.6}$$

the only complication being the calculation of N in terms of the given functions $\{\rho_i\}$. Substitution of Eq. (3.6) into Eq. (3.5) yields[11]

$$\begin{aligned}N^2 &= \tfrac{1}{4}\bar\eta[\sigma - i\gamma_5\pi + j^\mu\gamma_\mu]\\ &\quad\times[1 - (\sigma^2+\pi^2)^{-1}(\sigma + i\gamma_5\pi)k_\mu\gamma_5\gamma^\mu]\eta.\end{aligned} \tag{3.7}$$

So we have found a simple and general inversion of the bispinor algebra $\{\rho_i\}$ (see Ref. 12). We summarize this as a theorem whose proof is given by the above construction.

Inversion Theorem: The spinor ψ which generates the given bispinor algebra $\{\rho_i\}$ is determine by use of Eqs. (2.2), (3.6), and (3.7).

This inversion is not unique, as we may arbitrarily choose both the phase function φ and the constant spinor η. However, the choice of η is essentially irrelevant since the projection operators generate the required direction in spinor space[13] and a change in normalization of η is compensated by a corresponding change in N. Hence, regardless of the choice of η, the spinor wave function is uniquely determined up to an arbitrary phase. As expected, the phase of the spinor wave function cannot be determined from the observable quantities $\{\rho_i\}$.

The inversion theorem can also be used to cast any given Dirac spinor ψ into the general form given in Eq. (2.1). To proceed, first calculate the 16 bispinor densities ρ_i using Eq. (1.1). Next choose any constant spinor η (this choice is completely arbitrary). The overall phase φ and the normalization function N can now be determined using

$$\bar\eta\psi = e^{-i\varphi}N. \tag{3.8}$$

Note that the function N can also be determined by use of Eq. (3.7). Finally, the spinor ψ is then given by

$$\psi = e^{-i\varphi}(4N)^{-1}\rho^i\Gamma_i\eta = e^{-i\varphi}\Lambda_+S_+\eta. \tag{3.9}$$

This construction constitutes the proof that any spinor ψ can be written in the form of Eq. (2.1).

IV. CONCLUDING REMARKS

We have shown that an arbitrary spinor wave function can be written as the product of two projection operators acting on an arbitrary constant spinor and with this result have found a concise and general inversion of the bispinor algebra. We have then used this result to construct the specific factorized form for any given spinor wave function.

This work arose out of a study of nonlinear Dirac equations, and the application of the general spinor expansion (Sec. II) to this problem is the subject of ongoing research.

[1]For the 16 Dirac matrices Γ_i we use the conventions of J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw–Hill, New York, 1964).

[2]W. Pauli, Ann. Inst. H. Poincaré **6**, 109 (1936).

[3]M. Fierz, Z. Phys. **104**, 553 (1937).

[4]F. A. Kaempffer, Phys. Rev. D **23**, 918 (1981).

[5]The dual tensor is defined in the usual manner: $\tilde{\Sigma}_{\mu\nu} \equiv \frac{1}{2}\epsilon_{\mu\nu\rho\tau}\Sigma^{\rho\tau}$, where $\epsilon_{\mu\nu\rho\tau}$ is the completely antisymmetric tensor of rank 4 and $\epsilon^{0123} = -\epsilon_{0123} = 1$.

[6]Y. Takahashi, Phys. Rev. D **26**, 2169 (1982); Prog. Theor. Phys. **69**, 369 (1983). See also V. A. Zhelnorovich, Sov. Phys. Dokl. **24**, 899 (1979); A. A. Campolataro, Int. J. Theor. Phys. **19**, 99, 127 (1980).

[7]For the purposes of this paper, η need not be taken as a constant spinor. This restriction becomes useful when this expansion is substituted into the Dirac equation. In this case, the Dirac equation for ψ is replaced by differential equations for the functions $\{\varphi, R_i\}$.

[8]Note that the factorization theorem fails when Σ and Π are both zero.

[9]Strictly speaking, the properly normalized projection operators are $\frac{1}{2}S_\pm$ and $(1/2\Sigma)\Lambda_\pm$.

[10]Note that as defined $\bar{S}_+ \neq S_+$ (and also $[S_+, \Lambda_+] \neq 0$) for the case $\Pi \neq 0$, and therefore is itself not an observable. It may be possible to factor Eq. (2.1) in a manner different from Eq. (2.2) yielding projection operators $S_\pm$ and $\Lambda_\pm$, which are both Dirac self-adjoint and which mutually commute, but as yet I have been unable to do so.

[11]One can verify that the right-hand side of Eq. (3.7) is positive definite so long as $j_0 > 0$ and the set $\{\rho_i\}$ forms a bispinor algebra. Note that if $j_0 < 0$ there is *no* solution for ψ since $j_0 = \psi^\dagger\psi > 0$. However, in this case a ψ can be found which generates the bispinor algebra $\{-\rho_i\}$.

[12]Compare, for example, this inversion with the forms appearing in the papers of Ref. 6.

[13]We assume, of course, that $\Lambda_+ S_+ \eta \neq 0$.

A generalized Molien function for field theoretical Hamiltonians

Jeffrey W. Felix and Dorian M. Hatch
Brigham Young University, 296 ESC, Provo, Utah 84602

(Received 8 May 1984; accepted for publication 15 March 1985)

A generating function, or Molien function, the coefficients of which give the number of independent polynomial invariants in G, has been useful in the Landau and renormalization group theories of phase transitions. Here a generalized Molien function for a field theoretical Hamiltonian (with short-range interactions) of the most general form invariant in a group G is derived. This form is useful for more general renormalization group calculations. Its Taylor series is calculated to low order for the $F\Gamma_2^-$ representation of the space group $R\,\bar{3}c$ and also for the $l = 1$ (faithful) representation of SO(3).

I. INTRODUCTION

The idea of a generating function, sometimes known as a Molien function,[1] the coefficients in the Taylor's series expansion of which give meaningful information about a particular group, has proven[2–4] to be very useful in constructing free energies for use in the Landau theory of structural and magnetic phase transitions in solids. In Ref. 5, it was shown that an effective Hamiltonian or field theoretical Hamiltonian could be constructed for structural phase transitions of the form

$$H(c) = \sum_{m=0}^{\infty} \frac{1}{m!} \int \cdots \int d\mathbf{k}_1 \cdots d\mathbf{k}_m (2\pi)^{-(m-1)d} \times \delta(\mathbf{k}_1 + \cdots + \mathbf{k}_m) \sum_{L_m} H^m_{L_m}(\mathbf{k}_1,...,\mathbf{k}_m) \times c_{l_1}(\mathbf{k}_1) \cdots c_{l_m}(\mathbf{k}_m), \tag{1}$$

leading to a free energy of

$$F = -\frac{1}{\beta} \ln \int Dc\, e^{-H(c)}.$$

Here $\int Dc$ indicates a functional integration over the collection of $c_i(\mathbf{k})$ and $\mathbf{k}$ ranges continuously over a sphere with $k<\Lambda$, the cutoff parameter. L_m is the compound index $l_1 l_2 \cdots l_m$.

Furthermore, the form of $H(c)$ must be invariant when

$$c_i(\mathbf{k}) \to c_j(S^{-1}\mathbf{k}) D_{ji}(g), \tag{2}$$

or equivalently, when $\mathbf{k} \to S\mathbf{k}$, $H^m \to H^m D^m(g)$, where $g = (S\,|\,\mathbf{t}+\mathbf{t})$ is a space group element in G, the space group of the higher symmetry phase. It would be desirable if a generating function could be found for the more general field theoretical or Landau–Ginzburg–Wilson Hamiltonian of Eq. (1), given Eq. (2). In Ref. 6 the authors pointed out that a term not previously considered, but present in the most general invariant free energy form, contributes significantly to renormalization group behavior. A generalized Molien function for the H of Eq. (1) would be of aid in other such general considerations. Also, the $D(g)$ of Eq. (2) could be replaced by a general representation of any compact group, particularly any unitary compact group, and the following analysis would hold, if the group mean is suitably defined. (This does not take into account the antiunitary θ or complex conjugation, but a generalization is easily accomplished to include it.)

II. DERIVATION AND CALCULATION

Define a function $F(s,t)$, analytic in s,t, such that if

$$F(s,t) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} c_{nm} s^n t^m,$$

then c_{nm} will be the number of invariants of order n in the components of $\mathbf{k}$ and order m in c in $H(c)$. To find $F(s,t)$ we pursue a method which is motivated by procedures contained in Refs. 1, 2, and 7. The invariance group of a term in Eq. (1) includes the symmetry of Eq. (2) and in addition each term must be invariant under any exchange of indices, either on k or c. Such an invariance group is known formally as a wreath product; however, here we need only recognize the existence of both kinds of symmetries. The condition that a term of given order be invariant will be taken to mean that it transforms identically under an arbitrary product of a transformation g in G, and one in S_m, call it π, where S_m is the symmetric group in m objects. Note that the essential symmetry is in $\mathbf{k}$ space. One can be misled in attempting to find invariant forms in real space.[5,6]

To find the number of invariants of given order n and m we construct a general basis of the right order and find the subduction frequency of the identity representation on the representation induced by this basis.

For $m = 0$, there is only one basis functional, i.e., a constant, independent of c and k, hence

$$c_{n0} = \delta_{n0}. \tag{3}$$

For $m = 1$, the basis functionals in Eq. (1) are[5]

$$\Psi_{l_1} = c_{l_1}(0).$$

The permutational invariance subgroup is just S_1 so that

$$c_{n1} = \underset{g}{M} \sum_i D_{ii}(g)\delta_{n0} = \underset{g}{M}\, \chi(g)\delta_{n0}, \tag{4}$$

where

$$\underset{g}{M} = \frac{1}{|G|} \sum_{g\in G}$$

is the group mean, or its suitable generalization to an infinite group.

For $m \geq 2$, define a vector $\mathbf{a}$ in $N^{(m-1)d}$, the $(m-1)d$th Cartesian product of N, the set of non-negative integers, with components a_{ij}, $i = 1,2,...,m-1$, and $j = 1,2,...,d$, such that $\Sigma_{ij} a_{ij} = n$. Here d is the spatial dimension. Then a basis of functionals of order n,m in H, with $m \geq 2$, will have inte-

 0022-2488/85/071442-04$02.50

grands of the form

$$\Psi(\mathbf{a},L_m)=\prod_{i=1}^{m-1}\prod_{j=1}^{d}(k_{ij})^{a_{ij}}c_{l_1}(\mathbf{k}_1)\times c_{l_2}(\mathbf{k}_2)\cdots c_{l_m}(\mathbf{k}_m),$$

where $\mathbf{k}_m=-\mathbf{k}_1-\cdots-\mathbf{k}_{m-1}$ and k_{ij} is the jth component of $\mathbf{k}_i$. This particular Ψ has the advantage of already being symmetric in the k_{ij} for a fixed i. No sums are implied here or in the aftermath.

Suppose the basis of the $\mathbf{k}_i$ has been chosen so that $gk_{ij}=\rho_j k_{ij}$, where ρ_j is an eigenvalue of $V(g)$, a matrix in the vector representation of G. Then

$$g\Psi(\mathbf{a},L_m)=\prod_{ij}(\rho_j)^{a_{ij}}(k_{ij})^{a_{ij}}\sum_{L'_m}D_{l_1l'_1}(g)\cdots\times D_{l_ml'_m}(g)c_{l'_1}(\mathbf{k}_1)\cdots c_{l'_m}(\mathbf{k}_m).$$

Further,

$$\begin{aligned}\pi g\Psi(\mathbf{a},L_m)&=\prod_{ij}(\rho_j)^{a_{ij}}(k_{\bar\pi(i)j})^{a_{ij}}D_{l_1l'_{\bar\pi(1)}}(g)\cdots D_{l_ml'_{\bar\pi(m)}}(g)c_{l'_{\bar\pi(1)}}(\mathbf{k}_{\bar\pi(1)})\cdots c_{l'_{\bar\pi(m)}}(\mathbf{k}_{\bar\pi(m)})\\&=\prod_{ij}(\rho_j)^{a_{ij}}(k_{\bar\pi(i)j})^{a_{ij}}D_{l_1l'_{\bar\pi(1)}}(g)\cdots D_{l_ml'_{\bar\pi(m)}}(g)c_{l'_1}(\mathbf{k}_1)\cdots c_{l'_m}(\mathbf{k}_m).\end{aligned}$$

Here $\bar\pi=\pi^{-1}$.

Now suppose that $\bar\pi(q)=m$. Then in the above expression there occurs a factor

$$\begin{aligned}(k_{\bar\pi(q)j})^{a_{qj}}&=(-k_{1j}-k_{2j}-\cdots-k_{m-1j})^{a_{qj}}\\&=(-1)^{a_{qj}}\sum a_{qj}!(a_{qj1}!\cdots a_{qjm-1}!)^{-1}\times(k_{1j})^{a_{1j}}\cdots(k_{m-1j})^{a_{qjm-1}},\end{aligned}$$

where the sum is over all $a_{qji}\geq 0$ such that

$$\sum_{i=1}^{m-1}a_{qji}=a_{qj}.$$

Writing

$$\pi g\psi(\mathbf{a},L_m)=\sum_{a'L'_m}\Gamma(\mathbf{a},L_m;\mathbf{a}',L'_m)(\pi g)\psi(\mathbf{a}',L'_m),$$

we see that

$$\begin{aligned}\Gamma(\mathbf{a},L_m;\mathbf{a}',L'_m)=&\prod_j(-1)^{a_{qj}}a_{qj}!\prod_i(\rho_j)^{a_{ij}}\\&\times(a_{qji}!)^{-1}\delta(a'_{ij}-a_{\pi(i)j}-a_{qji})\\&\times D_{l_il'_{\bar\pi(i)}}(g),\end{aligned}\tag{5}$$

where all restrictions on the a_{ij} and a_{qji} as noted before hold, and we have defined $a_{mj}=0$ for all j. In this we have written

$$\prod_{ij}(k_{\bar\pi(i)j})^{a_{ij}}=\prod_{ij}(k_{ij})^{a_{\pi(i)j}},$$

where appropriate and $\delta(n)$ is the Kronecker delta δ_{n0}.

We can then form the trace of Eq. (5) to find the character of Γ, giving

$$\begin{aligned}\mathrm{tr}\,\Gamma(\pi g)=&\sum_{aL_m}\prod_j(-1)^{a_{qj}}a_{qj}!\prod_i(\rho_j)^{a_{\pi(i)j}}\\&\times(a_{qji}!)^{-1}\delta(a_{ij}-a_{\pi(i)j}-a_{qji})D_{l_il_{\bar\pi(i)}}(g).\end{aligned}\tag{6}$$

Now the sum over the L_m gives a factor, if π (and hence $\bar\pi$) has cycle structure $(\nu_1,\nu_2,\ldots,\nu_m)$,

$$(\chi(g))^{\nu_1}(\chi(g^2))^{\nu_2}\cdots(\chi(g^m))^{\nu_m},$$

which follows by writing

$$\prod_i D_{l_il_{\bar\pi(i)}}(g)=D_{l_1l_{\bar\pi(1)}}(g)\cdots D_{l_{\bar\pi^{(\nu_s-1)}(1)}l_1}(g)\cdots,$$

if, for example, 1 lies in a cycle of length ν_s, etc.

To carry out conveniently the sum over $\mathbf{a}$ we use the fact that a Kronecker delta $\delta(n)$ may be represented by

$$\delta(n)=\frac{1}{2\pi i}\oint_c dz\,z^{-(n+1)},\tag{7}$$

where the contour c includes the origin. With this we can extend the sum over $\mathbf{a}$ in Eq. (6) to all $\mathbf{a}\in\mathcal{N}^{(m-1)d}$, by including $\delta(n-\Sigma a_{ij})$ in this form. Then

$$\begin{aligned}\mathrm{tr}\,\Gamma(\pi g)=&\frac{1}{2\pi i}\oint_c dz\,z^{-(n+1)}\sum_a\prod_j(-1)^{a_{qj}}a_{qj}!\\&\times\prod_i\frac{1}{a_{qji}!}(\rho_j)^{a_{ij}}z^{a_{ij}}\delta(a_{ij}-a_{\pi(i)j}-a_{qji})\\&\times\prod_l(\chi(g^l))^{\nu_l}.\end{aligned}\tag{8}$$

Now suppose m lies in cycle of length l in π, i.e., $(q\cdots sm)$. The Kronecker delta in the sum over $\mathbf{a}$ in Eq. (8) constrains $a_{ij}\geq a_{\pi(i)j}$. Then

$$a_{qj}\geq a_{\pi(q)j}\geq a_{\pi^2(q)j}\geq\cdots\geq a_{sj}\geq a_{mj}=0,$$

such that

$$a_{\pi^{\mu-1}(q)j}-a_{\pi^\mu(q)j}=a_{qj\pi^{\mu-1}(q)}.$$

Note that $s=\pi^{l-2}(q)$. Also, for l_i in another cycle in $\pi=(l_1l_2\cdots l_t)$, the Kronecker delta restricts the sum to all a_{ij} such that

$$a_{i_1j}\geq a_{i_2j}\geq\cdots\geq a_{i_tj}\geq a_{i_1j},$$

which implies that they are all equal.

The portion of Eq. (8) involving a_{ij}, i in the cycle containing m, is then

$$\begin{aligned}\prod_j\sum_{a_{qj}=0}^{\infty}&(-1)^{a_{qj}}a_{qj}!(z\rho_j)^{a_{qj}}\sum_{a_{\pi(q)j}=0}^{a_{qj}}\frac{(z\rho_j)^{a_{\pi(q)j}}}{(a_{qj}-a_{\pi(q)j})!}\\&\times\sum_{a_{\pi^2(q)j}=0}^{a_{\pi(q)j}}\frac{(z\rho_j)^{a_{\pi^2(q)j}}}{(a_{\pi(q)j}-a_{\pi^2(q)j})!}\\&\cdots\sum_{a_{\pi^{l-2}(q)j}=0}^{a_{\pi^{l-3}(q)j}}\frac{(z\rho_j)^{a_{\pi^{l-2}(q)j}}}{(a_{\pi^{l-3}(q)j}-a_{\pi^{l-2}(q)j})!(a_{\pi^{l-2}(q)j})!}.\end{aligned}\tag{9}$$

Noting that

$$\sum_{i=0}^{p}\frac{x^i}{(p-i)!i!}=\frac{1}{p!}\sum_{i=0}^{p}\binom{p}{i}x^i1^{p-i}=\frac{(1+x)^p}{p!}$$

and performing the last $l-2$ sums in Eq. (9), it becomes

$$\prod_j \sum_{a_{qj}} (-1)^{a_{qj}}(z\rho_j)^{a_{qj}}(1+z\rho_j(1+z\rho_j(\cdots(1+z\rho_j))))^{a_{qj}}$$

$$=\prod_j (1+z\rho_j(\cdots(1+z\rho_j)))^{-1}$$

$$=\prod_j \left(\sum_{i=0}^{l-1}(z\rho_j)^i\right)^{-1}$$

$$=\prod_j \frac{1-z\rho_j}{1-(z\rho_j)^l}. \tag{10}$$

Furthermore the contribution to Eq. (8) from the sum over **a** for the cycles of length r [if the length of $(q\cdots sm)$ is not r] is [where $p=\nu_r r-(r-1)=(\nu_r-1)r+1$]

$$\prod_j \sum_{a_{1j}=0}^{\infty}(z\rho_j)^{ra_{1j}}\cdots\sum_{a_{pj}=0}^{\infty}(z\rho_j)^{ra_{pj}}$$

$$=\prod_j (1-z^r(\rho_j)^r)^{-\nu_r}$$

$$=\det(I-z^rV^r(g))^{-\nu_r}. \tag{11}$$

Combining the results of Eqs. (6)–(11) we have

$$\operatorname{tr}\Gamma(\pi g)=\frac{1}{2\pi i}\oint_c dz\, z^{-(n+1)}\det(I-zV(g))$$

$$\times\prod_{i=1}^{m}\det(I-z^iV^i(g))^{-\nu_i}(\chi(g^i))^{\nu_i}. \tag{12}$$

The contribution $c_{nm}(g)$ of the element g to c_{nm} is given by the number of times $\Gamma(\pi g)$ contains the identity representation of S_m. From Eq. (12) it is clear that $\operatorname{tr}\Gamma(\pi g)$ is a class function on S_m in that it depends only on the cycle structure of π. The number of elements in a class (ν) is $m!\Pi_i(i^{\nu_i}\nu_i!)^{-1}$ so that

$$c_{nm}(g)=\underset{\pi}{M}\ \operatorname{tr}\Gamma(\pi g)$$

$$=\sum_{(\nu)}\frac{1}{2\pi i}\oint_c dz\, z^{-(n+1)}$$

$$\times\det(I-zV(g))\prod_i\left(\frac{\chi(g^i)}{i\det(I-z^iV^i(g))}\right)^{\nu_i}\frac{1}{\nu_i!}, \tag{13}$$

where the sum over (ν) is restricted to all (ν) such that $\Sigma_i i\nu_i=m$ as i ranges from 1 to m. Using Eq. (7) to constrain the ν_i and then summing over all (ν) leaves Eq. (13)

$$c_{nm}(g)=\left(\frac{1}{2\pi i}\right)^2\oint\oint dz\,du\, z^{-(n+1)}u^{-(m+1)}$$

$$\times\det(I-zV(g))\prod_i\sum_{\nu_i}\frac{1}{\nu_i!}\left(\frac{u^i\chi(g^i)}{i\det(I-z^iV^i(g))}\right)^{\nu_i}$$

$$=(2\pi i)^{-2}\oint\oint dz\,du\, z^{-(n+1)}u^{-(m+1)}$$

$$\times\det(I-zV(g))\exp\left(\sum_{i=1}^{m}\frac{u^i\chi(g^i)}{i\det(I-z^iV^i(g))}\right). \tag{14}$$

Equation (14) was derived for $m\geq 2$ but can be extended to $m=1$. This may be seen by noting that

$$\frac{1}{2\pi i}\oint du\, u^{-(m+1)}f(u)=\frac{1}{m!}\left.\frac{df^m(u)}{du^m}\right|_{u=0}. \tag{15}$$

When Eq. (15) is applied to Eq. (14) for $m=1$, one obtains Eq. (4). Note, however, that Eq. (14), when $m=0$, gives

$$c_{n0}(g)=\frac{1}{n!}\frac{d^n}{ds^n}\det(I-sV(g))|_{s=0}, \tag{16}$$

rather than Eq. (3).

From Eq. (15) one can quickly see that the summation in Eq. (14) on i can be extended to infinity. This fact also then allows us to recognize Eq. (14) as just the coefficient of s^nt^m in the power series

$$F'(s,t,g)=\sum_{nm}c_{nm}(g)s^nt^m,$$

if we identify

$$F'(s,t,g)=\det(I-sV(g))\exp\left(\sum_{i=1}^{\infty}\frac{t^i\chi(g^i)}{i\det(I-s^iV^i(g))}\right).$$

Taking into account the discrepancy between Eqs. (3) and (16), and since

$$c_{nm}=\underset{g}{M}\ c_{nm}(g),$$

it then follows that

$$F(s,t)=\underset{g}{M}\ F(s,t,g),$$

where

$$F(s,t,g)=F'(s,t,g)-\det(I-sV(g))+1$$

or

$$F(s,t)=1+\underset{g}{M}\ \det(I-sV(g))$$

$$\times\left[\exp\left(\sum_{i=1}^{\infty}\frac{t^i\chi(g^i)}{i\det(I-s^iV^i(g))}\right)-1\right]. \tag{17}$$

It does not appear at present that Eq. (17) can be simplified further. As a check set $s=0$ in Eq. (17), leaving

$$F(0,t)=\underset{g}{M}\ \exp\left(\sum_{i=1}^{\infty}\operatorname{tr}\left(\frac{t^iD(g)^i}{i}\right)\right)$$

$$=\underset{g}{M}\ \exp(-\operatorname{tr}\ln(I-tD(g)))$$

$$=\underset{g}{M}\ \det(I-tD(g))^{-1},$$

which one recognizes as the expression for the simple Molien function.[2]

Equation (17) still is not entirely satisfactory for general calculations. However, it is useful in finding particular terms of arbitrary order m and n. This is so particularly if one uses computer algebra to perform the expansion. Indeed, for most representations of a space group G of interest, there are a finite number of $D(g)$ and $V(g)$ and one can readily find $F(s,t)$ to any desired order in s,t.

As an example, for the $F\Gamma_2^-$ representation of $R\,\bar{3}c$ (notation is that of Ref. 8)

$$F(s,t) = 1 + t^2(1 + 4s^2 + 9s^4 + 16s^6 + \cdots) + t^3(s + 10s^3 + 42s^5 + \cdots) + t^4(2 + 14s^2 + 103s^4 + \cdots) + t^5(3s + 50s^3 + \cdots) + t^6(3 + 31s^2 + \cdots) + \cdots,$$

where the computer algebra MACSYMA[9] has been employed.

As an example of the application of Eq. (17) to an infinite group, apply it to the faithful, or $l = 1$, representation of SO(3). The calculation is greatly simplified by noting that $F(s,t,g)$ is a class function, since it involves only similarity invariant traces and determinants. Then the sum over g reduces to an integral over the rotation angle, with an appropriate weighting factor. The result is

$$F(s,t) = 1 + (1 + s + 2s^2 + s^3 + 2s^4 + \cdots)t^2 + (s + s^2 + 5s^3 + \cdots)t^3 + (1 + s + 6s^2 + \cdots)t^4 + (s + \cdots)t^5 + (1 + \cdots)t^6 + \cdots.$$

Thus, for example, this representation has a Lifshitz invariant, i.e., $c_{12} = 1$.

[1]R. P. Stanley, Bull. Am. Math. Soc. **1**, 475 (1979).

[2]J. L. Birman and M. V. Jarič, J. Math. Phys. **18**, 1456, 1459 (1977).

[3]L. Michel and M. Mozrzymas, Lect. Notes Phys. **79**, 447 (1978).

[4]R. Berenson, J. Math. Phys. **23**, 1551 (1982).

[5]J. W. Felix and D. M. Hatch, Phys. Rev. B. (to be published); Jeffrey W. Felix, Ph. D. thesis, Brigham Young University, Utah, 1984.

[6]J. W. Felix and D. M. Hatch, Phys. Rev. Lett. **53**, 2425 (1984).

[7]G. Ya Lyubarskii, *Applications of Group Theory to Physics* (Pergamon, Oxford, 1961), p. 71.

[8]C. J. Bradley and A. P. Cracknell, *Mathematical Theory of Symmetry in Solids* (Oxford U. P., London, 1972).

[9]MACSYMA (©) 1976, 1983 MIT; Enhancements (©) 1983, Symbolics, Inc. All rights reserved.

A projection-based solution to the Sp(2N) state labeling problem

M. D. Gould
Department of Chemistry, University of Western Australia, Nedlands, Western Australia
E. G. Kalnins
Department of Mathematics, University of Waikato, Hamilton, New Zealand

(Received 24 May 1983; accepted for publication 23 November 1984)

A projection-based solution to the symplectic group state labeling problem is presented. The approach yields a nonorthogonal Gel'fand–Tsetlin basis for the irreducible representations of Sp($2n$). A method for evaluating the corresponding overlap coefficients is discussed. The action of the Sp($2n$) generators, in the basis obtained, is determined and some matrix element formulas are derived. The results obtained are comparable to the matrix element formulas for O(n) and U(n).

I. INTRODUCTION

The theory of Lie groups has been established as an invaluable tool in physical applications where they usually appear as the symmetry group of the system. The states of a physical system are then to comprise irreducible representations of the symmetry group. Lie groups afford not only convenient analytic methods but in practice are essential to the numerical solution of the equations of motion of the system by allowing the Hamiltonian to be broken into a convenient block form. Lie groups also provide suitable labels (i.e., quantum numbers) for physical states, even though such Lie groups need not be symmetry groups.

Thus, from the point of view of physical applications the principle problem to be solved in the representation theory of a semisimple (compact) Lie group is the complete determination of the basis states of an irreducible representation. The first major step in this direction was made by Gel'fand and Tsetlin,[1] who constructed, with a full set of labels, a complete set of basis vectors for the irreducible representations of the orthogonal and unitary groups. The work of Gel'fand and Tsetlin on U(n) was extended by Baird and Biedenharn,[2] who revealed the group theoretic nature of the Gel'fand–Tsetlin results. The work of Baird and Biedenharn has recently been extended to O(n) by Gould.[3]

The solution to the O(n) and U(n) state labeling problem, as proposed by Gel'fand and Tsetlin,[1] relies on the fact that O(n) and U(n) admit a so-called canonical[4] chain of subgroups whose Casimir invariants provide a complete set of labels for the irreducible representations. For general subgroup chains, however, this method of labeling is incomplete and it is necessary to supplement the Casimir invariants of the subgroup chain one is considering with additional labeling operators. In such a case there still remains the problem of finding the eigenvalues of these extra invariant operators, which are known to be irrational in general (and thus the action of the group generators in such a basis is likely to be complicated). This behavior is typical of the general state labeling problem. Examples are afforded by the subgroup embeddings U($2n$)$\supset$Sp($2n$), U(n)$\supset$O(n), U($n+m$)$\supset$U(n) $\times$U(m), Sp($2n$)$\supset$Sp($2n-2$), etc.

An alternative approach to the state labeling problem is to use the method of projection, which has proved in the past to be a powerful tool for handling the multiplicity problem numerically and, in some cases, analytically. The methods of projection have been successfully employed by Elliot[5] to the U(3)$\supset$O(3) state labeling problem. More recently the methods of projection have been applied to give a solution to the Clebsch–Gordan multiplicity problem for a semisimple Lie group G (i.e., the $G\times G\supset G$ state labeling problem). A detailed account of the various methods of projection can be found in Moshinsky *et al.*,[6] Asherova and Smirnov,[7] and Edwards and Gould.[8]

In this paper we present a projection-based solution to the symplectic group state labeling problem. Our method consists of embedding an irreducible representation of the symplectic group Sp($2n$) in a suitable representation of the unitary group U($2n$). A basis for the irreducible representations of Sp($2n$) is then obtained by (central) projection from a suitable set of Gel'fand–Tsetlin (GT) basis states for U($2n$). This leads to a GT-type labeling scheme for Sp($2n$) in analogy with the solutions to the O(n) and U(n) state labeling problems. The principle feature of our approach is that the action of the Sp($2n$) generators in the basis obtained is simple and comparable to the O(n) and U(n) cases. By expanding the Sp($2n$) generators in terms of U($2n$) generators we are able to deduce the action of the Sp($2n$) generators in the basis obtained. We give explicit matrix element formulas for certain generators but only give the general form of action for the remaining generators. However, as it turns out, the matrix elements of the elementary generators (which generate the symplectic group Lie algebra) are not difficult to obtain and will be evaluated in a forthcoming publication. It is evident that the approach of this paper may be extended to other subgroup embeddings such as, for example, the problem of obtaining a weight basis for the irreducible representations of O(n).

The solution to the symplectic group state labeling problem, as proposed in this paper, suffers from the disadvantage that the basis obtained is nonorthogonal. To this end we have found it convenient to introduce a dual Gel'fand–Tsetlin pattern labeling. One (upper) pattern, which has direct group theoretical significance, refers to the representation labels of the group Sp($2n$) and its subgroups Sp($2n-2$), ..., Sp(2). The other (lower) pattern carries the labels of the U($2n$) GT states from which we are projecting. Thus our basis is orthogonal with respect to the upper patterns but not with respect to the lower ones. However, the projection operators used may be

 0022-2488/85/071446-12$02.50

constructed as a polynomial in the universal Casimir elements of the subgroups Sp(2n), Sp(2$n-2$), ..., Sp(2) and thus the overlap coefficients may, in principle, be evaluated using the known matrix element formulas of the U(2n) generators. The problem of evaluating the overlap coefficients and various simplifications is discussed in the final section of the paper.

The dual pattern labeling of this paper may be compared to the approach of Louck *et al.*[9,10] to the U(N) tensor operator problem. In this case two GT patterns appear. One, which has direct group theoretical significance, refers to the components of the tensor operator. The other is an operator pattern which was shown in Ref. 8 to correspond to a projection-type labeling in analogy with the labeling scheme proposed in this paper. We remark that this dual pattern labeling for the symplectic groups also appears in the work of Zhelobenko[11] but without any group theoretical significance.

Other developments in connection with the symplectic group have been made by Lohe and Hurst,[12] who have advocated the use of modified boson operators as a method of constructing basis states for the irreducible representations of Sp(2n) in analogy with the boson polynomials used in the theory of U(n). Explicit matrix element formulas, in certain degenerate representations of Sp(2n), have recently been obtained by Klimyk.[13] The method of raising and lowering operators to construct a basis for the irreducible representations of Sp(2n) has been advocated by Bincer[14] and Mickelsson.[15] The symplectic groups also figure prominently in Cartan's classification of homogeneous spaces, which afford certain degenerate representations of Sp(2n) which have been studied by Pajas and Raczka[16] and Kalnins and Gould.[17]

II. FUNDAMENTALS

We begin by introducing the symplectic group Sp(2n) as a subgroup of the unitary group U(2n). The $(2n)^2$ generators a_{ij} $(i,j=1,...,2n)$ of the Lie group U(2n) satisfy the commutation relations

$$[a_{ij},a_{kl}]=\delta_{kj}a_{il}-\delta_{il}a_{kj}$$

and are, moreover, required to satisfy the Hermiticity condition

$$a_{ij}^{\dagger}=a_{ji}$$

on finite-dimensional (i.e., unitary) representations of the group. In order to define the symplectic subgroup of U(2n) we introduce a nonsingular antisymmetric metric $g_{ij}=-g_{ji}$ $(i,j=1,...,2n)$. One may then take for the infinitesimal generators of the Lie group Sp(2n) the $n(2n+1)$ independent operators

$$\alpha_{ij}=g_{ip}a_{pj}+g_{jp}a_{pi}=\alpha_{ji},$$

where we have summed over repeated index p from 1 to $2n$. These generators satisfy the commutation relations

$$[\alpha_{ij},\alpha_{kl}]=g_{kj}\alpha_{il}-g_{il}\alpha_{kj}+g_{ki}\alpha_{jl}-g_{jl}\alpha_{ki}.$$

Without loss of generality we choose the symplectic group metric g_{ij} to be given by

$$g_{ij}=\begin{cases}\delta_{j,i+1}, & i \text{ odd},\\ -\delta_{j,i-1}, & i \text{ even},\end{cases}\quad i,j=1,...,2n.$$

We remark that as far as the representation theory is concerned the actual choice of symplectic group metric is immaterial and there exist several other standard choices, all of which should lead to equivalent formalisms.

It is useful to introduce the inverse metric g^{ij} defined by

$$g^{ij}g_{jk}=\delta^{i}{}_{k}.$$

This equation is to be understood in the sense of the summation convention, which we employ throughout the paper, where any repeated index is to be summed from 1 to $2n$ (unless otherwise stated). Note that the metric g_{ij} satisfies the property $g_{ij}g_{jk}=-\delta_{ik}$ (i.e., $g^2=-I$) from which it follows that $g^{ij}=-g_{ij}=g_{ji}$. Using the inverse metric we define generators

$$\alpha^{i}{}_{j}=g^{ik}\alpha_{kj}, \tag{1}$$

which satisfy the commutation relations

$$[\alpha^{i}{}_{j},\alpha^{k}{}_{l}]=\delta^{k}{}_{j}\alpha^{i}{}_{l}-\delta^{i}{}_{l}\alpha^{k}{}_{j}+g^{ik}\alpha_{jl}-g_{jl}\alpha^{ki}, \tag{2}$$

where we define

$$\alpha^{ki}=g^{ij}\alpha^{k}{}_{j}.$$

We note moreover that the generators (1) satisfy the Hermiticity condition

$$(\alpha^{i}{}_{j})^{\dagger}=\alpha^{j}{}_{i}.$$

The advantage of working with the generators (1) is that they are automatically in Cartan–Weyl form. We choose as a Cartan subalgebra (CSA) the diagonal generators

$$\alpha^{i}{}_{i}=a_{ii}-(g_{ip})^2a_{pp},\quad i=1,...,2n$$

(where the repeated index p is to be summed from 1 to $2n$). However, only n of these operators are linearly independent so we only need consider the Cartan generators

$$\begin{aligned}h_i&=\alpha^{2i-1}{}_{2i-1}=a_{2i-1,2i-1}-a_{2i,2i}\\&=-\alpha^{2i}{}_{2i},\quad i=1,...,n,\end{aligned}$$

whose eigenvalues provide a unique labeling for the system of weights. Note, with our choice of metric, that the CSA for Sp(2n) is embedded in the CSA for U(2n).

To see that the generators (1) are in Cartan–Weyl form we note that the commutation relations (2) imply the result

$$[h_i,\alpha^{k}{}_{l}]=(\delta^{k}{}_{2i-1}-\delta^{l}{}_{2i-1}+\delta^{l}{}_{2i}-\delta^{k}{}_{2i})\alpha^{k}{}_{l}. \tag{3}$$

If we introduce the fundamental weights Δ_r $(r=1,...,n)$ consisting of 1 in the rth position and zeros elsewhere one sees immediately from Eq. (3) that the roots for the symplectic group Lie algebra are given by the weights

$$\pm(\Delta_i+\Delta_j),\quad i\leqslant j\quad\text{and}\quad\pm(\Delta_i-\Delta_j),\quad i<j.$$

We take as a system of positive roots the weights

$$\Delta_i+\Delta_j\quad(i\leqslant j),\quad \Delta_i-\Delta_j\quad(i<j).$$

The corresponding generators are given by

$$\alpha^{2i-1}{}_{2j}\quad(i\leqslant j)\quad\text{and}\quad\alpha^{2i-1}{}_{2j-1}\quad(i<j), \tag{4}$$

respectively, which we henceforth refer to as raising generators. Note that the raising generators (4) are given in terms of U(2n) generators by

$$\alpha^{2i-1}{}_{2j} = a_{2i-1,2j} + a_{2j-1,2i} = \alpha^{2j-1}{}_{2i},$$
$$\alpha^{2i-1}{}_{2j-1} = a_{2i-1,2j-1} - a_{2j,2i} = -\alpha^{2j}{}_{2i}. \tag{5}$$

By taking the Hermitian conjugate of Eqs. (4) and (5) we obtain the set of lowering generators.

We draw particular attention to the generators $\alpha^{2i-1}{}_{2i}, \alpha^{2i}{}_{2i-1}$ which may be expressed in terms of U(2n) generators according to

$$\alpha^{2i-1}{}_{2i} = 2a_{2i-1,2i}, \quad \alpha^{2i}{}_{2i-1} = 2a_{2i,2i-1}. \tag{6}$$

The operators

$$h_n = \alpha^{2n-1}{}_{2n-1}, \ \alpha^{2n-1}{}_{2n}, \ \alpha^{2n}{}_{2n-1} \tag{7}$$

form the generators of the subgroup Sp(2) of Sp(2n). The generators (7) together with the Sp($2n-2$) generators $\alpha^i{}_j (i,j = 1, ..., 2n-2)$ form the generators of the subgroup Sp($2n-2$)$\times$Sp(2) of Sp(2n).

We note that the symplectic group Lie algebra is generated (as a Lie algebra) by the elementary generators

$$\alpha^{2i-1}{}_{2i+1}, \quad \alpha^{2i+1}{}_{2i-1}, \quad \alpha^{2n-1}{}_{2n}, \quad \alpha^{2n}{}_{2n-1},$$
$$i = 1, ..., n-1. \tag{8}$$

Every symplectic group generator $\alpha^i{}_j$ may be obtained by repeated commutation with generators of the form (8).

With regard to the group U(2n) we follow the notation of Gould.[18] We choose as a CSA for the Lie algebra of U(2n) the Abelian Lie algebra spanned by the diagonal generators a_{ii} ($i = 1,...,2n$) whose eigenvalues uniquely label the weights of U(2n). With respect to the usual lexicographical ordering imposed on the weights we see that the U(2n) generators a_{ij} with $i<j$ (resp. $i>j$) are raising (resp. lowering) generators.

We let L (resp. L_0) denote the Lie algebra of U(2n) [resp. Sp(2n)] and we let H (resp. H_0) denote the CSA of L (resp. L_0). The weights for L (resp. L_0) may be identified with the CSA dual H^* (resp. $H^*{}_0$) in an obvious manner. We let B (resp. B_0) denote the nilpotent Lie subalgebra of L (resp. L_0) generated by the raising generators and we let N (resp. N_0) denote the nilpotent Lie subalgebra of L (resp. L_0) generated by the lowering generators. We furthermore set

$$\overline{N} = N \oplus H, \quad \overline{B} = B \oplus H,$$
$$\overline{N}_0 = N_0 \oplus H_0, \quad \overline{B}_0 = B_0 \oplus H_0.$$

Note that the subalgebras $\overline{N}, \overline{B}$ (resp. $\overline{N}_0, \overline{B}_0$) are reductive Lie algebras. In this notation the Lie algebras L, L_0 may be written

$$L = H \oplus N \oplus B = \overline{N} \oplus B = N \oplus \overline{B},$$
$$L_0 = H_0 \oplus N_0 \oplus B_0 = \overline{N}_0 \oplus B_0 = N_0 \oplus \overline{B}_0.$$

We let U (resp. U_0) denote the universal enveloping algebra of L (resp. L_0) and we denote the universal enveloping algebras of H, N, B, H_0, N_0, B_0 etc., by $U(H)$, $U(N)$, $U(B)$, $U(H_0)$, $U(N_0)$, $U(B_0)$, respectively.

We now recall some basic facts on the structure of universal enveloping algebras (see, e.g., Humphreys[19]). According to the PBW theorem the universal enveloping algebra U of L may be written

$$U = U(N)U(H)U(B). \tag{9}$$

Now the Lie algebra L is generated (as a Lie algebra) by the CSA H together with the elementary generators

$$x_i = a_{i,i+1}, \quad y_i = a_{i+1,i}, \quad i = 1, ..., 2n-1.$$

The nilpotent subalgebra B (resp. N) is generated as an algebra by the set $\{x_i\}^{2n-1}{}_{i=1}$ (resp. $\{y_i\}^{2n-1}{}_{i=1}$). Again, by the PBW theorem, we may choose as a basis for the universal enveloping algebra $U(B)$ the identity $1 \in \mathbb{C}$ together with the set of all basis monomials of degree k ($k = 1,2,3,...$)

$$x_{i_1} x_{i_2} \cdots x_{i_k}, \quad 1 \leqslant i_r \leqslant 2n-1.$$

Letting k range over all positive integers and letting the integers i_r ($r = 1, ..., k$) take all possible values in the range 1, ..., $2n-1$ we thereby get a basis for $U(B)$. A similar analysis may be applied to the universal enveloping algebras $U(H)$ and $U(N)$. In view of Eq. (9) we may choose as a basis for U the set of all monomials of the form

$$u = nhb,$$

where n, h, and b are basis monomials for $U(N)$, $U(H)$, and $U(B)$, respectively.

We remark that we may impose a total ordering on the basis monomials with respect to their degrees. One may also impose a partial ordering on the basis monomials according to their weights under the adjoint action of H in U. A similar analysis may be applied to the algebras U_0, $U(H_0)$, $U(N_0)$, and $U(B_0)$.

We conclude this section by setting up an (associative) algebra homomorphism of $U(B)$ into U_0 which will be needed in the following section. From Eq. (5) we may write

$$\alpha^{2i}{}_{2i+1} = a_{2i,2i+1} + a_{2i+2,2i-1},$$
$$\alpha^{2i-1}{}_{2i} = 2a_{2i-1,2i}. \tag{10}$$

We then define a mapping

$$\theta: U(B) \to U_0, \tag{11}$$

defined by

$$\theta(a_{2i,2i+1}) = \alpha^{2i}{}_{2i+1},$$
$$\theta(a_{2i-1,2i}) = \tfrac{1}{2}\alpha^{2i-1}{}_{2i},$$
$$\theta(1) = 1, \quad i = 1, ..., n-1,$$

which we extend to an algebra homomorphism to all of $U(B)$; that is, if $b = x_{i_1} \cdots x_{i_k}$ is a basis monomial of $U(B)$ we define $\theta(b) = \theta(x_{i_1}) \cdots \theta(x_{i_n})$ and extend linearly. It is easily verified that the mapping θ, as defined above, satisfies the algebra homomorphism requirements

$$\theta(\alpha b_1 + \beta b_2) = \alpha\theta(b_1) + \beta\theta(b_2),$$
$$\theta(b_1 b_2) = \theta(b_1)\theta(b_2), \quad \text{for all } b_1, b_2 \in U(B),$$

and is well defined. One may check moreover that θ is one-to-one (although this fact will not be required).

In view of Eq. (10) we may write $\theta(x_i)$ in the form $\theta(x_i) = x_i + n_i$, where $n_i \in N$. We see from this that if b is a basis monomial of $U(B)$ then we may write

$$\theta(b) = b + w, \tag{12}$$

where w is a sum of basis monomials in U with $U(2n)$ weight strictly less than b.

III. PROJECTED GEL'FAND BASIS FOR Sp(2n)

In this section we obtain a (nonsymmetry-adapted) basis for the irreducible representations of Sp(2n) by a method of projection from the unitary group Gel'fand–Tsetlin (GT) basis. We begin by recalling the solution to the unitary group state labeling problem obtained by Gel'fand and Tsetlin.[1]

The U(2n) generators a_{ij}, where i and j are restricted to values 1, ..., m (for some positive integer m less than 2n) form the generators of the unitary subgroup U(m) of U(2n). We see therefore that U(2n) admits the canonical chain of subgroups

$$\mathrm{U}(2n)\supset\mathrm{U}(2n-1)\supset\mathrm{U}(2n-2)\supset\cdots\supset\mathrm{U}(1). \tag{13}$$

Following Baird and Biedenharn,[2] the Casimir invariants, for each subgroup occurring in the chain (13), provide a complete set of commuting (Hermitian) operators whose normalized eigenstates form an orthonormal basis (ONB) for the irreducible representations of U(2n). Now the eigenvalues of the Casimir invariants for the subgroup U(m) uniquely label the irreducible representations of U(m). An alternative characterization of the irreducible representations of U(m) is in terms of their highest weights $(\lambda_{1m}, \lambda_{2m},..., \lambda_{mm})$, where the λ_{im} are integers satisfying the inequalities $\lambda_{1\mathrm{m}}\geqslant\lambda_{2\mathrm{m}}\geqslant\cdots\geqslant\lambda_{\mathrm{mm}}$.

By virtue of Weyl's subgroup branching laws the highest weights of two groups U($m+1$) and U(m) occurring in the chain (13) are related by the inequalities

$$\lambda_{1,m+1}\geqslant\lambda_{1m}\geqslant\lambda_{2,m+1}\geqslant\lambda_{2m}\geqslant\cdots\geqslant\lambda_{mm}\geqslant\lambda_{m+1,m+1}.$$

The set of partitions for the chain (13) is most conveniently arranged into a GT pattern which labels the GT basis states for the irreducible representations of U(2n). More details are given in the paper by Baird and Biedenharn.[2]

The crucial property that makes the GT scheme work for U(2n) is that in the reduction of an irreducible representation of U($m+1$) into irreducible representations of U(m) all irreducible representations occur with unit multiplicity. This property is also shared by the orthogonal groups for which a GT scheme exists (see, e.g., Gould[3]).

One would ideally like to obtain a similar solution to the symplectic group state labeling problem. One method is to consider the subgroup chain

$$\mathrm{Sp}(2n)\supset\mathrm{Sp}(2n-2)\supset\cdots\supset\mathrm{Sp}(2)\supset\mathrm{U}(1), \tag{14}$$

but it is well known (see Zhelobenko[11]) that the Casimir invariants for the subgroups occurring in the chain (14) do not give a complete labeling. The situation may be improved by considering the refinement

$$\mathrm{Sp}(2n)\supset\mathrm{Sp}(2)\times\mathrm{Sp}(2n-2)\supset\mathrm{Sp}(2n-2)$$
$$\supset\cdots\supset\mathrm{Sp}(2)\times\mathrm{Sp}(2)\supset\mathrm{Sp}(2)\supset\mathrm{U}(1). \tag{15}$$

The subgroup chain (15) in fact works for the cases $n\leqslant 2$, where we have the local isomorphisms $\mathrm{Sp}(4)\cong\mathrm{O}(5)$, $\mathrm{Sp}(2)\times\mathrm{Sp}(2)\cong\mathrm{O}(4)$, and $\mathrm{Sp}(2)\cong\mathrm{O}(3)$. However, for $n>2$ the chain (15) fails in general to provide a complete set of labels.

This failure is due to the fact that in the reduction of an irreducible representation of Sp(2n) into irreducible representations of its subgroups Sp($2n-2$) or Sp(2)×Sp($2n-2$) multiplicities may occur and extra invariants are required to completely specify the irreducible representations.

One method of obtaining a solution to the symplectic group state labeling problem is to supplement the Casimir invariants of the chain (15) with an additional set of labeling invariants. However, there still remains the problem of obtaining the eigenvalues of these additional invariants which are known to be irrational in general (and thus the action of the generators in such a basis is likely to be complicated). We propose here an alternative solution based on projection.

For our purposes it suffices to consider irreducible representations $V(\lambda)$ of U(2n) with highest weights λ of the special form

$$\lambda=(\lambda_1,\lambda_2,...,\lambda_n,0,0,...,0).$$

The space $V(\lambda)$ constitutes a reducible representation of the subgroup Sp(2n). The branching rules for the reduction of $V(\lambda)$ into irreducible representations of Sp(2n) are given by Hamermesh[20] and Zhelobenko.[11] In general the irreducible representations of Sp(2n) occurring in the space $V(\lambda)$ occur with multiplicities [the U(2n)⊃Sp(2n) state labeling problem].

We recall however that the space $V(\lambda)$ contains exactly one copy of the irreducible representation of Sp(2n) with highest weight $\lambda=(\lambda_1,...,\lambda_n)$, which we denote by $V_0(\lambda)$. Thus the space $V_0(\lambda)$ may be obtained by central projection from $V(\lambda)$. To this end let $\Pi(\lambda)$ denote the set of all Sp(2n) highest weights occurring in $V(\lambda)$ but excluding $\lambda=(\lambda_1,...,\lambda_n)$. Then set

$$P^{\lambda}{}_{n}=\prod_{\nu\in\Pi(\lambda)}\left[\frac{\sigma_2-\chi_\nu(\sigma_2)}{\chi_\lambda(\sigma_2)-\chi_\nu(\sigma_2)}\right], \tag{16}$$

where $\sigma_2=\alpha^i{}_j\alpha^j{}_i$ is the second-order invariant of Sp(2n) and

$$\chi_\nu(\sigma_2)=2\sum_{r=1}^{n}\nu_r(\nu_r+2n-2r)$$

is the eigenvalue of σ_2 in the irreducible representation $V_0(\nu)$ of Sp(2n) (see, e.g., Green[21]). We have explicitly included the subscript n on the left-hand side of (16) to indicate we are considering the group Sp(2n). We have

$$V_0(\lambda)=P^{\lambda}{}_{n}V(\lambda). \tag{17}$$

We remark that the proof of Eq. (17) follows from noticing that the universal Casimir element separates λ from the weights in $\Pi(\lambda)$, a fact which is easily deduced from the known form of the weights in $\Pi(\lambda)$ (see, e.g., Hamermesh[20]).

Thus we may obtain a set of vectors spanning the irreducible representation $V_0(\lambda)$ by considering the central projector (16) applied to the U(2n) GT basis states of the space $V(\lambda)$. However, such a basis will be overcomplete and we need to consider a certain restricted set of GT vectors to yield a complete basis for the space $V_0(\lambda)$. To this end we restrict ourselves to Gel'fand vectors of the form

$$\left|\begin{matrix} \lambda_{1n} & \lambda_{2n} & \cdots & \lambda_{nn} & 0 & 0 & \cdots & 0 \\ \mu_{1n} & \mu_{2n} & \cdots & \mu_{nn} & 0 & \cdots & 0 & \\ \lambda_{1,n-1} & \lambda_{2,n-1} & \cdots & \lambda_{n-1,n-1} & 0 & \cdots & 0 & \\ \mu_{1,n-1} & \mu_{2,n-2} & \cdots & \mu_{n-1,n-1} & 0 & \cdots & 0 & \\ \ddots & \ddots & & & \iddots & \iddots & & \\ \lambda_{1,2} & \lambda_{2,2} & 0 & 0 & & & & \\ \mu_{1,2} & \mu_{2,2} & 0 & & & & & \\ \lambda_{1,1} & 0 & & & & & & \\ \mu_{1,1} & & & & & & & \end{matrix}\right\rangle \tag{18}$$

From the Gel'fand betweenness conditions the integers $\mu_{i,j}$ and $\lambda_{i,j}$ in the above pattern must satisfy

$$\lambda_{1,m} \geqslant \mu_{1,m} \geqslant \lambda_{2,m} \geqslant \mu_{2,m} \geqslant \cdots \geqslant \lambda_{m,m} \geqslant \mu_{m,m} \geqslant 0, \tag{19}$$

$$\mu_{1,m} \geqslant \lambda_{1,m-1} \geqslant \mu_{2,m} \geqslant \lambda_{2,m-1} \geqslant \cdots \geqslant \lambda_{m-1,m-1} \geqslant \mu_{m,m} \geqslant 0.$$

For simplicity we denote the GT state (18) by

$$\left|\begin{matrix} \lambda_{1n} & \lambda_{2n} & \cdots & & \lambda_{nn} \\ \mu_{1n} & \mu_{2n} & \cdots & & \mu_{nn} \\ \lambda_{1n-1} & \lambda_{2n-1} & \cdots & \lambda_{n-1,n-1} & \\ \mu_{1n-1} & \mu_{2n-1} & \cdots & \lambda_{n-1,n-1} & \\ \vdots & & & & \\ \mu_{12} & \mu_{22} & & & \\ \lambda_{11} & & & & \\ \mu_{11} & & & & \end{matrix}\right) \cdot \tag{20}$$

We remark that these are the patterns appearing in the work of Zhelobenko[11] (although no group theoretic meaning is attached to such patterns in his work).

It is our aim now to show that we may obtain a complete set of basis states for the space $V_0(\lambda)$ by central projection from the GT states (20). For simplicity we denote the space spanned by the GT basis states of the special form (18) by $A_0(\lambda)$ which we refer to as the space of allowed GT states.

Now from the work of Zhelobenko,[11] the number of allowed Gel'fand states (20) is precisely equal to dim $V_0(\lambda)$; that is,

$$\dim A_0(\lambda) = \dim V_0(\lambda)\,. \tag{21}$$

Thus in order to prove our result it suffices to prove that the central projector $P^\lambda{}_n$ is one-to-one on $A_0(\lambda)$, that is,

$$(\ker P^\lambda{}_n) \cap A_0(\lambda) = (0),$$

where

$$\ker P^\lambda{}_n = \{ v \in V(\lambda) \,|\, P^\lambda{}_n v = 0 \}\,.$$

Equation (21) then guarantees that $P^\lambda{}_n A_0(\lambda) = V_0(\lambda)$.

Note that the GT state (20) has U($2n$) weight $(\rho_1,\rho_2,\ldots,\rho_{2n})$, where

$$\rho_{2i-1} = \sum_{j=1}^{i} \mu_{j,i} - \sum_{j=1}^{i-1} \lambda_{j,i-1},$$

$$\rho_{2i} = \sum_{j=1}^{i} \lambda_{j,i} - \sum_{j=1}^{i} \mu_{j,i}, \quad i = 1,\ldots,n.$$

Now, since the CSA H_0 of Sp($2n$) is contained in the CSA H of U($2n$) we see that the state (20) is also a weight state of Sp($2n$) with weight $(\nu_1,\ldots,\nu_n)$, where

$$\nu_i = \rho_{2i-1} - \rho_{2i}, \quad i = 1,\ldots,n.$$

In particular the maximal allowable GT state

$$\Omega^\lambda{}_0 = \left|\begin{matrix} \lambda_{1n} & \lambda_{2n} & \cdots & & \lambda_{nn} \\ \lambda_{1n} & \lambda_{2n} & \cdots & & \lambda_{nn} \\ \lambda_{1n} & \lambda_{2n} & \cdots & \lambda_{n-1n} & \\ \lambda_{1n} & \lambda_{2n} & \cdots & \lambda_{n-1n} & \\ \vdots & & & & \\ \lambda_{1n} & & & & \\ \lambda_{1n} & & & & \end{matrix}\right) \tag{22}$$

has Sp($2n$) weight $\lambda = (\lambda_1,\lambda_2,\ldots,\lambda_n)$ which is the highest weight of $V_0(\lambda)$. It is easily shown that the GT state (22) in fact constitutes a highest weight state for Sp($2n$) by verifying that the elementary raising generators [see Eq. (8)] vanish on the state $\Omega^\lambda{}_0$. Thus we have immediately $\Omega^\lambda{}_0 \in V_0(\lambda)$ whence

$$P^\lambda{}_n \Omega^\lambda{}_0 = \Omega^\lambda{}_0\,.$$

We note moreover that $\Omega^\lambda{}_0$ has U($2n$) weight $(\lambda_1,0,\lambda_2,0,\ldots,\lambda_n,0)$ which is conjugate under the Weyl group to the U($2n$) highest weight $(\lambda_1,\lambda_2,\ldots,\lambda_n,0,\ldots,0)$ and hence occurs with unit multiplicity in $V(\lambda)$.

Some of the properties of the space $A_0(\lambda)$ are summarized in the following (notation as in Sec. II).

Lemma 1: (a) If $\Omega \in A_0(\lambda)$ is a U($2n$) weight vector then there exists a basis monomial $b \in U(B)$ such that $b\Omega = \alpha \Omega^\lambda{}_0$, $0 \neq \alpha \in \mathbb{C}$.

(b) If $\Omega \in A_0(\lambda)$ is arbitrary then there exists $b \in U(\overline{B})$ such that $b\Omega = \Omega^\lambda{}_0$.

(c) $A_0(\lambda)$ is a module over the algebras $U(N)$ and $U(\overline{N})$ and is cyclically generated by $\Omega^\lambda{}_0$; that is,

$$A_0(\lambda) = U(\overline{N})\Omega^\lambda{}_0 = U(N)\Omega^\lambda{}_0.$$

Proof: (a) Our proof of this result is based on a lengthy induction argument (which is not relevant to the remainder of this paper) and is presented in Appendix A for clarity of presentation.

(b) The proof of (b) is an immediate consequence of (a) since one may project out weight states from a general state with elements from $U(H)$.

(c) From the known action of the U($2n$) generators on GT states (see Baird and Biedenharn[2] and Gould[22]) it is clear that the space $A_0(\lambda)$ is stable under the action of N and $\overline{N}$. Hence $A_0(\lambda)$ constitutes a module over $U(N)$ and $U(\overline{N})$. Now set

$$A'_0(\lambda) = U(\overline{N})\Omega^{\lambda}{}_0 \subseteq A_0(\lambda).$$

We prove $A'_0(\lambda) = A_0(\lambda)$ by a contradiction argument. For suppose on the contrary $A'_0(\lambda) \neq A_0(\lambda)$ and choose $\Omega \in A_0(\lambda)$ orthogonal to $A'_0(\lambda)$. Then we have

$$0 = \langle \Omega \,|\, U(\overline{N})\Omega^{\lambda}{}_0 \rangle = \langle U(\overline{B})\Omega \,|\, \Omega^{\lambda}{}_0 \rangle . \tag{23}$$

But by (b) there exists $b \in U(\overline{B})$ such that $b\Omega = \Omega^{\lambda}{}_0$ whence (23) implies $0 = \langle b\Omega \,|\, \Omega^{\lambda}{}_0 \rangle = \langle \Omega^{\lambda}{}_0 | \Omega^{\lambda}{}_0 \rangle = 1$ and a contradiction has been reached. Thus our assumption was false and we must have $A'_0(\lambda) = U(\overline{N})\Omega^{\lambda}{}_0 = A_0(\lambda)$. From the PBW theorem we may write

$$U(\overline{N}) = U(N)U(H),$$

whence

$$A_0(\lambda) = U(\overline{N})\Omega^{\lambda}{}_0 = U(N)U(H)\Omega^{\lambda}{}_0 = U(N)\Omega^{\lambda}{}_0 .$$

Q.E.D

Now from Sec. II there exists an algebra homomorphism θ: $U(B) \rightarrowtail U_0$ [see Eq. (11)]. This result together with the previous lemma implies the following.

Lemma 2: If $\Omega \in A_0(\lambda)$ there exists $u \in U_0$ such that

$$\langle \Omega^{\lambda}{}_0 | u\Omega \rangle = 1.$$

Proof: Let $\Omega \in A_0(\lambda)$ be arbitrary. Then Ω may be decomposed into a sum of weight vectors

$$\Omega = \sum_{i=1}^{k} \Omega_{\mu_i},$$

where Ω_{μ_i} has U(2n) weight μ_i. Assume the weights $\mu_1,\dots,\mu_k$ are ordered in decreasing order with respect to the partial ordering induced by the positive roots (i.e., lexical ordering). Thus μ_1 is maximal in the set of weights $\{\mu_i\}^k{}_{i=1}$. Then there exists $h \in U(H)$ such that $h\Omega = \Omega_{\mu_1}$. Then by Lemma 1(a) there exists a basis monomial $b \in U(B)$ such that

$$bh\Omega = b\Omega_{\mu_1} = \alpha\Omega^{\lambda}{}_0, \quad \alpha \neq 0.$$

If we denote the U(2n) weight of the state $\Omega^{\lambda}{}_0$ by λ_0 then we see that the basis monomial b has weight $\lambda_0 - \mu_1$ and we may write $bh = h'b$ for suitable $h' \in U(H)$. Thus we have

$$\alpha = \langle \Omega^{\lambda}{}_0 | h'b\Omega \rangle = \langle h'\Omega^{\lambda}{}_0 | b\Omega \rangle = \beta \langle \Omega^{\lambda}{}_0 | b\Omega \rangle , \tag{24}$$

for some $\beta \neq 0$. Now we let $u = \theta(b) \in U_0$. From Eq. (12) we have $u = b + w$ where w is a sum of basis monomials in U with U(2n) weight strictly less than $\lambda_0 - \mu_1$. We may thus write

$$u\Omega = b\Omega + w\Omega = b\Omega + \sum_{i=1}^{k} w\Omega_{\mu_i}.$$

Since μ_1 is maximal in the set of U(2n) weights $\{\mu_j\}^k{}_{j=1}$ and since w has U(2n) weight strictly less than $\lambda_0 - \mu_1$ it follows that $w\Omega_{\mu_i}$ has weight, strictly less than $\lambda_0 - \mu_1 + \mu_i$, which cannot equal λ_0. Thus each state $w\Omega_{\mu_i}$ must be orthogonal to $\Omega^{\lambda}{}_0$, from which we obtain, in view of (24),

$$\langle \Omega^{\lambda}{}_0 | u\Omega \rangle = \langle \Omega^{\lambda}{}_0 | b\Omega \rangle = \alpha/\beta \neq 0.$$

Replacing $u \in U_0$ by $(\beta/\alpha)u$ the result is seen to follow. Q.E.D.

We are now in a position to prove our main result.

Theorem 1: $[\ker P^{\lambda}{}_n] \cap A_0(\lambda) = (0)$ and $P^{\lambda}{}_n A_0(\lambda) = V_0(\lambda)$. In particular if $\{\Omega_i\}$ is a basis for $A_0(\lambda)$ then the projected states $\widetilde{\Omega}_i = P^{\lambda}{}_n\Omega_i$ constitute a basis $\{\widetilde{\Omega}_i\}$ for $V_0(\lambda)$.

Proof: Clearly $P^{\lambda}{}_n A_0(\lambda) \subseteq V_0(\lambda)$. We prove $P^{\lambda}{}_n$ is one-to-one on $A_0(\lambda)$ using a contradiction argument. Suppose on the contrary there exists $\Omega \in A_0(\lambda)$ such that $P^{\lambda}{}_n\Omega = 0$. But Lemma 2 implies there exists $u \in U_0$ such that $\langle u\Omega \,|\, \Omega^{\lambda}{}_0 \rangle = 1$. Now since $P^{\lambda}{}_n$ commutes with the action of Sp(2n) we have

$$0 = P^{\lambda}{}_n\Omega = uP^{\lambda}{}_n\Omega = P^{\lambda}{}_n u\Omega.$$

Thus

$$0 = \langle P^{\lambda}{}_n u\Omega \,|\, \Omega^{\lambda}{}_0 \rangle = \langle u\Omega \,|\, P^{\lambda}{}_n\Omega^{\lambda}{}_0 \rangle$$
$$= \langle u\Omega \,|\, \Omega^{\lambda}{}_0 \rangle = 1$$

and a contradiction has been reached. Thus our assumption was false and $P^{\lambda}{}_n$ must be one-to-one on $A_0(\lambda)$; that is, $[\ker P^{\lambda}{}_n] \cap A_0(\lambda) = (0)$. Since $P^{\lambda}{}_n A_0(\lambda) \subseteq V_0(\lambda)$ Eq. (21) implies that $P^{\lambda}{}_n A_0(\lambda) = V_0(\lambda)$.

Thus if $\{\Omega_i | i = 1,\dots,d_\lambda = \dim V_0(\lambda)\}$ denotes a basis for $A_0(\lambda)$ then the projected states $P^{\lambda}{}_n\Omega_i$ constitute a basis for $V_0(\lambda)$. Q.E.D.

As a particular case of the above result we may choose the GT states (20) as a basis for $A_0(\lambda)$ and the projected states constitute a basis for $V_0(\lambda)$ which we denote by

$$\left|\begin{matrix} \lambda_{1,n} & \lambda_{2,n} & \cdots & & \lambda_{nn} \\ \mu_{1n} & \mu_{2n} & \cdots & & \mu_{nn} \\ \lambda_{1n-1} & \lambda_{2n-1} & \cdots & \lambda_{n-1,n-1} & \\ \mu_{1n-1} & \mu_{2n-1} & \cdots & \mu_{n-1,n-1} & \\ \vdots & & & & \\ \lambda_{12} & \lambda_{22} & & & \\ \mu_{12} & \mu_{22} & & & \\ \lambda_{11} & & & & \\ \mu_{11} & & & & \end{matrix}\right\rangle = P^{\lambda}{}_n \left|\begin{matrix} \lambda_{1n} & \lambda_{2n} & \cdots & & \lambda_{nn} \\ \mu_{1n} & \mu_{2n} & \cdots & & \mu_{nn} \\ \lambda_{1n-1} & \lambda_{2n-1} & \cdots & \lambda_{n-1,n-1} & \\ \mu_{1n-1} & \mu_{2n-1} & \cdots & \mu_{n-1,n-1} & \\ \vdots & & & & \\ \lambda_{12} & \lambda_{22} & & & \\ \mu_{12} & \mu_{22} & & & \\ \lambda_{11} & & & & \\ \mu_{11} & & & & \end{matrix}\right\rangle . \tag{25}$$

These states form a complete basis (of weight vectors) for the space $V_0(\lambda)$ which we refer to as the projected Gel'fand–Tsetlin (PGT) basis. We remark however that the PGT basis (25) is not orthonormal and moreover is not symmetry adapted to either of the subgroup chains (14) or (15). We shall consider the problem of constructing a symmetry-adapted basis for the space $V_0(\lambda)$ in the next section.

The labels appearing in the PGT basis for the space $V_0(\lambda)$ are those of the GT states from which we project. This method of labeling may be compared with a recently announced solution (see Edwards and Gould[8]) to the Clebsch–Gordan multiplicity problem (for arbitrary semisimple Lie algebras) where the highest weight states occurring in the tensor product representation $V(\lambda) \otimes V(\mu)$ are obtained by central projection from states of the form $e_i \otimes e^{\mu}{}_{+}$, where $\{e_i\}$ denotes a basis for the irreducible representation $V(\lambda)$ and $e^{\mu}{}_{+}$ is the maximal weight vector of $V(\mu)$. Thus the maximal weight states (and hence the irreducible representations they generate) may be labeled by the vector $e_i \otimes e^{\mu}{}_{+}$ (or equivalently e_i) from which these states are projected. In the case of $U(N)$ where $\{e_i\}$ is the GT basis, this leads to a GT pattern labeling for the irreducible representations occurring in the tensor product representation $V(\lambda) \otimes V(\mu)$. In terms of the equivalent problem of determining all tensor operators for a semisimple Lie algebra this leads to a GT pattern labeling for tensor operators which is closely related to the operator patterns of Biedenharn *et al.*[9,10] More details are given in Ref. 8.

Although the basis (25) is nonorthogonal one may obtain, at least in principle, the overlap coefficients for this basis from the known matrix elements of the U(2n) generators together with the explicit form (16) for the projection operators $P^{\lambda}{}_{n}$. Whether this leads to an efficient algorithm for adaption to computers (or ideally analytic manipulation) remains to be seen. However having obtained the overlap coefficients the action of the symplectic group generators in the basis (25) may be obtained from the known matrix elements of the U(2n) generators, viz.,

$$\begin{aligned}\alpha_{ij}|(\lambda)\rangle &= \alpha_{ij}P^{\lambda}{}_{n}|(\lambda)) \\ &= P^{\lambda}{}_{n}\alpha_{ij}|(\lambda)) \\ &= P^{\lambda}{}_{n}\,[g_{ip}a_{pj} + g_{jp}a_{pi}]\,|(\lambda)),\end{aligned}$$

where (λ) denotes an allowed GT pattern for Sp(2n). The trouble with this method, however, is that the Sp(2n) generators α_{ij} do not leave the space $A_0(\lambda)$ invariant and hence it is necessary to obtain the matrix elements of $P^{\lambda}{}_{n}$ in the GT basis for $V(\lambda)$ [and not just for $A_0(\lambda)$]. This deficiency will be removed, by considering a symmetry-adapted basis, in the following section.

IV. SYMMETRY-ADAPTED BASIS FOR Sp(2n)

As mentioned in Sec. III the trouble with the PGT basis (25) is that it is not symmetry adapted to the subgroup chain (14). In order to obtain a symmetry-adapted basis we need to apply the above projective scheme recursively for each of the subgroups Sp(2n), Sp(2$n-2$),...,Sp(2).

We denote the PGT state (25) by the simpler notation

$$\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle = P^{\lambda}{}_{n}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right),$$

where (λ) and (μ) denote the patterns

$$(\lambda) = \begin{pmatrix} \lambda_{1n} & \cdots & \lambda_{nn} \\ \lambda_{1n-1} & \cdots & \lambda_{n-1n-1} \\ \vdots & & \\ \lambda_{12} & \lambda_{22} & \\ \lambda_{11} & & \end{pmatrix},$$

$$(\mu) = \begin{pmatrix} \mu_{1n} & \cdots & \mu_{nn} \\ \mu_{1n-1} & \cdots & \mu_{n-1n-1} \\ \vdots & & \\ \mu_{12} & \mu_{22} & \\ \mu_{11} & & \end{pmatrix}.$$

The numbers $\lambda_{i,j}$ and $\mu_{i,j}$ satisfy the betweenness conditions of Eq. (19). Corresponding to each row in the (λ) (i.e., upper) pattern

$$\lambda_m = (\lambda_{1m}, \lambda_{2m}, \ldots, \lambda_{mm}),$$

we construct the associated Sp(2m) projector $P^{\lambda_m}{}_{m}$ [cf. Eq. (16) for Sp(2n)]. Clearly these subgroup projectors satisfy the rules

$$(P^{\lambda_m}{}_{m})^2 = P^{\lambda_m}{}_{m}, \quad P^{\lambda_m}{}_{m}P^{\lambda_k}{}_{k} = P^{\lambda_k}{}_{k}P^{\lambda_m}{}_{m}\,. \tag{26}$$

We now consider the compound projector

$$P_{(\lambda)} = \prod_{m=1}^{n} P^{\lambda_m}{}_{m}\,. \tag{27}$$

In view of Eq. (26) it is clear that the projection operators (27) obey the rule

$$P_{(\lambda)}P_{(\lambda')} = P_{(\lambda')}P_{(\lambda)}, \quad P^2{}_{(\lambda)} = P_{(\lambda)}\,.$$

By repeated application of Theorem 1 it is easily deduced that the states $P_{(\lambda)}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right)$ form a basis for the irreducible representation $V_0(\lambda)$ which, by our construction, is symmetry adapted to the subgroup chain (14). For ease of notation we denote these symmetry-adapted states by

$$\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0 = P_{(\lambda)}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right). \tag{28}$$

The group theoretical interpretation of the (upper) (λ) pattern is now obvious and refers to the highest weights (or equivalently the eigenvalues of Casimir invariants) of the groups in the subgroup chain (14) in analogy with the GT states for U(n) and O(n). This implies that two symmetry-adapted states (28) are orthogonal unless they have the same upper pattern, i.e.,

$${}_0\left\langle\begin{matrix}(\lambda')\\(\mu')\end{matrix}\middle|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0 = 0, \quad \text{unless } (\lambda') = (\lambda).$$

The (lower) (μ) pattern is a multiplicity label which, in our approach, refers to the representation labels of the groups U(2$n-1$),U(2$n-3$),...,U(3),U(1) of the U(2n) GT vectors (20) from which we are projecting. This dual-pattern labeling may be compared to the pattern calculus of Biedenharn *et al.*,[9,10] developed for the tensor operator problem of U(N). In this latter approach two GT patterns appear, one (which has group theoretical significance) for labeling the components of the tensor operator and the other a multiplicity label (i.e., operator pattern). This dual-pattern idea also appears in the work of Zhelobenko[11] but without any group theoretical significance.

We note that the basis (28) is not symmetry adapted to the subgroup chain (15). However the generators of the subgroup

$$G = \mathrm{Sp}(2)\times\mathrm{Sp}(2)\times\cdots\times\mathrm{Sp}(2) \quad (n \text{ times})$$

of Sp($2n$) have a simple action on the basis states (28). To see this consider the infinitesimal generators of the subgroup G [see Eqs. (6) and (7)]

$$\alpha^{2m-1}{}_{2m} = 2a_{2m-1,2m}, \quad \alpha^{2m}{}_{2m-1} = 2a_{2m,2m-1}, \tag{29}$$

$$h_m = a_{2m-1,2m-1} - a_{2m,2m}, \quad m = 1,\ldots,n.$$

The Cartan generators h_m are diagonal in the basis (28) with eigenvalue given by

$$h_m \left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0 = \left[2\sum_{j=1}^{m}\mu_{j,m} - \sum_{j=1}^{m}\lambda_{j,m} - \sum_{j=1}^{m-1}\lambda_{j,m-1}\right]\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0. \tag{30}$$

The generators (29) of the group G moreover commute with the Casimir invariants of the subgroup chain (14) so that we may write

$$\alpha^{2m-1}{}_{2m}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0 = \alpha^{2m-1}{}_{2m}P_{(\lambda)}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle = P_{(\lambda)}\alpha^{2m-1}{}_{2m}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle.$$

Note that the generators (29) leave the space $A_0(\lambda)$ (to which the GT state $\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle$ belongs) invariant, and moreover the generators $\alpha^{2m-1}{}_{2m}$ (and $\alpha^{2m}{}_{2m-1}$) can only effect the mth row $\mu_m = (\mu_{1m},\ldots,\mu_{mm})$ of the (lower) (μ) pattern. Using the known matrix element formulas of the U($2n$) generators (see, e.g., Gould[22]) we have

$$\alpha^{2m-1}{}_{2m}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle = 2a_{2m-1,2m}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle = 2\sum_{r=1}^{m}N^m{}_r\left|\begin{matrix}(\lambda)\\(\mu)+\Delta^m{}_r\end{matrix}\right\rangle,$$

where $(\mu)+\Delta^m{}_r$ is the pattern obtained from (μ) by the shifts

$$\mu_{k,l}\to\mu_{k,l}, \quad \mu_{r,m}\to\mu_{r,m}+1, \quad \text{for } (k,l)\neq(r,m).$$

From the known matrix elements of the U($2n$) generators we deduce the result (see Appendix B)

$$N^m{}_r = (\mu_{r,m}+m+1-r)^{1/2} \times\left(\frac{(-1)^{m+1}\Pi_{p=1}^{m}(\lambda_{p,m}-\mu_{r,m}+r-p)\Pi_{l=1}^{m-1}(\mu_{r,m}-\lambda_{l,m-1}+l-r+1)}{\Pi_{l=1,\,\neq r}^{m}(\mu_{r,m}-\mu_{l,m}+l-r)(\mu_{r,m}-\mu_{l,m}+l-r-1)}\right)^{1/2}. \tag{31}$$

Thus we obtain the result

$$\alpha^{2m-1}{}_{2m}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0 = 2\sum_{r=1}^{m}N^m{}_r\left|\begin{matrix}(\lambda)\\(\mu)+\Delta^m{}_r\end{matrix}\right\rangle_0,$$

with $N^m{}_r$ as in Eq. (31). Similarly for the lowering generator $\alpha^{2m}{}_{2m-1}$ we obtain

$$\alpha^{2m}{}_{2m-1}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0 = 2\sum_{r=1}^{m}\overline{N}^m{}_r\left|\begin{matrix}(\lambda)\\(\mu)-\Delta^m{}_r\end{matrix}\right\rangle_0,$$

where

$$\overline{N}^m{}_r = (\mu_{r,m}+2m-r-1)^{1/2}\left(\frac{(-1)^{m+1}\Pi_{p=1}^{m}(\lambda_{p,m}-\mu_{r,m}+r-p+1)\Pi_{l=1}^{m-1}(\mu_{r,m}-\lambda_{l,m-1}+l-r)}{\Pi_{l=1,\,\neq r}^{m}(\mu_{r,m}-\mu_{l,m}+l-r)(\mu_{r,m}-\mu_{l,m}+l-r-1)}\right)^{1/2}. \tag{32}$$

Thus using the matrix element formulas (31) and (32) one may obtain, in principle, a basis symmetry adapted to the subgroup chain (15). This can be done using either a lowering operator method or by construction of projection operators for the subgroup chain (15) in analogy with the projection operators of Eq. (27) for the subgroup chain (14). We aim to consider this aspect of the problem in a future publication.

Although the generators (29) of the subgroup G have a simple action on the basis states (28) the action of the remaining generators α_{ij} is not so clear. It suffices, in principle, to determine the action of the elementary generators (8). The matrix elements of the remaining generators can then be obtained using repeated commutation. However, unlike the generators (29), the elementary generators (8) do not commute with the projection operators $P_{(\lambda)}$ so it is necessary to proceed indirectly. From Eq. (5) we may expand the generators $\alpha^{2m-1}{}_{2m+1}$ and $\alpha^{2m+1}{}_{2m-1}$ in terms of U($2n$) generators according to

$$\alpha^{2m-1}{}_{2m+1} = a_{2m-1,2m+1} - a_{2m+2,2m},$$
$$\alpha^{2m+1}{}_{2m-1} = a_{2m+1,2m-1} - a_{2m,2m+2}. \tag{33}$$

Using the known action of the U($2n$) generators on the GT basis states (20) together with the shift properties of the Sp($2n$) generators (33) on the representation labels of the subgroup Sp($2m$), we deduce that the action of the generators (33) on the basis states (28) is of the form

$$\alpha^{2m-1}{}_{2m+1}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0 = \sum_{r,l=1}^{m}N^{m,+}{}_{r,l}\left|\begin{matrix}(\lambda)+\Delta^m{}_r\\(\mu)+\Delta^m{}_r\end{matrix}\right\rangle_0 + \sum_{r=1}^{m}\sum_{l=1}^{m+1}N^{m,-}{}_{r,l}\left|\begin{matrix}(\lambda)-\Delta^m{}_r\\(\mu)-\Delta^{m+1}{}_l\end{matrix}\right\rangle_0, \tag{34}$$

$$\alpha^{2m+1}{}_{2m-1}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0 = \sum_{r,l=1}^{m}\overline{N}^{m,+}{}_{r,l}\left|\begin{matrix}(\lambda)-\Delta^m{}_r\\(\mu)-\Delta^m{}_l\end{matrix}\right\rangle_0 + \sum_{r=1}^{m}\sum_{l=1}^{m+1}\overline{N}^{m,-}{}_{r,l}\left|\begin{matrix}(\lambda)+\Delta^m{}_r\\(\mu)+\Delta^{m+1}{}_l\end{matrix}\right\rangle_0 . \tag{35}$$

We hope to be permitted to evaluate the matrix elements in Eqs. (34) and (35) [which follow directly from the U($2n$) matrix element formulas] in a future publication. Since the basis states (28) are nonorthogonal there still remains the problem of evaluating overlap coefficients, to which we now turn.

V. OVERLAP COEFFICIENTS

The overlap coefficients for the basis states (28) are given by the matrix elements of the projection operator (27) between GT basis states in the space $A_0(\lambda)$, viz.,

$$_0\left\langle\begin{matrix}(\lambda')\\(\mu')\end{matrix}\middle|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0 = \left(\begin{matrix}(\lambda')\\(\mu')\end{matrix}\right| P_{(\lambda)}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right). \tag{36}$$

Since the projection operator $P_{(\lambda)}$ may be expressed as a polynomial in the second-order Casimir invariants for the groups in the chain (14) one may in principle calculate these overlap coefficients using the matrix element formulas for the U($2n$) generators. Whether this yields an effective algorithm for adaption to computers (or ideally analytic manipulation) remains to be seen. Nevertheless we may deduce some elementary properties of the coefficients (36) from general considerations.

We have already noted that the overlap coefficient (36) vanishes unless the (upper) (λ) patterns coincide. Thus we have

$$_0\left\langle\begin{matrix}(\lambda')\\(\mu')\end{matrix}\middle|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0 = \delta_{(\lambda'),(\lambda)}\,{}_0\left\langle\begin{matrix}(\lambda)\\(\mu')\end{matrix}\middle|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0 . \tag{37}$$

Next, since the states (28) are eigenstates of the Cartan generators h_m Eq. (30) implies that the overlap coefficient (37) vanishes unless $N_m(\mu) = N_m(\mu')$ where we define

$$N_m(\mu) = \sum_{i=1}^{m}\mu_{i,m} .$$

Thus we obtain

$$_0\left\langle\begin{matrix}(\lambda)\\(\mu')\end{matrix}\middle|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0 = \delta_{\underline{N}(\mu),\underline{N}(\mu')}\,{}_0\left\langle\begin{matrix}(\lambda)\\(\mu')\end{matrix}\middle|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0 , \tag{38}$$

where $\underline{N}(\mu) = (N_1(\mu), N_2(\mu),\ldots,N_n(\mu))$. Note that the maximal state

$$\left|\begin{matrix}(\lambda)\\(\lambda)\end{matrix}\right\rangle_0 = \left|\begin{matrix}(\lambda)\\(\lambda)\end{matrix}\right)$$

[see Eq. (22)] satisfies

$$_0\left\langle\begin{matrix}(\lambda)\\(\mu')\end{matrix}\middle|\begin{matrix}(\lambda)\\(\lambda)\end{matrix}\right\rangle_0 = \delta_{(\mu'),(\lambda)} . \tag{39}$$

Now in view of Schur's lemma we may write

$$\left(\begin{matrix}(\lambda)\\(\mu')\end{matrix}\right| P^{\lambda_n}{}_n P^{\lambda_{n-1}}{}_{n-1}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right) = \alpha\left(\begin{matrix}(\lambda)_{n-1}\\(\mu')_{n-1}\end{matrix}\right| P^{\lambda_{n-1}}{}_{n-1}\left|\begin{matrix}(\lambda)_{n-1}\\(\mu)_{n-1}\end{matrix}\right),$$

for some constant α, where the patterns $(\lambda)_{n-1}$, $(\mu)_{n-1}$ are obtained from the patterns (λ), (μ), respectively, by omission of the top rows. We call the constant α the reduced Sp($2n$):Sp($2n-2$) overlap coefficient and write it in the form

$$\left\langle\begin{matrix}\lambda_n\\ \lambda_{n-1}\\ \mu_n':\mu_n\end{matrix}\right\rangle \tag{40}$$

to indicate that it depends only on the labels $\lambda_n, \lambda_{n-1}, \mu_{n-1}$, and μ'_{n-1}. Thus we may write

$$_0\left\langle\begin{matrix}(\lambda)\\(\mu')\end{matrix}\middle|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle_0 = \left\langle\begin{matrix}\lambda_n\\ \lambda_{n-1}\\ \mu_n':\mu_n\end{matrix}\right\rangle {}_0\left\langle\begin{matrix}(\lambda)_{n-1}\\(\mu')_{n-1}\end{matrix}\middle|\begin{matrix}(\lambda)_{n-1}\\(\mu)_{n-1}\end{matrix}\right\rangle_0 ,$$

showing that the Sp($2n$) overlap coefficients may be written as an Sp($2n$):Sp($2n-2$) reduced overlap coefficient times an Sp($2n-2$) overlap coefficient. Thus it suffices to evaluate only the reduced overlap coefficients (40) [for each group in the chain (14)]. If we choose the representation labels of the subgroup Sp($2n-2$) to be maximal, Eq. (39) implies

$$\left\langle\begin{matrix}\lambda_n\\ \lambda_{n-1}\\ \mu_n':\mu_n\end{matrix}\right\rangle = \left(\begin{matrix}\lambda_n\\ \lambda_{n-1}\\ (\lambda)_{\max}\\ \mu_n'\\ (\lambda)_{\max}\end{matrix}\right| P^{\lambda_n}{}_n \left|\begin{matrix}\lambda_n\\ \lambda_{n-1}\\ (\lambda)_{\max}\\ \mu_n\\ (\lambda)_{\max}\end{matrix}\right).$$

In other words it suffices to evaluate the matrix elements of the projector $P^{\lambda_n}{}_n$ between the GT states of the space $A_0(\lambda)$ which are maximal in the subgroup Sp($2n-2$) (i.e., semimaximal states). This observation clearly reduces the problem of evaluating the overlap coefficient (37).

From Eqs. (37) and (39) we deduce that the reduced overlap coefficients satisfy

$$\left\langle\begin{matrix}\lambda_n\\ \lambda_{n-1}\\ \mu_n':\mu_n\end{matrix}\right\rangle = 0, \quad \text{unless } N_n(\mu) = N_n(\mu') .$$

Moreover for the maximal reduced overlap coefficients (i.e., $\lambda_{i,n-1} = \lambda_{i,n}, i = 1,\ldots,n-1$ and $\mu_{i,n} = \lambda_{i,n}$, for $i = 1,\ldots,n$) we have

$$\left\langle\begin{matrix}\lambda_n\\ \lambda_n\\ \mu_n':\lambda_n\end{matrix}\right\rangle = \delta_{\mu_n',\lambda_n} .$$

Since the matrix elements of the U($2n$) generators in the GT basis may be chosen real we deduce that the reduced overlap coefficients are real and satisfy

$$\left\langle\begin{matrix}\lambda_n\\ \lambda_{n-1}\\ \mu_n':\mu_n\end{matrix}\right\rangle = \left\langle\begin{matrix}\lambda_n\\ \lambda_{n-1}\\ \mu_n:\mu_n'\end{matrix}\right\rangle .$$

Also, due to the properties of projectors, we must have

$$1 \geqslant \left\langle\begin{matrix}\lambda_n\\ \lambda_{n-1}\\ \mu_n':\mu_n\end{matrix}\right\rangle^2 \geqslant 0 .$$

We do not consider the problem of overlap coefficients any further here. We remark however that there still remain other general features of the labeling scheme we have advocated which may be studied from general considerations. One of these is the property of asymptotic orthogonality where the basis (28) becomes orthogonal in a certain limit of large quantum numbers. This property is satisfied by the solution to the Clebsch–Gordan multiplicity problem (for semisimple Lie algebras) given by Edwards and Gould.[8] Another example of this asymptotic behavior is afforded by Elliot's[5,23] well-known solution to the U(3)⊃O(3) state labeling problem where the projected O(3) states of Elliot rapidly approach orthogonality when the U(3) quantum numbers get sufficiently large. It would be of interest to determine whether such an asymptotic orthogonality is satisfied by the states (28). In terms of reduced overlap coefficients this would require that the coefficients (40) approach unity as the quantum numbers $\lambda_{i,n}, \lambda_{j,n-1}, \mu_{i,n} = \mu'_{i,n}$ become suitably large. This problem has been considered by Biedenharn *et al.*,[9] for the U(N) tensor operator problem, who investigate the behavior of certain coupling coefficients in various limits. It would be of interest to see whether analogous results could be obtained for the reduced overlap coefficients (40).

ACKNOWLEDGMENTS

The authors would like to thank Professor C. A. Hurst for his valuable discussions and for bringing to our attention the work of Zhelobenko.[11]

APPENDIX A

We prove here part (a) of Lemma 1. We adopt the notation of Secs. II and III of the paper. Let $\Omega \in A_0(\lambda)$ be a U($2n$) (allowable) weight state. We prove that there exists a basis monomial $b \in U(B)$ such that $b\Omega = \alpha \Omega^{\lambda}{}_0$, $0 \neq \alpha \in \mathbb{C}$, by induction on n. Since we shall be proceeding down the subgroup chain we adopt the notation of Secs. II and III, except that we add a subscript n to everything to indicate precisely which unitary group we are considering. The notation we adopt is obvious in the present context.

The result holds for $n = 1$ since U(2) is trivial. Proceeding inductively assume the result holds for U($2n-2$); that is, if Ω is any allowable U($2n-2$) GT state of U($2n-2$) weight ν, in the irreducible representation $V(\lambda_{n-1})$ of U($2n-2$), then there exists a basis monomial $b \in U_{n-1}(B)$ such that

$$b\Omega = \alpha \Omega^{\lambda_{n-1}}{}_0, \quad 0 \neq \alpha \in \mathbb{C},$$

where

$$\Omega^{\lambda_{n-1}}{}_0 = \left| \begin{matrix} \lambda_{1,n-1}, \lambda_{2,n-1}, \ldots, \lambda_{n-1,n-1} \\ [\max] \end{matrix} \right) \tag{A1}$$

is the Sp($2n-2$) maximal weight state of Sp($2n-2$) weight $(\lambda_{1,n-1}, \ldots, \lambda_{n-1,n-1})$ [cf. Eq. (22)]. We recall, from the remarks of Sec. III, that $\Omega^{\lambda_{n-1}}{}_0$ is the unique vector in $V(\lambda_{n-1})$ with U($2n-2$) weight

$$\lambda_{0,n-1} = (\lambda_{1,n-1}, 0, \lambda_{2,n-1}, 0, \ldots, \lambda_{n-1,n-1}, 0). \tag{A2}$$

Now let Ω be any allowable U($2n$) basis state of U($2n$) weight ν. Then Ω may be expressed as a sum of Gel'fand states

$$\Omega = \sum_{\mu_n, \lambda_{n-1}, (\tau)} \xi(\mu_n, \lambda_{n-1}, (\tau)) \left| \begin{matrix} \lambda_n \\ \mu_n \\ \lambda_{n-1} \\ (\tau) \end{matrix} \right), \tag{A3}$$

where the sum is over all allowable GT states of weight ν. Choose λ_{n-1} to be maximal (under the lexical ordering) such that $\xi(\mu_n, \lambda_{n-1}, (\tau)) \neq 0$ [for some (τ) and μ_n]. Then for μ_n fixed we see that

$$\Omega' = \sum_{(\tau)} \xi(\mu_n, \lambda_{n-1}, (\tau)) \left| \begin{matrix} \lambda_n \\ \mu_n \\ \lambda_{n-1} \\ (\tau) \end{matrix} \right)$$

is a U($2n-2$) weight state of weight $(\nu_1, \nu_2, \ldots, \nu_{2n-2})$. Hence, by our inductive hypothesis, there exists a basis monomial $b \in U_{n-1}(B)$, of U($2n-2$) weight $\lambda_{0,n-1} - \nu$, such that

$$b\Omega' = \alpha \left| \begin{matrix} \lambda_n \\ \mu_n \\ \lambda_{n-1} \\ (\max) \end{matrix} \right), \quad 0 \neq \alpha \in \mathbb{C}.$$

For other labels λ'_{n-1}, occurring in the sum (A_3), we necessarily have

$$b \left| \begin{matrix} \lambda_n \\ \mu_n \\ \lambda'_{n-1} \\ (\tau) \end{matrix} \right) = 0, \quad \lambda_{n-1} \neq \lambda'_{n-1},$$

since this state has U($2n$) weight $\lambda_{0,n-1}$ [see Eq. (A2)] which is conjugate under the Weyl group to the maximal weight $(\lambda_{n-1}, \dot{0})$ and hence, in view of the maximal nature of λ_{n-1}, cannot occur unless $\lambda'_{n-1} = \lambda_{n-1}$ [recall[19] that the weights in $V(\lambda'_{n-1})$ consist of all integral weights $\nu_{n-1} < \lambda'_{n-1}$ together with their Weyl group conjugates].

For all labels μ'_n occurring in the sum (A3) such that $\xi(\mu'_n, \lambda_{n-1}, (\tau)) \neq 0$, for some (τ), we have, since the state (A1) is the unique U($2n-2$) weight state of weight $\lambda_{0,n-1}$, that

$$b \sum_{(\tau)} \xi(\mu'_n, \lambda_{n-1}, (\tau)) \left| \begin{matrix} \lambda_n \\ \mu'_n \\ \lambda_{n-1} \\ (\tau) \end{matrix} \right) = \alpha' \left| \begin{matrix} \lambda_n \\ \mu'_n \\ \lambda_{n-1} \\ (\max) \end{matrix} \right), \quad \alpha' \in \mathbb{C}.$$

Thus we must have

$$b\Omega = \sum_{\mu_n} \alpha(\mu_n) \left| \begin{matrix} \lambda_n \\ \mu_n \\ \lambda_{n-1} \\ (\max) \end{matrix} \right), \tag{A4}$$

for suitable scalars $\alpha(\mu_n) \in \mathbb{C}$.

Now choose μ_n to be maximal (under the lexical ordering) such that $\alpha(\mu_n) \neq 0$ (by the above such a μ_n exists) and set

$$b' = (a_{2n-3,2n-1})^{\mu_{n-1,n} - \lambda_{n-1,n-1}} \cdots (a_{3,2n-1})^{\mu_{2,n} - \lambda_{2,n-1}} (a_{1,2n-1})^{\mu_{1,n} - \lambda_{1,n-1}}.$$

From the known action of the U($2n$) generators we deduce

$$b'\left|\begin{matrix}\lambda_n\\ \mu_n\\ \lambda_{n-1}\\ (\max)\end{matrix}\right)\neq 0.$$

Moreover this state has weight $(\mu_{1,n},0,\mu_{2,n},\ldots,\mu_{n-1,n},0,\ \mu_{n,n})$ which is conjugate under the Weyl group to the U($2n-1$) maximal weight $(\mu_{1,n},\mu_{2,n},\ldots,\mu_{n,n},\dot{0})$ and thus occurs with unit multiplicity. We thus deduce

$$b'\left|\begin{matrix}\lambda_n\\ \mu_n\\ \lambda_{n-1}\\ (\max)\end{matrix}\right) = \alpha\left|\begin{matrix}\lambda_n\\ \mu_n\\ (\max)\end{matrix}\right),\quad 0\neq\alpha\in\mathbb{C}.$$

For other labels $\mu_n'\neq\mu_n$ occurring in the sum (A4) we deduce, in view of the maximality of μ_n, that

$$b'\left|\begin{matrix}\lambda_n\\ \mu_n'\\ \lambda_{n-1}\\ (\max)\end{matrix}\right) = 0,\quad \text{for } \mu_n'\neq\mu_n .$$

We thus obtain

$$b'b\Omega = \beta\left|\begin{matrix}\lambda_n\\ \mu_n\\ (\max)\end{matrix}\right),\quad 0\neq\beta\in\mathbb{C}.$$

Now set

$$b'' = (a_{2n-1,2n})^{\lambda_{n,n}-\mu_{n,n}} \times (a_{2n-3,2n})^{\lambda_{n-1,n}-\mu_{n-1,n}}\cdots(a_{1,2n})^{\lambda_{1,n}-\mu_{1,n}}.$$

Again, in view of the known action of the U($2n$) generators on GT basis states, we deduce, as before

$$b''\left|\begin{matrix}\lambda_n\\ \mu_n\\ (\max)\end{matrix}\right) = \beta'\left|\begin{matrix}\lambda_n\\ \\ (\max)\end{matrix}\right),\quad \beta'\neq 0.$$

Thus we have

$$b''b'b\Omega = \gamma\left|\begin{matrix}\lambda_n\\ (\max)\end{matrix}\right) = \gamma\Omega^{\lambda}{}_0,\quad 0\neq\gamma\in\mathbb{C}.$$

Thus we have established the result that given a U($2n$) weight vector $\Omega\in A_0(\lambda)$, of weight ν, there exists $b\in U(B)$, of weight $\lambda_0-\nu$ [where λ_0 is the U($2n$) weight of $\Omega^{\lambda}{}_0$] such that

$$b\Omega = \gamma\Omega^{\lambda}{}_0,\quad 0\neq\gamma\in\mathbb{C}.$$

Now b may be expressed as a sum of basis monomials of weight $\lambda_0-\nu$:

$$b = \sum_i \alpha_i b_i,\quad \alpha_i\in\mathbb{C}.$$

Then for some i we must have $b_i\Omega\neq 0$. Since $\Omega^{\lambda}{}_0$ is the unique vector in $V(\lambda)$ of weight λ_0 we must have

$$b_i\Omega = \alpha\Omega^{\lambda}{}_0,\quad 0\neq\alpha\in\mathbb{C}.$$

Our argument is now complete and the result is proved. We remark that, in view of the simplicity of the final result, there probably exists a simpler proof of this result.

APPENDIX B

Let $|\lambda_{ij})$ denote a GT basis state for U($2n$). Then from the known matrix element formulas of the U($2n$) generators (see, e.g., Gould[22]) we have

$$a_{2m-1,2m}|\lambda_{ij}) = \sum_{r=1}^{2m-1} N^{2m-1}{}_r\,|\lambda_{ij}+\Delta^{2m-1}{}_r),$$

where

$$N^{2m-1}{}_r = \left(\frac{(-1)^{2m-1}\Pi_{p=1}^{2m}(\lambda_{p,2m}-\lambda_{r,2m-1}+r-p)\Pi_{l=1}^{2m-2}(\lambda_{r,2m-1}-\lambda_{l,2m-2}+l-r+1)}{\Pi_{l=1}^{2m-1}(\lambda_{r,2m-1}-\lambda_{l,2m-1}+l-r)(\lambda_{r,2m-1}-\lambda_{l,2m-1}+l-r+1)}\right)^{1/2}. \tag{B1}$$

Now for the special representations of U($2n$) we are considering, we have

$$\lambda_{i,2m}=\lambda_{i,2m-1}=0,\quad \text{for } i>m,\quad \lambda_{i,2m-2}=0,\quad \text{for } i>m-1. \tag{B2}$$

For this case only the matrix elements $N^{2m-1}{}_r$ for $r\leqslant m$ are nonzero. Substituting Eq. (B2) into Eq. (B1) we obtain (for $r\leqslant m$)

$$N^{2m-1}{}_r = \left(\frac{(-1)\Pi_{p=1}^{m}(\lambda_{p,2m}-\lambda_{r,2m-1}+r-p)\Pi_{l=1}^{m-1}(\lambda_{r,2m-1}-\lambda_{l,2m-2}+l-r+1)}{\Pi_{l=1,\,\neq r}^{m}(\lambda_{r,2m-1}-\lambda_{l,2m-1}+l-r)(\lambda_{r,2m-1}-\lambda_{l,2m-1}+l-r+1)},\right.$$

$$\left.\frac{\Pi_{p=m+1}^{2m}(r-p-\lambda_{r,2m-1})\Pi_{l=m}^{2m-2}(\lambda_{r,2m-1}+l-r+1)}{\Pi_{l=m+1}^{2m-1}(\lambda_{r,2m-1}+l-r)(\lambda_{r,2m-1}+l-r+1)}\right)^{1/2}$$

$$= \left(\frac{(-1)^{m+1}\Pi_{p=1}^{m}(\lambda_{p,2m}-\lambda_{r,2m-1}+r-p)\Pi_{l=1}^{m-1}(\lambda_{r,2m-1}-\lambda_{l,2m-2}+l-r+1)}{\Pi_{l=1,\,\neq r}^{m}(\lambda_{r,2m-1}-\lambda_{l,2m-1}+l-r)(\lambda_{r,2m-1}-\lambda_{l,2m-1}+l-r+1)}\right)^{1/2}(\lambda_{r,2m-1}+m+1-r)^{1/2}. \tag{B3}$$

In the dual pattern notation of Sec. IV we set $\lambda_{r,m}=\lambda_{r,2m},\mu_{r,m}=\lambda_{r,2m-1}\,(r=1,\ldots,m)$, and $\lambda_{r,m-1}=\lambda_{r,2m-2}\,(r=1,\ldots,m-1)$ and we denote the matrix element (B3) by $N^m{}_r$. Thus we obtain, in the notation of Sec. IV

$$a_{2m-1,2m}\left|\begin{matrix}(\lambda)\\ (\mu)\end{matrix}\right) = \sum_{r=1}^{m} N^m{}_r\left|\begin{matrix}(\lambda)\\ (\mu)+\Delta^m{}_r\end{matrix}\right),$$

where

$$N^{m}{}_{r} = [\mu_{r,m} + m + 1 - r]^{1/2}\left(\frac{(-1)^{m+1}\Pi_{p=1}^{m}(\lambda_{p,m} - \mu_{r,m} + r - p)\Pi_{l=1}^{m-1}(\mu_{r,m} - \lambda_{l,m-1} + l - r + 1)}{\Pi_{\substack{l=1\\ \neq r}}^{m}(\mu_{r,m} - \mu_{l,m} + l - r)(\mu_{r,m} - \mu_{l,m} + l - r + 1)}\right)^{1/2},$$

which gives the matrix element formula of Eq. (31) as required. A similar analysis may be applied to the matrix element formula (32).

[1]I. M. Gel'fand and M. L. Tsetlin, Dokl. Akad. Nauk. SSSR **71**, 825 (1950).
[2]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[3]M. D. Gould, J. Math. Phys. **22**, 2376 (1981).
[4]J. C. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965).
[5]J. P. Elliot, Proc. R. Soc. London Ser. A **245**, 128, 562 (1958); J. P. Elliot and M. Harvey, *ibid.* **272**, 557 (1962).
[6]M. Moshinsky, J. Patera, R. T. Sharp, and P. Winternitz, Ann. Phys. (NY) **95**, 13a (1975); see also B. R. Judd, and W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. **15**, 1787 (1974).
[7]R. M. Asherova and Yu. F. Smirnov, Nucl. Phys. A **144**, 116 (1970).
[8]S. A. Edwards and M. D. Gould, J. Math. Phys. **25**, 2848 (1984).
[9]J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970); **13**, 1985 (1972); M. A. Lohe, L. C. Biedenharn, and J. D. Louck, *ibid.* **18**, 1883 (1977).
[10]L. C. Biedenharn, J. D. Louck, E. Chacón, and M. Ciftan, J. Math. Phys. **13**, 1957 (1972); J. D. Louck and L. C. Biedenharn, *ibid.* **14**, 1336 (1973).
[11]D. P. Zhelobenko, Russ. Math. Surv. **17**, 1 (1962); Transl. Math. Monogr. **40** (1973).
[12]M. A. Lohe and C. A. Hurst, J. Math. Phys. **12**, 1882 (1971).
[13]A. U. Klimyk, J. Math. Phys. A: Math. Gen. **15**, 3009 (1982).
[14]A. M. Bincer, J. Math. Phys. **23**, 347 (1982).
[15]J. Mickelsson, Rep. Math. Phys. **3**, 193 (1972).
[16]P. Pajas and R. Raczka, J. Math. Phys. **9**, 1188 (1968).
[17]E. G. Kalnins and M. D. Gould (to appear).
[18]M. D. Gould, J. Aust. Math. Soc. B **20**, 401 (1978).
[19]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer–Verlag, Berlin, 1972).
[20]M. Hamermesh, *Group Theory and Its Applications* (Addison–Wesley, Reading, MA, 1962).
[21]H. S. Green, J. Math. Phys. **12**, 2106 (1971).
[22]M. D. Gould, J. Math. Phys. **22**, 15 (1981).
[23]J. P. Draayer, D. L. Pursey, and S. A. Williams, Nucl. Phys. A **119**, 577, (1968).

On the denominator function for canonical SU(3) tensor operators

L. C. Biedenharn
Duke University, Durham, North Carolina 27706 and Los Alamos National Laboratory, Theoretical Division, Los Alamos, New Mexico 87545

M. A. Lohe
University of Adelaide, Adelaide, Australia

J. D. Louck
Los Alamos National Laboratory, Theoretical Division, Los Alamos, New Mexico 87545

(Received 21 November 1984; accepted for publication 25 January 1985)

The definition of a canonical unit SU(3) tensor operator is given in terms of its characteristic null space as determined by group-theoretic properties of the intertwining number. This definition is shown to imply the canonical splitting conditions used in earlier work for the explicit and unique (up to $\pm$ phases) construction of all SU(3) WCG coefficients (Wigner–Clebsch–Gordan). Using this construction, an explicit SU(3)-invariant denominator function characterizing completely the canonically defined WCG coefficients is obtained. It is shown that this denominator function (squared) is a product of linear factors which may be obtained explicitly from the characteristic null space times a ratio of polynomials. These polynomials, denoted G_q^t, are defined over three (shift) parameters and three barycentric coordinates. The properties of these polynomials (hence, of the corresponding invariant denominator function) are developed in detail: These include a derivation of their degree, symmetries, and zeros. The symmetries are those induced on the shift parameters and barycentric coordinates by the transformations of a 3×3 array under row interchange, column interchange, and transposition (the group of 72 operations leaving a 3×3 determinant invariant). Remarkably, the zeros of the general G_q^t polynomial are in position and multiplicity exactly those of the SU(3) weight space associated with irreducible representation $[q-1,t-1,0]$. The results obtained are an essential step in the derivation of a fully explicit and comprehensible algebraic expression for all SU(3) WCG coefficients.

I. INTRODUCTION AND REVIEW

Applications of symmetry group techniques in quantal physics depend in an essential way on the algebraic construction, for a given group, of suitable analogs to the Wigner–Clebsch–Gordan coefficients[1] for SU(2) (the quantal angular momentum group). From an operator perspective this is the problem of explicitly constructing all canonical unit tensor operators, a problem that in this detail has been resolved, so far, only for SU(2).

We show in this paper that for the unitary group SU(3)—a symmetry group of fundamental importance in physics—the canonical construction of any unit tensor operator is uniquely defined (to within $\pm$ phase conventions) by an associated invariant function: the denominator function of the tensor operator [see Eq. (2.26) below].

The explicit construction of the denominator function is the necessary first step in the complete construction of all canonical unit tensor operators. In contrast to the construction of the denominator function for SU(2)—which is an elementary product of linear factors[1]—the task of constructing the denominator function for SU(3) is far from elementary and embodies considerable group-theoretic information. This single function encodes the canonical resolution of the multiplicity problem for SU(3), and implies, among other things the Littlewood–Richardson numbers for SU(3) as we shall discuss below.

The purpose of the present paper is the detailed development and study of the denominator function for the symmetry group SU(3). We show that, in the general case, the denominator function is a ratio of two explicitly defined polynomials and develop in detail the many remarkable properties not only of the denominator function, but also of these polynomials [see Eqs. (3.3) and (3.6)] themselves. We demonstrate, in fact, that these polynomials are characterized by surprisingly elegant symmetry properties.

Let us briefly sketch the relevant background for the problem of determining the Wigner–Clebsch–Gordan (WCG) coefficients for the group SU(3), so that we can be more precise about the properties of the denominator function.

The unitary irreps (irreducible representations) of SU(3) are uniquely labeled by two-rowed Young frames corresponding to the partition $[m_{13},m_{23},0]$, with the integers m_{i3} obeying $m_{13}\geqslant m_{23}\geqslant 0$. The orthonormal basis vectors $\left|\begin{smallmatrix}[m]\\(m)\end{smallmatrix}\right\rangle$ spanning each irrep space are uniquely labeled by the Gel'fand–Weyl pattern[2]

$$\begin{pmatrix}[m]\\(m)\end{pmatrix}=\begin{pmatrix}m_{13} & & m_{23} & & m_{33}\\ & m_{12} & & m_{22} & \\ & & m_{11} & & \end{pmatrix}, \tag{1.1}$$

where the integers m_{ij} obey the betweenness conditions

$$\begin{aligned}&m_{13}\geqslant m_{12}\geqslant m_{23}\geqslant m_{22}\geqslant m_{33}=0,\\ &m_{12}\geqslant m_{11}\geqslant m_{22}.\end{aligned} \tag{1.2}$$

The Hilbert space $\mathscr{H}$ on which the tensor operators act is defined to be the direct sum of the vector spaces carrying the unitary irreps of SU(3), each irrep (and hence each Gel'fand–Weyl vector) occurring once and only once.

In order to preserve the underlying symmetry between the labels 1,2,3— and moreover to make this symmetry evident—it is convenient, in defining the space on which the tensor operators act, to allow basis vectors having $m_{33} \geqslant 0$, thus admitting carrier spaces of U(3) irreps labeled by the partition $[m_{13},m_{23},m_{33}]$ (three-rowed Young frames). The Hilbert space $\mathscr{H}$ then consists of a direct sum of such U(3) irrep spaces, each such space occurring exactly once. The inequivalent SU(3) irreps are then obtained by declaring an equivalence relation on U(3) irreps:

$$[m_{13}+k,m_{23}+k,m_{33}+k] \sim [m_{13},m_{23},m_{33}], \tag{1.3}$$

for k a finite integer.

Let us now recall the definition of Ref. 3 of a tensor operator on $\mathscr{H}$ belonging to SU(3) symmetry. A tensor operator is a set of linear operators $O(M)$ indexed by SU(3) Gel'fand–Weyl patterns (M) and obeying the equivariance condition

$$U_g O(M) U_g^{-1} = O(g(M)), \tag{1.4}$$

for every $g \in \mathrm{SU}(3)$, where U_g is the unitary transformation of $\mathscr{H}$ associated with g.

An *irreducible* tensor operator is indexed by patterns belonging to a single SU(3) irrep $[M_{13},M_{23},0]$. The unit tensor operators are not uniquely specified by the SU(3) labels (M) alone. A linear basis for the tensor operators in SU(3) is provided by the *unit* tensor operators (operators with unit norm phased conventionally) having scalar operators as multipliers.[4] The unit tensor operators are continuous, and may be uniquely determined by giving all their matrix elements on the basis $\left|\begin{smallmatrix}[m]\\(m)\end{smallmatrix}\right\rangle$ of $\mathscr{H}$.

It has been shown[2,4] that irreducible unit tensor operators having the same (M) index may be distinguished by a second triangular pattern of integers (also obeying the constraints of a Gel'fand–Weyl pattern) called the operator pattern:

$$\begin{pmatrix}[M]\\(\Gamma)\end{pmatrix} \equiv \begin{pmatrix} M_{13} & & M_{23} & & 0 \\ & \Gamma_{12} & & \Gamma_{22} & \\ & & \Gamma_{11} & & \end{pmatrix}. \tag{1.5}$$

The significance of this operator pattern may be seen in this way: the action of a tensor operator is a transformation of $\mathscr{H}$ into $\mathscr{H}$; for unit tensor operators, the operator label (Γ) specifies the shift $(\Delta_1,\Delta_2,\Delta_3)$ induced by the operator when acting on vectors belonging to the irrep space $\mathscr{H}([m]) \subset \mathscr{H}$. That is,

$$\begin{aligned} O_\Gamma : \mathscr{H}([m]) &= \mathscr{H}([m_{13},m_{23},m_{33}]) \mapsto \mathscr{H}([m+\Delta]) \\ &= \mathscr{H}([m_{13}+\Delta_1,m_{23}+\Delta_2,m_{33}+\Delta_3]), \end{aligned} \tag{1.6}$$

where

$$\begin{aligned} &\Delta_1 = \Gamma_{11}, \quad \Delta_2 = \Gamma_{12}+\Gamma_{22}-\Gamma_{11}, \\ &\Delta_3 = M_{13}+M_{23}+M_{33}-\Gamma_{12}-\Gamma_{22}. \end{aligned} \tag{1.7}$$

The shift labels Δ do not in general distinguish among all unit tensor operators with the same Gel'fand–Weyl label, but the complete operator pattern (Γ) does provide such a distinction. Since the Gel'fand–Weyl pattern and the operator pattern share a common row (the irrep label $[M]$), it is convenient to designate a unique element of the irreducible unit tensor operator by writing the two patterns together by inverting the operator pattern and placing it above the Gel'fand–Weyl pattern, writing the common labels $[M]$ only once:

$$\left\langle \begin{matrix} & & \Gamma_{11} & & \\ & \Gamma_{12} & & \Gamma_{22} & \\ M_{13} & & M_{23} & & M_{33} \\ & M_{12} & & M_{22} & \\ & & M_{11} & & \end{matrix} \right\rangle, \tag{1.8}$$

or more briefly by

$$\left\langle \begin{matrix} (\Gamma) \\ [M] \\ (M) \end{matrix} \right\rangle,$$

in which the common irrep label $[M]$ is set apart from the patterns

$$(M) = \begin{pmatrix} M_{12} & & M_{22} \\ & M_{11} & \end{pmatrix}$$

and

$$(\Gamma) = \begin{pmatrix} & \Gamma_{11} & \\ \Gamma_{12} & & \Gamma_{22} \end{pmatrix}.$$

Let us summarize: A given element, denoted

$$\left\langle \begin{matrix} \cdot \\ [M] \\ \cdot \end{matrix} \right\rangle,$$

of a canonical SU(3) unit tensor operator denoted $\langle [M] \rangle = \langle M_{13} \;\; M_{23} \;\; M_{33} \rangle$ is uniquely labeled by three patterns: (i) a Young frame pattern (partition), $[M] = [M_{13},M_{23},M_{33}]$, which specifies the SU(3) equivalence class irrep carried by the operator; (ii) a Gel'fand–Weyl pattern,

$$\begin{pmatrix}[M]\\(M)\end{pmatrix} \equiv \begin{pmatrix} M_{13} & & M_{23} & & M_{33} \\ & M_{12} & & M_{22} & \\ & & M_{11} & & \end{pmatrix}, \tag{1.1'}$$

which specifies the *vector component* of the operator [irreducible transformation property under the group action (1.4)]; and (iii) an operator pattern,

$$\begin{pmatrix}(\Gamma)\\ [M]\end{pmatrix} \equiv \begin{pmatrix} & & \Gamma_{11} & & \\ & \Gamma_{12} & & \Gamma_{22} & \\ M_{13} & & M_{23} & & M_{33} \end{pmatrix}, \tag{1.5'}$$

which specifies the operator component of the tensor operator (shift action in $\mathscr{H}$).

It is important to note that a tensor operator possesses two distinct weights: (i) an ordinary weight in U(3), which is denoted by W, associated with the Gel'fand-Weyl pattern (1.1'),

$$\begin{aligned} &W = (W_1,W_2,W_3) \text{ with } W_1 = M_{11}, \\ &W_2 = M_{12}+M_{22}-M_{11}, \\ &W_3 = M_{13}+M_{23}+M_{33}-M_{12}-M_{22}; \end{aligned} \tag{1.9}$$

and (ii) a "shift" weight, which is denoted by Δ, associated with the operator pattern (1.5′),

$$\Delta = \Delta\begin{pmatrix}[\Gamma]\\(M)\end{pmatrix} = (\Delta_1,\Delta_2,\Delta_3), \quad (1.10)$$

with component Δ_i defined in analogy with the W_i as given explicitly by Eqs. (1.7).

The two weights, W and Δ, do not completely specify a unique vector component or a unique operator component of the tensor operator $\langle [M]\rangle$; both weights have multiplicity (see below). In general, it requires *both* the operator pattern and Gel'fand–Weyl pattern to specify a unique element of the operator as discussed above. (The canonical unit tensor operator $\langle [M]\rangle$ has $(\mathrm{Dim}\,[M])^2$ elements.)

Let us now give the multiplicity of a given weight. Equivalently, this is the question: How many operator patterns $\binom{[M]}{(\Gamma)}$ are there having given irrep labels $[M_{13},M_{23},M_{33}]$ and given shift pattern $(\Delta_1,\Delta_2,\Delta_3)$? The answer is provided by the following formula for the multiplicity $\mathscr{M}$ (see Ref. 5):

$$\mathscr{M} = \begin{cases}(M_{23} - M_{33} + 1) - (\lambda_1 + \lambda_2 + \lambda_3), \\ \quad \text{for } \Delta_1 + \Delta_2 + \Delta_3 = \Sigma_i M_{i3}, \quad M_{13} \geqslant \Delta_i \geqslant M_{33}, \\ 0, \quad \text{otherwise,}\end{cases} \quad (1.11a)$$

where λ_i is the step function defined by

$$\lambda_i = \max(0, M_{23} - \Delta_i). \quad (1.11b)$$

(This step function will prove very useful in the formulas developed in later sections.)

It is useful to define a notation enumerating the set of operator patterns having the same shift pattern Δ, and having, of course, the same irrep labels $[M]$. We denote elements of this ordered set of operator patterns by (Γ_t), where $t = 1,2,\ldots,\mathscr{M}$ (see Ref. 5). The pattern (Γ_1)—called the "stretched pattern" (by analogy to a similar concept in nuclear spectroscopy)—has the form

$$\begin{pmatrix}(\Gamma_1)\\ [M]\end{pmatrix} \equiv \begin{pmatrix} & & \Delta_1 & & \\ & \Delta_1 + \Delta_2 - M_{33} - \lambda_3 & & M_{33} + \lambda_3 & \\ M_{13} & & M_{23} & & M_{33}\end{pmatrix}. \quad (1.12a)$$

Denoting this stretched pattern by

$$\begin{pmatrix}(\Gamma_1)\\ [M]\end{pmatrix} = \begin{pmatrix} & & \Delta_1 & & \\ & \gamma_{12} & & \gamma_{22} & \\ M_{13} & & M_{23} & & M_{33}\end{pmatrix}, \quad (1.12b)$$

one finds, for the general pattern,

$$\begin{pmatrix}(\Gamma_t)\\ [M]\end{pmatrix} = \begin{pmatrix} & & \Delta_1 & & \\ & \gamma_{12} - t + 1 & & \gamma_{22} + t - 1 & \\ M_{13} & & M_{23} & & M_{33}\end{pmatrix}, \text{ for } t = 1,2,\ldots,\mathscr{M}. \quad (1.12c)$$

The final pattern $(\Gamma_{\mathscr{M}})$ has the form

$$\begin{pmatrix}(\Gamma_{\mathscr{M}})\\ [M]\end{pmatrix} = \begin{pmatrix} & & \Delta_1 & & \\ & \Delta_1 + \Delta_2 + \lambda_1 + \lambda_2 - M_{23} & & M_{23} - \lambda_1 - \lambda_2 & \\ M_{13} & & M_{23} & & M_{33}\end{pmatrix}. \quad (1.12d)$$

Let us consider now the intertwining number function $I_{[M],\Delta}$, which is defined on the set P of all U(3) irrep labels $[m] = [m_{13},m_{23},m_{33}]$. (Thus P denotes the set of all partitions with three parts, including 0.) By definition, the intertwining number $I_{[M],\Delta}([m])$ is the dimensionality of the vector space intertwining the product representation $[M]\times[m]$ with the irrep $[m + \Delta]$. The intertwining number is, accordingly, exactly the same as the so-called Littlewood–Richardson number $g_{[M],[m],[m+\Delta]}$.

The intertwining number function plays a basic role in implementing the definition of a canonical tensor operator. One can see this, in part, from the fact that the number of operator patterns, $\mathscr{M}$, for the operator $\langle [M]\rangle$ with shift Δ, is equal to or greater than the interwining number $I_{[M],\Delta}([m])$,

$$\mathscr{M} \geqslant I_{[M],\Delta}([m]), \quad \text{for all } [m]\in P. \quad (1.13)$$

We show below that equality is achieved for any $[M] = [M_{13},M_{23},0]$ and $\Delta = (\Delta_1,\Delta_2,\Delta_3)$ whenever $m_{13} - m_{23} \geqslant \Delta_2$ and $m_{23} - m_{33} \geqslant \Delta_3$.

Let us regard $[M]$ and Δ as fixed, and seek the level sets P_k of the function $I_{[M],\Delta}$:

$$P_k = \{\,[m]\in P \,|\, I_{[M],\Delta}([m]) = \mathscr{M} - k\,\}, \quad k = 0,1,\ldots,\mathscr{M}. \quad (1.14)$$

The coordinate system most appropriate for graphing these level sets[6] is the barycentric system (x_1,x_2,x_3) with $x_1 + x_2 + x_3 = 0$ (also called the Möbius plane). Here the coordinates for the irrep $[m]$ are given by

$$\begin{aligned} x_1 &= m_{23} - m_{33} + 1, \\ x_2 &= m_{33} - m_{13} - 2, \\ x_3 &= m_{13} - m_{23} + 1. \end{aligned} \quad (1.15)$$

That this coordinate system is the natural one is a consequence of the invariance of the intertwining number $I_{[M],\Delta}([m])$ to "translations"; that is, $I_{[M],\Delta}([m + a]) = I_{[M],\Delta}([m])$ for each $[m]\in P$ and each

$$[m + a] \equiv [m_{13} + a, m_{23} + a, m_{33} + a]\in P.$$

It is useful to develop this use of barycentric coordinates in more detail.

Let us denote the Möbius plane by $\mathbb{M}$ and the subset of *lattice points* (points having integral coordinates) of $\mathbb{M}$ by $\mathbb{L}$. Then the transformation $[m] \mapsto x$ given by Eq. (1.15) is a map ϕ from $P \to \mathbb{L}$; that is, $x = \phi([m])$. We denote by $\mathbb{L}^+$ the image of P under ϕ; that is, $\mathbb{L}^+ = \phi(P)$. Similarly, we define $\mathbb{L}_k$ by $\mathbb{L}_k = \phi(P_k)$. Equivalently, if we define the intertwining number function $\mathscr{I}_{[M],\Delta}$ with domain $\mathbb{L}^+$ by

$$\mathscr{I}_{[M],\Delta}(x) = I_{[M],\Delta}([m]), \quad \text{for } x = \phi([m]), \tag{1.16a}$$

then

$$\mathbb{L}_k = \{x \in \mathbb{L}^+ \mid \mathscr{I}_{[M],\Delta}(x) = \mathscr{M} - k\}, \quad k = 0,1,...,\mathscr{M}. \tag{1.16b}$$

The level sets $\mathbb{L}_k$ have a principal role in the tensor operator problem, as discussed below. Since it is the level sets $\mathbb{L}_k$ that we give in subsequent figures, let us note also that the sets P_k may be recovered from the sets $\mathbb{L}_k$ by

$$P_k = \{[m] \in P \mid \phi([m]) \in \mathbb{L}_k\}. \tag{1.17}$$

It is important to note that the level sets $\mathbb{L}_k$ (equivalently, P_k) are determined *completely* by group-theoretic information, namely, by the set of values of the Littlewood–Richardson numbers.

The role of the level sets $\mathbb{L}_k$, hence, of the intertwining number function in the tensor operator problem, can be best understood from an example. Consider tensor operators $\langle 630 \rangle$ transforming as the irrep $[M] = [630]$ and having the shift pattern $\Delta = (333)$. There are four canonical unit tensor operators, which are denoted in the notation (1.12) by the operator patterns

$$\Gamma_1 = \begin{pmatrix} & 3 & \\ 6 & & 0 \end{pmatrix}, \quad \Gamma_2 = \begin{pmatrix} & 3 & \\ 5 & & 1 \end{pmatrix},$$

$$\Gamma_3 = \begin{pmatrix} & 3 & \\ 4 & & 2 \end{pmatrix}, \quad \Gamma_4 = \begin{pmatrix} & 3 & \\ 3 & & 3 \end{pmatrix}. \tag{1.18}$$

As remarked earlier [Eq. (1.6)], the meaning of these operator patterns is, in part, one of specifying the action of each of the unit tensor operators

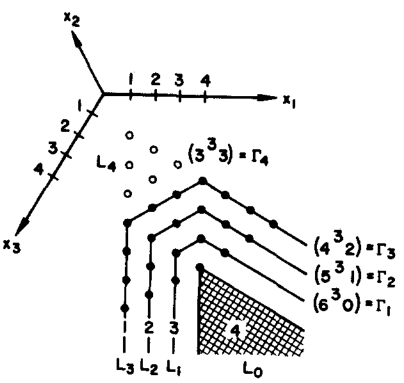


FIG. 1. Graph of the intertwining number function $\mathscr{I}_{[630],(333)}$. The level sets $\mathbb{L}_0, \mathbb{L}_1,...,\mathbb{L}_4$ consist, respectively, of all lattice points $\mathbb{L}_0$ in the cross-hatched region, all lattice points $\mathbb{L}_t$ $(t = 1,2,3)$ on the broken line, and the six lattice points $\mathbb{L}_4$ shown as open circles. The boundary line between the characteristic null space set $\mathbb{N}_t = \mathbb{L}_t \cup \cdots \cup \mathbb{L}_4$ $(t = 1,2,3,4)$ and the non-null region for each operator is labeled by the corresponding operator pattern Γ_t.

$$\left\langle \begin{matrix} & \Gamma_t & \\ 6 & 3 & 0 \\ & \cdot & \end{matrix} \right\rangle, \quad t = 1,2,3,4, \tag{1.19a}$$

on each subspace $\mathscr{H}([m]) \subset \mathscr{H}$,

$$\left\langle \begin{matrix} & \Gamma_t & \\ 6 & 3 & 0 \\ & \cdot & \end{matrix} \right\rangle : \mathscr{H}([m]) \mapsto \mathscr{H}([m] + (3,3,3)), \tag{1.19b}$$

or possibly,

$$\left\langle \begin{matrix} & \Gamma_t & \\ 6 & 3 & 0 \\ & \cdot & \end{matrix} \right\rangle : \mathscr{H}([m]) \mapsto 0. \tag{1.19c}$$

The possibility that some subspaces $\mathscr{H}([m]) \subset \mathscr{H}$ may belong to the null space of a given tensor operator *must* be admitted; indeed, this structural property of a unit tensor operator is *essential* as we next show by considering the level sets $\mathbb{L}_0, \mathbb{L}_1, \mathbb{L}_2, \mathbb{L}_3, \mathbb{L}_4$ described in Fig. 1 and defined by [see Eq. (1.16b)]

$$\mathbb{L}_k = \{x \in \mathbb{L}^+ \mid \mathscr{I}_{[630],(3\,3\,3)}(x) = 4 - k\}. \tag{1.20}$$

A relation between the level sets $\{\mathbb{L}_t\}$ and properties of the unit tensor operators

$$\left\langle \begin{matrix} (\Gamma_t) \\ [M] \\ \cdot \end{matrix} \right\rangle$$

results from the fact[1] that the matrix elements of these unit tensor operators taken between the initial state $\left| \begin{smallmatrix} [m] \\ (m) \end{smallmatrix} \right\rangle$ and the final state $\left| \begin{smallmatrix} [m+\Delta] \\ (m') \end{smallmatrix} \right\rangle$ are WCG coefficients for U(3). For the example at hand, these WCG coefficients are the matrix elements denoted by

$$\left\langle \begin{matrix} m_{13}+3 & m_{23}+3 & m_{33}+3 \\ & \cdot & \end{matrix} \right| \left\langle \begin{matrix} & \Gamma_t & \\ 6 & 3 & 0 \\ & \cdot & \end{matrix} \right\rangle \left| \begin{matrix} m_{13} & m_{23} & m_{33} \\ & \cdot & \end{matrix} \right\rangle, \tag{1.21}$$

which, by definition, are zero should $\mathscr{H}([m])$ belong to the null space of

$$\left\langle \begin{matrix} & \Gamma_t & \\ 6 & 3 & 0 \\ & \cdot & \end{matrix} \right\rangle.$$

In expression (1.21), the symbol $\cdot$ denotes the various Gel'fand–Weyl patterns that enumerate $U(2) \subset U(1)$ subgroup labels. For each specified $[m] \in P$, there is a set of WCG coefficients (1.21) corresponding to each value $t = 1,2,3,4$, and each set, when not all zeros, is a row of a finite, real, proper orthogonal matrix (rows in the real orthogonal matrix that reduces fully the direct product representation $[630] \times [m]$).

The group-theoretic information provided by Fig. 1 is that for all partitions $[m] \in P$ such that $\phi([m]) \in \mathbb{L}_k$ there exist exactly $4 - k$ sets of orthogonal and normalizable (nonzero sets) of WCG coefficients. Thus, for all $[m]$ such that $\phi([m]) \in \mathbb{L}_0$, all four sets, $t = 1,2,3,4$ of WCG coefficients (1.21) exist and are orthogonal and normalized; however, for all $[m] \in P$ such that $\phi([m]) \in \mathbb{L}_1$, only three such sets of orthogonal and normalized WCG coefficients exist. The fourth set, which existed and was orthogonal and normalizable for

$\phi([m]) \in \mathbb{L}_0$, must consist of all zeros for $\phi([m]) \in \mathbb{L}_1$. Thus, all subspaces $\mathscr{H}([m]) \subset \mathscr{H}$ with $[m] \in P$ such that $\phi([m]) \in \mathbb{L}_1$ *must* belong to the null space of one of the unit tensor operators (1.19a). Continuing in this way, we obtain the following group-theoretic information from Fig. 1: All subspaces $\mathscr{H}([m]) \subset \mathscr{H}$ such that $[m] \in P$ and $\phi([m]) \in \mathbb{L}_k$ must belong to the null space of exactly k of the unit tensor operators (1.19a), each $k = 1,2,3,4$.

We can now state what we mean by a canonical set of unit tensor operators (1.19a). Let $\mathscr{N}_k$ denote the vector space $\mathscr{N}_k \subset \mathscr{H}$ defined by

$$\mathscr{N}_k = \sum_{[m]} \oplus \mathscr{H}([m]), \tag{1.22a}$$

where the summation is over all $[m] \in P$ such that

$$\phi([m]) \in \mathrm{N}_k \equiv \mathbb{L}_k \cup \cdots \cup \mathbb{L}_4, \tag{1.22b}$$

each $k = 1,2,3,4$. We call the set of all unit tensor operators

$$\left\langle \begin{matrix} & \Gamma_t & \\ 6 & 3 & 0 \\ & \cdot & \end{matrix} \right\rangle \quad (t = 1,2,3,4)$$

canonical if and only if

$$\left\langle \begin{matrix} & \Gamma_t & \\ 6 & 3 & 0 \\ & \cdot & \end{matrix} \right\rangle : \mathscr{N}_t \to \mathbf{0}. \tag{1.23}$$

The set inclusion property

$$\mathrm{N}_1 \supset \mathrm{N}_2 \supset \mathrm{N}_3 \supset \mathrm{N}_4 \tag{1.24a}$$

and, correspondingly [since $\phi(\mathscr{N}_k) = \mathrm{N}_k$],

$$\mathscr{N}_1 \supset \mathscr{N}_2 \supset \mathscr{N}_3 \supset \mathscr{N}_4, \tag{1.24b}$$

then imply that the conditions (discussed above) imposed on the WCG coefficients by the structure of the level sets $\mathbb{L}_k$ in Fig. 1 are satisfied. The association $\Gamma_t \to \mathscr{N}_t$ between operator labels and vector spaces is discussed further below.

Let us note especially that the vector spaces $\mathscr{N}_t$ are direct sums of whole irrep spaces contained in $\mathscr{H}$ and other individual vectors in $\mathscr{H}$, not in $\mathscr{N}_t$, may possibly be in the null space of a canonical unit tensor operator. For this reason, we refer to $\mathscr{N}_t$ as the characteristic null space of the unit tensor operator

$$\left\langle \begin{matrix} & \Gamma_t & \\ 6 & 3 & 0 \\ & \cdot & \end{matrix} \right\rangle.$$

The results described above for the special case $[M] = [630], \Delta = (333)$ generalize: The unit tensor operators

$$\left\langle \begin{matrix} (\Gamma_1) \\ [M] \\ \cdot \end{matrix} \right\rangle, \left\langle \begin{matrix} (\Gamma_2) \\ [M] \\ \cdot \end{matrix} \right\rangle, \ldots, \left\langle \begin{matrix} (\Gamma_{\mathscr{M}}) \\ [M] \\ \cdot \end{matrix} \right\rangle, \tag{1.25a}$$

with the same shift pattern

$$\Delta(\Gamma_1) = \Delta(\Gamma_2) = \cdots = \Delta(\Gamma_{\mathscr{M}}) = \Delta = (\Delta_1, \Delta_2, \Delta_3),$$

are distinguished by the nested property

$$\mathscr{N}_1 \supset \mathscr{N}_2 \supset \cdots \supset \mathscr{N}_{\mathscr{M}} \tag{1.25b}$$

of their corresponding characteristic null spaces. In this general case, the vector space $\mathscr{N}_t$ is defined by

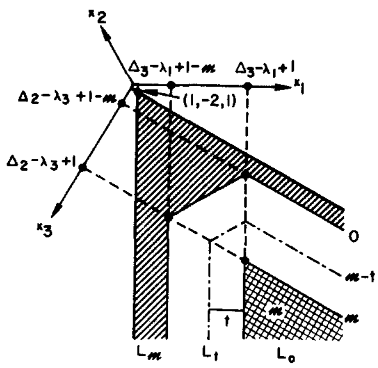


FIG. 2. Graph of the intertwining number function $\mathscr{I}_{[M],\Delta}$. The level sets $\mathbb{L}_0, \mathbb{L}_1, \ldots, \mathbb{L}_{\mathscr{M}}$ consist, respectively, of all lattice points $\mathbb{L}_0$ in the cross-hatched region, all lattice points $\mathbb{L}_t$ $(t = 1,2,\ldots,\mathscr{M}-1)$ on the broken line, and all lattice points $\mathbb{L}_{\mathscr{M}}$ in the shaded region, the value of $\mathscr{I}$ being $\mathscr{M}, \mathscr{M}-t$ $(t = 1,\ldots,\mathscr{M}-1)$, 0. In general $\Delta_3 - \lambda_1 + 1 - \mathscr{M} > 0$ and $\Delta_2 - \lambda_3 + 1 - \mathscr{M} > 0$; in case that $\Delta_3 - \lambda_1 + 1 - \mathscr{M} = 0$ (resp. $\Delta_2 - \lambda_3 + 1 - \mathscr{M} = 0$), the x_1-boundary line (resp. x_3-boundary line) is $x_1 = 1$ (resp. $x_3 = 1$).

$$\mathscr{N}_t = \sum_{[m]} \oplus \mathscr{H}([m]), \tag{1.25c}$$

where the summation is over all partitions $[m] \in P$ such that

$$\phi([m]) \in \mathrm{N}_t \equiv \mathbb{L}_t \cup \cdots \cup \mathbb{L}_{\mathscr{M}}, \tag{1.25d}$$

where $\mathbb{L}_k$ is the level set defined by Eq. (1.16b). These level sets $\mathbb{L}_t$ are given explicitly for the general case in Fig. 2 and the set N_t is shown in Fig. 3.

We can now give the definition of a canonical unit tensor operator for the general case: The set of unit tensor operators

$$\left\langle \begin{matrix} \Gamma_t \\ [M] \\ \cdot \end{matrix} \right\rangle \text{ with } \Delta \begin{pmatrix} [M] \\ (\Gamma_t) \end{pmatrix} = (\Delta_1, \Delta_2, \Delta_3), \quad t = 1,\ldots,\mathscr{M} \tag{1.26a}$$

is canonical if and only if

$$\left\langle \begin{matrix} (\Gamma_t) \\ [M] \\ \cdot \end{matrix} \right\rangle : \mathscr{N}_t \to \mathbf{0}. \tag{1.26b}$$

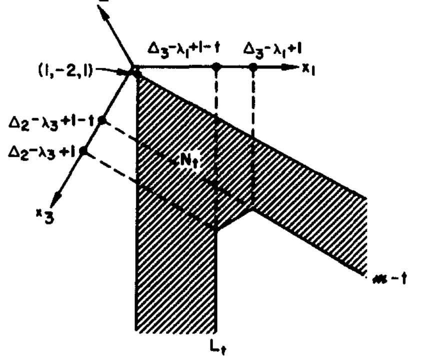


FIG. 3. The general null space set N_t. The set N_t consists of all lattice points in the lexical region $\mathbb{L}^+$ $(x_1 \geqslant 1, x_2 \leqslant -2, x_3 \geqslant 1)$ with boundary consisting of the broken line $\mathbb{L}_t$. The points in this set, which includes the boundary, correspond to all irrep label $[m]$ such that $\phi([m]) \in \mathrm{N}_t$, hence, $\mathscr{H}([m]) \subset \mathscr{N}_t$.

The characteristic null space $\mathscr{N}_t$ is defined fully by Eqs. (1.25) and Fig. 2; equivalently, it is defined by

$$\phi(\mathscr{N}_t) = \mathrm{N}_t. \tag{1.26c}$$

[An explicit definition of $\mathscr{N}_t$ without the use of figures is given in Eqs. (1.33) below.]

The preceding definition of a canonical unit tensor operator is quite abstract and contains no hint as to how the matrix elements of such an operator are to be calculated. We have given earlier (see Refs. 5–8, particularly, Ref. 5) such a calculational tool, which we now review briefly so as to present (in Theorems 1.1 and 1.2 below) the relationship of that method to the definition of a canonical unit tensor operator given here and based directly on the characteristic null space concept.

In our earlier work,[5] we have given an explicit procedure for calculating all SU(3) WCG coefficients based on conditions called the "canonical splitting conditions." These conditions are that certain SU(3):U(2) projective operators must be *zero operators*. [The matrix elements of the general projective operator are often called isoscalar factors or SU(3):U(2) reduced matrix elements.] The requirement of these zero operators in a canonical definition of the WCG coefficients was based on rather subtle properties of the pattern calculus.[5,6,8] Here we derive these conditions directly from the definition of a canonical unit tensor operator given above.

A unit U(3):U(2) projective operator is labeled by the U(3) operator pattern of the U(3) unit tensor operator

$$\left\langle \begin{matrix} & & (\Gamma_t) & & \\ \mathrm{M}_{13} & & \mathrm{M}_{23} & & \mathrm{M}_{33} \\ & \mathrm{M}_{12} & & \mathrm{M}_{22} & \\ & & \mathrm{M}_{11} & & \end{matrix} \right\rangle \tag{1.27a}$$

with which it is associated, and any one of the U(2) operator patterns associated with the U(2) unit tensor operators

$$\left\langle \begin{matrix} & \Gamma'_{11} & \\ M_{12} & & M_{22} \\ & M_{11} & \end{matrix} \right\rangle, \quad \Gamma'_{11} = M_{22}, M_{22}+1,\ldots,M_{12}. \tag{1.27b}$$

The explicit notation is

$$\begin{bmatrix} (\Gamma_t) \\ [M] \\ (\Gamma') \end{bmatrix}, (\Gamma') = \begin{pmatrix} M_{12} & & M_{22} \\ & \Gamma'_{11} & \end{pmatrix}. \tag{1.27c}$$

It follows then that each of the patterns

$$\begin{pmatrix} (\Gamma_t) \\ [M] \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} [M] \\ (\Gamma') \end{pmatrix}$$

satisfies the betweenness constraints of a Gel'fand–Weyl pattern. Moreover, the significance of these patterns is exactly that inherited from its associated unit tensor operator. In particular, there are two shift patterns associated with the unit U(3):U(2) projective operator (1.27c): the U(3) shift pattern,

$$\Delta = \Delta \begin{pmatrix} [M] \\ (\Gamma_t) \end{pmatrix} = (\Delta_1, \Delta_2, \Delta_3) \tag{1.28a}$$

as defined in Eqs. (1.7); and the U(2) shift pattern,

$$\begin{aligned} &\Delta' = \Delta(\Gamma') = (\Delta'_1, \Delta'_2) \quad \text{with} \\ &\Delta'_1 = \Gamma'_{11}, \Delta'_2 = M_{12} + M_{22} - \Gamma'_{11}. \end{aligned} \tag{1.28b}$$

The action of the unit U(3):U(2) projective operator in $\mathscr{H}$ is also inherited from its parent U(3) unit tensor operator and its daughter U(2) unit tensor operator. However, such an operator is a U(1) invariant and has no action on the U(1) label m_{11} of the U(3)$\supset$U(2)$\supset$U(1) basis vectors of $\mathscr{H}$. It is convenient therefore to introduce the basis vector

$$\left| \begin{matrix} m_{13} & & m_{23} & & m_{33} \\ & m_{12} & & m_{22} & \end{matrix} \right\rangle, \tag{1.29a}$$

which denotes the set of basis vectors in $\mathscr{H}$ given by

$$\left\{ \left| \begin{matrix} m_{13} & & m_{23} & & m_{33} \\ & m_{12} & & m_{22} & \\ & & m_{11} & & \end{matrix} \right\rangle \middle| m_{11} = m_{22},\ldots,m_{12} \right\}. \tag{1.29b}$$

Accordingly, the symbol (1.29a) denotes an equivalence class of vectors in $\mathscr{H}$.

Let us next give the action of the projective operator (1.27c) on the generic basis vector (1.29a): If

$$\begin{bmatrix} (\Gamma_t) \\ [M] \\ (\Gamma') \end{bmatrix} \neq 0, \tag{1.30a}$$

then this unit projective operator effects the following shift action in $\mathscr{H}$:

$$\begin{aligned} &\begin{bmatrix} (\Gamma_t) \\ [M] \\ (\Gamma') \end{bmatrix} : \left| \begin{matrix} m_{13} & m_{23} & m_{33} \\ m_{12} & m_{22} & \end{matrix} \right\rangle \\ &\quad \to (\#) \left| \begin{matrix} m_{13}+\Delta_1 & m_{23}+\Delta_2 & m_{23}+\Delta_3 \\ m_{12}+\Delta'_1 & m_{22}+\Delta'_2 & \end{matrix} \right\rangle, \end{aligned} \tag{1.30b}$$

where # is defined by

$$\begin{aligned} \# = (\mathrm{Dim}[m_{12}+\Delta'_1, m_{22}+\Delta'_2])^{-1} \sum_{m'_{11}, m_{11}, M_{11}} &\left\langle \begin{matrix} m_{13}+\Delta_1 & m_{23}+\Delta_2 & m_{33}+\Delta_3 \\ m_{12}+\Delta'_1 & m_{22}+\Delta'_2 & \\ & m'_{11} & \end{matrix} \right| \left\langle \begin{matrix} & (\Gamma_t) & \\ M_{13} & M_{23} & M_{33} \\ M_{12} & M_{22} & \\ & M_{11} & \end{matrix} \right\rangle \left| \begin{matrix} m_{13} & m_{23} & m_{33} \\ m_{12} & m_{22} & \\ & m_{11} & \end{matrix} \right\rangle \\ \times &\left\langle \begin{matrix} m_{12}+\Delta'_1 & m_{22}+\Delta'_2 \\ m'_{11} & \end{matrix} \right| \left\langle \begin{matrix} & \Gamma'_{11} & \\ M_{12} & & M_{22} \\ & M_{11} & \end{matrix} \right\rangle \left| \begin{matrix} m_{12} & m_{22} \\ m_{11} & \end{matrix} \right\rangle, \end{aligned} \tag{1.30c}$$

where the summation is over all m'_{11}, m_{11}, M_{11} corresponding to lexical Gel'fand–Weyl patterns [see Ref. 1, Vol. 8, Eq. (3.227)].

Our reason for giving relations (1.30) in this detail is that we can immediately conclude the following.

Lemma 1.1: If

$$\begin{bmatrix} (\Gamma_t) \\ [M] \\ (\Gamma') \end{bmatrix} \neq \mathbf{0},$$

then the characteristic null space of this U(3):U(2) projective and the canonical unit tensor operator

$$\left\langle \begin{matrix} (\Gamma_t) \\ [M] \\ \cdot \end{matrix} \right\rangle$$

coincide; that is, the null space of the projective operator is also $\mathscr{N}_t$.

We can now prove two principal results—Theorems 1.1 and 1.2.

Theorem 1.1: Suppose there exists a subset $\mathbb{K}_t$ of the lexical region $\mathbb{L}^+$ of the Möbius plane with the following property: For each lexical U(3):U(2) Gel'fand–Weyl pattern

$$\begin{pmatrix} m_{13} & & m_{23} & & m_{33} \\ & m_{12} & & m_{22} & \end{pmatrix} \text{ such that } \phi([m]) \in \mathbb{K}_t, \tag{1.31a}$$

the pattern

$$\begin{pmatrix} m_{13}+\Delta_1 & & m_{23}+\Delta_2 & & m_{33}+\Delta_3 \\ & m_{12}+\Delta'_1 & & m_{22}+\Delta'_2 & \end{pmatrix} \tag{1.31b}$$

is nonlexical (violates betweenness), where this pattern is the one associated with the shift action in $\mathscr{H}$ of the U(3):U(2) unit projective operator

$$\begin{bmatrix} (\Gamma_t) \\ [M] \\ (\Gamma') \end{bmatrix}. \tag{1.31c}$$

Then, necessarily,

$$\mathbb{K}_t \subseteq \mathbb{N}_t. \tag{1.31d}$$

Proof: Since, by assumption, the pattern (1.31b) is nonlexical and the pattern (1.31a) is lexical, it is necessary that $\mathscr{H}([m]) \in \mathscr{N}_t$ for all $[m]$ such that $\phi([m]) \in \mathbb{K}_t$. But $\mathscr{H}([m]) \in \mathscr{N}_t$ implies that $\phi([m]) \in \mathbb{N}_t$. ■

Using Theorem 1.1, we can now prove a second principal result, which relates the definition of a canonical unit tensor operator given here in terms of characteristic null space to the canonical "splitting conditions" discovered much earlier.[5–8] [It is convenient now to put $M_{13} = p$, $M_{23} = q$, $M_{33} = 0$, and $(\Gamma') = \left(\begin{smallmatrix} \alpha & & \beta \\ & \gamma & \end{smallmatrix}\right)$ for comparison with our earlier results.]

Theorem 1.2: Let $t \in \{2,3,\ldots,\mathscr{M}\}$. If the entries in the lexical pattern

$$\begin{pmatrix} p & & q & & 0 \\ & \alpha & & \beta & \\ & & \gamma & & \end{pmatrix}$$

satisfy either

$$\alpha \geqslant \gamma \geqslant \beta + (p - t + 2) \tag{1.32a}$$

or

$$\alpha - (p - t + 2) \geqslant \gamma \geqslant \beta, \tag{1.32b}$$

then the following U(3):U(2) projective operator is the zero operator:

$$\begin{bmatrix} (\Gamma_t) \\ \begin{matrix} p & & q & & 0 \\ & \alpha & & \beta & \\ & & \gamma & & \end{matrix} \end{bmatrix} = \mathbf{0} \tag{1.32c}$$

Proof: Let us note first that $t = 1$ is not included in the above theorem because relations (1.32a) and (1.32b) are not satisfied for any lexical pattern.

Before giving the proof itself, let us give explicitly the irrep labels $[m] \in P$ such that $\phi([m]) \in \mathbb{N}_t$; that is, such that $\mathscr{H}([m]) \in \mathscr{N}_t$. These irrep labels are those satisfying any of the following sets of conditions (obtained from Fig. 3):

(i) $m_{13} = m_{23} + k, \quad k = 0,1,\ldots,\Delta_2 - \lambda_3 - t;$ (1.33a)

(ii) $m_{23} = m_{33} + k, \quad k = 0,1,\ldots,\Delta_3 - \lambda_1 - t;$ (1.33b)

(iii) $m_{13} = m_{33} + k,$

$$k = \Delta_2 + \Delta_3 - \lambda_1 - \lambda_3 - 2t + 2, \ldots, \Delta_2 + \Delta_3 - \lambda_1 - \lambda_3 - t, \tag{1.33c}$$

$$m_{23} \leqslant m_{33} + \Delta_3 - \lambda_1 - 1,$$

$$m_{13} \leqslant m_{23} + \Delta_2 - \lambda_3 - 1,$$

where conditions (iii) are empty for $t = 1$.

Let us next define the set $\mathscr{K}_t$ by

$$\mathscr{K}_t = \left\{ [m] \in P \;\middle|\; \begin{matrix} m_{13} = m_{23} + \Delta_2 - \lambda_3 - t + 1 \quad \text{or} \\ m_{23} = m_{33} + \Delta_3 - \lambda_1 - t + 1 \end{matrix} \right\} \tag{1.34a}$$

and $\mathbb{K}_t$ by

$$\mathbb{K}_t = \phi(\mathscr{K}_t). \tag{1.34b}$$

Note that $\mathbb{K}_t$ is a subset of $\mathbb{N}_{t-1}$, but *not* of $\mathbb{N}_t$.

We prove the theorem by showing that the pattern

$$\begin{pmatrix} m_{13}+\Delta_1 & & m_{23}+\Delta_2 & & m_{33}+\Delta_3 \\ & m_{12}+\gamma & & m_{22}+\alpha+\beta-\gamma & \end{pmatrix} \tag{1.35a}$$

is nonlexical for all lexical patterns

$$\begin{pmatrix} p & & q & & 0 \\ & \alpha & & \beta & \\ & & \gamma & & \end{pmatrix} \text{ and } \begin{pmatrix} m_{13} & & m_{23} & & m_{33} \\ & m_{12} & & m_{22} & \end{pmatrix} \tag{1.35b}$$

such that conditions (1.32a) or (1.32b) are fulfilled and $[m] \in \mathscr{K}_t$. The desired proof then follows from Theorem 1.1, since $\mathbb{K}_t$ is not contained in $\mathbb{N}_t$.

Consider first that $m_{13} = m_{23} + \Delta_2 - \lambda_3 - t + 1$ so that $m_{13} + \Delta_1 = m_{23} + p + q - \Delta_3 - \lambda_3 - t + 1$, where we have used $\Delta_1 + \Delta_2 + \Delta_3 = p + q$. Since the minimum value of $m_{12} + \gamma$ is $(m_{12} + \gamma)_{\min} = m_{23} + \beta + p - t + 2$, we find that $m_{13} + \Delta_1 < (m_{12} + \gamma)_{\min}$ for all β such that $\beta \geqslant q - \Delta_3 - \lambda_3$. Since $\beta \geqslant 0$ and $q - \Delta_3 - \lambda_3 \leqslant 0$ [see Eq. (1.11a) with $M_{23} = q$], this last inequality is always satisfied. Thus, the entry $m_{13} + \Delta_1$ in the final pattern is always less than the smallest entry that can occur for $m_{12} + \gamma$. The pattern is therefore nonlexical for all $[m_{13}, m_{23}, m_{33}]$ having $m_{13} = m_{23} + \Delta_2 - \lambda_3 - t + 1$. The proof for $m_{23} = m_{33} + \Delta_3 - \lambda_1 - t + 1$ is given similarly. ■

Remarks: (a) The essential structural ingredient underlying the occurrence of zero projective operators may be summarized by the following statement: Certain U(3):U(2) unit projective operators must be the zero operator in order that the null space of all nonzero unit projective operators agrees exactly with the null space of the parent unit tensor operator.

(b) We have given in our previous work[8] an algorithm for calculating all WCG coefficients based on the occurrence of the zero projective operators given in Theorem 1.2. This construction is unique, up to $\pm$ phase choices. Thus, the definition of a canonical unit tensor operator given here, and based on the characteristic null space, gives a unique structural resolution of the multiplicity problem implied by group-theoretic information alone except possibly for the assignment of the operator patterns (Γ_t) themselves. We discuss this further below.

In our review of the properties of tensor operators, we have discussed the concept of a operator pattern, and we have mentioned how a specific operator pattern (Γ_t) in a given multiplicity set is to be associated with a specific operator according to the characteristic null space $\mathcal{N}_t$ of that operator. Whereas the Δ pattern associated to a given operator (and operator pattern) is clearly related to an intrinsic property of the operator (the shift induced by the operator), this association of a specific operator pattern (Γ_t)—even though well defined and unique—nonetheless seems to involve some arbitrariness. To put the matter in different words: Why could we not have associated the patterns (Γ_t) in some other ordering? It is a remarkable fact that this ordering itself is uniquely induced from the limiting properties of the matrix elements of a canonical unit U(3):U(2) projective operator:

$$\lim_{m_{22}\to-\infty}\lim_{m_{33}\to-\infty}\left\langle \begin{matrix} m_{13}+\Delta_1 & m_{23}+\Delta_2 & m_{33}+\Delta_3 \\ m_{12}+\Delta_1' & m_{22}+\Delta_2' & \end{matrix} \right| \left[\begin{matrix} \Gamma_{11} \\ \Gamma_{12}\quad\Gamma_{22} \\ M_{13}\quad M_{23}\quad M_{33} \\ M_{12}\quad M_{22} \\ \Gamma_{11}' \end{matrix} \right] \left| \begin{matrix} m_{13} & m_{23} & m_{33} \\ m_{12} & m_{22} & \end{matrix} \right\rangle = \delta_{\Gamma_{12},M_{12}}\delta_{\Gamma_{22},M_{22}}. \quad (1.36)$$

[Because the matrix elements in this result depend only on the differences $m_{13}-m_{23}$, $m_{23}-m_{33}$, $m_{13}-m_{22}$, $m_{12}-m_{22}$, $m_{i3}-m_{j2}$ $(i=1,2,3; j=1,2)$, the limits are well defined.] We have discussed earlier[6] how this limiting property [which implies that a U(3) WCG coefficient limits similarly to zero or a U(2) WCG coefficient] is used to assign the operator patterns (Γ_t). Since this result is not used directly here, we assume it for the purpose of assigning definite operator patterns. The important result we wish to note in this review is that this association of a given operator pattern (Γ_t) to a specific operator (defined by its null space) is fully as intrinsic, and unique, as the assignment of the shift labels Δ.

We can now assert the sense in which the word "canonical" is applied to this construction of tensor operators.

A canonical set of unit tensor operators (hence, Wigner–Clebsch–Gordan coefficients) is a set of unit tensor operators that is labeled by operator patterns and associated with the characteristic null spaces as described above. Thus, not only do the operator patterns enumerate all irreducible tensor operators, but they do so in a way that gives precisely and uniquely a basis for all linear transformations between *any* two irrep spaces; equivalently, this solves the "Wigner–Clebsch–Gordan coefficient problem" for SU(3) in a structurally meaningful way with no free choices (aside from $\pm$ phase conventions).

A global, coordinate-free algebraic, presentation of these concepts has recently been obtained.[9,10] It has been shown that the set of all tensor operators for SU(3) can be identified as a simple algebra—with no (nontrivial) two-sided ideal—which is a quotient algebra of the enveloping algebra of so_8. The Hilbert space on which these operators act (precisely the same $\mathscr{H}$ discussed above) has been shown to be a carrier space of a simple irreducible unitary representation of the Lie algebra so(6,2)—a noncompact real form of so_8 having Cartan index -4.

We are now in a position to be more precise about the denominator function. To every operator component of the tensor operator $\langle [M] \rangle$, there is associated a denominator function; since the association is to be invariant under SU(3), the function is denoted by D_Γ; that is, only the operator pattern enters.

As shown below, the function D_Γ is, in fact, a *norm* for the set of operator components having a specified Δ pattern; the values of D_Γ are accordingly non-negative real numbers. The domain of the function D_Γ is the set of irrep labels $[m]\in P$ [the labels of the U(3) invariant vector spaces $\mathscr{H}([m])\subset\mathscr{H}$]. Again, because of the translational invariance of the intertwining function, only the barycentric coordinates (1.15) enter.

The most important property of the denominator function is that it vanishes on the characteristic null space associated with the operator pattern Γ. Since the characteristic null spaces are uniquely associated with operator patterns that define unique elements of a given tensor operator (splitting all multiplicities), the denominator function is an invariant structural definition of the operator itself.

The matrix elements of a canonical unit U(3):U(2) projective operator have the following symbolic form:

$$\frac{(\text{NPCF})^{1/2}(\text{numerator polynomial})}{[\text{U(3)-invariant polynomial}]^{1/2}[\text{U(2)-invariant polynomial}]^{1/2}}, \quad (1.37)$$

where NPCF denotes a (known) numerator pattern calculus factor, which is obtained from some general (and very useful) rules known as the pattern calculus.[8,11] NPCF is the same for all operators in a multiplicity set (independent of t). The numerator polynomial may be written as a polynomial in the U(2) variables m_{12}, m_{22} with coefficients that are polynomials in the m_{i3} $(i=1,2,3)$, that is, U(3)-invariant coefficients. It is convenient to define the denominator function D_Γ such that the U(3)-invariant coefficient of the highest power, $m_{12}^a m_{22}^b$, in the numerator polynomial is unity. This gives a denominator function that is a ratio of U(3)-invariant polynomials. After factoring out certain linear factors, which themselves are identified from the characteristic null space, we are led to a denominator function of the (symbolic) form:

$$D^2_{\Gamma_t} = \text{(ratio of Dim factors) (linear factors)} G_q^{t-1}/G_q^t, \tag{1.38}$$

where each G_q^t is a polynomial ($G_q^0 = 1$) [see Eq. (3.3)].

The G_q^t functions have remarkable symmetry properties, which are developed in detail in Sec. V. Probably the most striking of these symmetries is that the zeros of the G_q^t functions occur in SU(3) weight space patterns with the multiply occupied weight space points corresponding precisely to multiple zeros! This reemergence of unitary symmetry was, to us, both unexpected and surprising.

Let us now outline the plan of the paper.

In Sec. II, we give the derivation of the denominator function by an explicit construction of all projective operators of maximal U(2) shift in a canonical set (Γ_t) [Eq. (2.5a)]. The method exploits the canonical splitting conditions to implement, uniquely, an *ordered* Gram–Schmidt process.

The denominator functions determined in this way are, inherently, very complicated objects indeed. It is therefore essential to develop further properties that help one understand the nature of this object. One such property is the expression of the denominator function as a ratio of polynomials. This leads to the G_q^t function (a polynomial), which is defined in Sec. III. The basic structural form of the denominator function is given in Eq. (3.3).

We study the G_q^t function in Sec. IV and develop a general reduction formula, Eq. (4.15). In Sec. V, we prove that the G_q^t function has remarkable symmetry properties (Theorem 5.1). The fact that G_q^t is a polynomial is demonstrated in Sec. VI. Further properties of G_q^t are discussed in Sec. VII. The zeros of the G_q^t function are determined in Sec. VIII and our most striking result—that the zeros belong to weight space patterns—is demonstrated (Theorems 8.1–8.3). Section IX contains our conclusions.

The proofs of the various properties of the denominator function, and of the G_q^t function, are necessarily rather long, detailed, and, inevitably, arduous to work out. Because of this length, we have chosen to put in a second paper a complete reformulation of the G_q^t function as an explicit multiple series, defining a new type of special function (this aspect of the problem has been developed and examined by Milne and collaborators).[12] We obtained this reformulation as a conjecture when the results proved in this present paper were first announced at the 1975 Nijmegen Conference.[13] A detailed elaboration of the symmetry properties of the G_q^t functions (contained in the present paper) proved essential to the proof that the conjectured function actually *is* identical to G_q^t, a task that eluded our efforts at completion for almost a decade.

II. DETERMINANTAL FORM FOR THE DENOMINATOR FUNCTION

The purpose of the present section is to sketch the derivation of the denominator $D^2_{\Gamma_t}$ characterizing the SU(3) Wigner operator

$$\left\langle \begin{matrix} & \Gamma_t & \\ p & q & 0 \\ & \cdot & \end{matrix} \right\rangle, \quad t = 1,2,\ldots,\mathscr{M}. \tag{2.1}$$

We give the function $D^2_{\Gamma_t}$ explicitly in Eqs. (2.22), (2.26), and (2.27) below. Since our primary objective here is to develop the properties of these functions, one may for the purpose of the present investigation take these equations to be the definition of the function $D^2_{\Gamma_t}$. It is important, however, to point out how these results relate to and extend our earlier work on the canonical SU(3) tensor operators (see, in particular, Refs. 5, 6, 8, and 14).

The denominator function $D^2_{\Gamma_t}$ occurs in the characterization of the SU(3):U(2) projective function denoted by

$$\begin{bmatrix} & \Gamma_t & \\ p & q & 0 \\ & \cdot & \end{bmatrix}, \tag{2.2}$$

and may be obtained by considering the following product of multiplicity-free projective operators:

$$\begin{aligned} &\begin{bmatrix} & \Gamma' & \\ p-q & 0 & 0 \\ & p-q \quad 0 & \\ & p-q & \end{bmatrix} \begin{bmatrix} & \Gamma'' & \\ q & q & 0 \\ & q \quad q & \\ & q & \end{bmatrix} \\ &\quad = \sum_{t=1}^{\mathscr{M}} \left\{ \begin{pmatrix} p & q & 0 \\ & \Gamma_t & \end{pmatrix} \begin{pmatrix} & \max & \\ p-q & 0 & 0 \\ & \Gamma' & \end{pmatrix} \begin{pmatrix} q & q & 0 \\ & \Gamma'' & \end{pmatrix} \right\} \\ &\quad \times \begin{bmatrix} & \Gamma_t & \\ p & q & 0 \\ & p \quad q & \\ & p & \end{bmatrix}. \end{aligned} \tag{2.3}$$

In this expression the bracket symbol $\{\cdots\}$ denotes an invariant operator, which will also be obtained in the present work, but whose detailed properties will not be required here. The symbols Γ', Γ'', and Γ_t denote operator patterns of the respective projective operators. Here Γ' and Γ'' are multiplicity-free, since there is a one-to-one correspondence between the operator patterns Γ', Γ'' and the corresponding shift patterns

$$[\Delta'] = [\Delta_1', \Delta_2', \Delta_3'] \quad \text{and} \quad [\Delta''] = [\Delta_1'', \Delta_2'', \Delta_3''].$$

In general the shift pattern for the projective operator $[p\ q\ 0]$ is not multiplicity-free; that is, for prescribed Γ' and Γ'', the shift pattern

$$[\Delta] = [\Delta_1, \Delta_2, \Delta_3] = [\Delta_1' + \Delta_1'', \Delta_2' + \Delta_2'', \Delta_3' + \Delta_3'']$$

does not uniquely determine the operator pattern for $[p\ q\ 0]$ but rather a set of distinct patterns $\Gamma_1, \Gamma_2, \ldots, \Gamma_{\mathscr{M}}$. The summation on the right-hand side of Eq. (2.3) extends over this set. The explicit assignment of the operator patterns in the order given corresponds to maximal, ..., minimal characteristic null space as has been discussed in detail in Refs. 5, 6, and 8 and reviewed briefly in the Introduction.

The operator identity (2.3) becomes an explicit relation between matrix elements of projective operators when it acts on the SU(3)$\supset$U(2) basis vector characterized by the irrep labels $[m]=[m_{13},m_{23},m_{33}]$, $[m']=[m_{12},m_{22}]$. (The final SU(3)$\supset$U(2) irrep labels are then $[m_{13}+\Delta_1, m_{23}+\Delta_2, m_{33}+\Delta_3]$ and $[m_{12}+q, m_{22}+q]$.) The results for the individual projective operators on the left-hand side of Eq. (2.3) are given by [see Eqs. (3.50) and (3.55) of Ref. 8]

$$\begin{bmatrix} & \Gamma'' & \\ q & q & 0 \\ & q \quad q & \\ & q & \end{bmatrix}\begin{pmatrix}[m]\\ [m']\end{pmatrix} = (-1)^{q-\Delta''_2}(\mathrm{NPCF})^{1/2} \times\left[D\begin{pmatrix}[\Delta''_1\Delta''_2\Delta''_3]\\ [q\,q\,0]\end{pmatrix}([m])\right]^{-1}, \tag{2.4a}$$

$$\begin{bmatrix} & \Gamma'' & \\ p-q & 0 & 0 \\ & p-q \quad 0 & \\ & p-q & \end{bmatrix}\begin{pmatrix}[m]\\ [m']\end{pmatrix} = (-1)^{p-q-\Delta'_1}(\mathrm{NPCF})^{1/2} \times\left[D_2\begin{pmatrix}[p-q\,0]\\ [p-q\,0]\end{pmatrix}([m'])D\begin{pmatrix}[\Delta'_1\Delta'_2\Delta'_3]\\ [p-q\,0\,0]\end{pmatrix}([m])\right]^{-1}. \tag{2.4b}$$

In these expressions NPCF denotes the numerator pattern calculus factor for the operator in question. (The rules for evaluating these factors for a prescribed operator have been described in Refs. 8 and 11.) The denominator factors occurring in these expressions are also given in Ref. 8 (in terms of the same notation) and are used in deriving Eq. (2.9a) below.

The *form* of the projective operator appearing on the right-hand side of relation (2.3) may be shown to be

$$\begin{bmatrix} & \Gamma_t & \\ p & q & 0 \\ & p \quad q & \\ & p & \end{bmatrix}\begin{pmatrix}[m]\\ [m']\end{pmatrix} = (\mathrm{NPCF})^{1/2}\mathscr{P}_t(m_{22},[m]) \times\left[D_2\begin{pmatrix}[p\,q]\\ [p\,q]\end{pmatrix}([m'])D\begin{pmatrix} & \Gamma_t & \\ p & q & 0\end{pmatrix}([m])\right]^{-1}, \tag{2.5a}$$

where $\mathscr{P}_t(m_{22},[m])$ is a polynomial of (total) degree $\mathscr{M}-t$ in $m_{13}-m_{22}, m_{23}-m_{22}, m_{33}-m_{22}$ with coefficients that are rational polynomials in the differences $m_{13}-m_{23}, m_{23}-m_{33}, m_{33}-m_{13}$, and where

$$D^2_{\Gamma_t}([m]) = D^2\begin{pmatrix} & \Gamma_t & \\ p & q & 0\end{pmatrix}([m]) \tag{2.5b}$$

is the unknown denominator function that we seek to determine. For this purpose the only property of $\mathscr{P}_t$ we need initially is the fact that the coefficient of the leading term, $m_{22}^{\mathscr{M}-t}$ is unity—this is actually a convention we make in order that expression (2.5a) fully defines the denominator function (2.5b) of interest here.

In order to identify the denominator function $D^2_{\Gamma_t}$ using Eq. (2.3)–(2.5), we point out the following features involved in the calculation.

(i) The U(2) denominator function in Eq. (2.5a) is related to the SU(2) denominator function in Eq. (2.4b) by

$$D_2\begin{pmatrix}[p\,q]\\ [p\,q]\end{pmatrix}([m']) = D_2\begin{pmatrix}[p-q\,0]\\ [p-q\,0]\end{pmatrix}([m']+[q\,q]). \tag{2.6a}$$

(ii) The numerator pattern calculus factors appearing in Eqs. (2.4) combine to yield the numerator pattern calculus factor in Eq. (2.5a) times linear factors arising from opposing arrows in the following two arrow patterns:

$\Delta'_1 \quad \Delta'_2 \quad \Delta'_3 \qquad \Delta''_1 \quad \Delta''_2 \quad \Delta''_3$

$p-q \quad 0 \qquad\qquad q \quad q$

$$\rightarrow(-1)^{\mu_3}\prod_{i=1}^{3}(p_{i3}+\Delta''_i-p_{22}-q)_{\mu_i}, \tag{2.6b}$$

where

$$p_{i3}=m_{i3}+3-i,\quad p_{j2}=m_{j2}+2-j, \tag{2.6c}$$

$$\mu_i=\max(\Delta'_i, q-\Delta''_i)\quad \text{with}\quad \Delta_i=\Delta'_i+\Delta''_i, \tag{2.6d}$$

and $(x)_a$ denotes a rising factorial [see the list of symbols given in Eqs. (2.8)].

Using the results given by Eqs. (2.4)–(2.6) in Eq. (2.3) in its matrix element version, we obtain the following relation:

$$(-1)^{p-\Delta'_1-\Delta''_2+\mu_3}\prod_{i=1}^{3}(p_{i3}+\Delta'_i-p_{22}-q)_{\mu_i}\left\{\left[\mathrm{D}\begin{pmatrix}[\Delta'_1\,\Delta'_2\,\Delta'_3]\\ [p-q\,0\,0]\end{pmatrix}\otimes D\begin{pmatrix}[\Delta''_1\,\Delta''_2\,\Delta''_3]\\ [q\,q\,0]\end{pmatrix}\right]([m])\right\}^{-1}$$

$$=\sum_{t=1}^{\mathscr{M}}\left\{\begin{pmatrix}p & q & 0\\ & \Gamma_t & \end{pmatrix}\begin{pmatrix} & \max & \\ p-q & 0 & 0\\ & \Gamma' & \end{pmatrix}\begin{pmatrix}q & q & 0\\ & \Gamma'' & \end{pmatrix}\right\}([m]+[\Delta])\times\mathscr{P}_t(m_{22},[m])\left[D\begin{pmatrix} & \Gamma_t & \\ p & q & 0\end{pmatrix}([m])\right]^{-1}. \tag{2.7}$$

Before extracting the denominator function from Eq. (2.7), it is convenient to define the following notations.

(i) q is an integer that may assume values $0,1,2,\ldots$.

(ii) p is an integer that may assume for each q the values $q, q+1,\ldots$.

(iii) Δ is any 3-tuple of nonnegative integers $(\Delta_1,\Delta_2,\Delta_3)$ such that for specified q and p

$$0\leqslant\Delta_i\leqslant p\quad\text{and}\quad \Delta_1+\Delta_2+\Delta_3=p+q. \tag{2.8a}$$

(iv) λ is the 3-tuple of integers $(\lambda_1,\lambda_2,\lambda_3)$ such that for specified q and Δ_i each λ_i is defined by

$$\lambda_i=\max(0,\ q-\Delta_i). \tag{2.8b}$$

(v) $\mathscr{M}$ is defined by

$$\mathscr{M}=q+1-(\lambda_1+\lambda_2+\lambda_3) \tag{2.8c}$$

and is the multiplicity of the shift pattern Δ in irrep $[p\,q\,0]$.

(vi) n is the 3-tuple of integers (n_1,n_2,n_3) with domain $\mathbb{D}$ defined for specified p,q,Δ by

$$\mathbb{D}(p,q,\Delta) = \left\{(n_1,n_2,n_3)\middle| \begin{matrix} n_i\in\{\lambda_i,\lambda_i+1,\ldots,\sigma_i\} \\ n_1+n_2+n_3=q \end{matrix}\right\}, \quad (2.8d)$$

where σ_i is defined by

$$\sigma_i = \min(q,p-\Delta_i).$$

(vii) $n-\lambda$ is the 3-tuple of integers $(n_1-\lambda_1,n_2-\lambda_2,n_3-\lambda_3)=(\mu_1,\mu_2,\mu_3)$ with domain $\mathbb{D}'$ defined for specified p,q,Δ by

$$\mathbb{D}'(p,q,\Delta) = \left\{(\mu_1,\mu_2,\mu_3)\middle| \begin{matrix} \mu_i\in\{0,1,\ldots,\sigma_i-\lambda_i\} \\ \mu_1+\mu_2+\mu_3=\mathscr{M}-1 \end{matrix}\right\}. \quad (2.8e)$$

(viii) $\mathbb{M}\supset\mathbb{L}\supset\mathbb{L}^+$ denote, respectively, the Möbius plane, the set of lattice points of $\mathbb{M}$ (subset of $\mathbb{M}$ with integral coordinates), and the subset of lattice points such that $x_1\geqslant 1, x_2\leqslant -2, x_3\geqslant 1$ [the points corresponding to irreps $[m]$ of U(3)]. Here $x=(x_1,x_2,x_3)$ denotes a point $x\in\mathbb{M}$, or $x\in\mathbb{L}$, or $x\in\mathbb{L}^+$, as appropriate.

(ix) $(z)_a = z(z+1)\cdots(z+a-1)$ for each $a=0,1,2,\ldots$ and indeterminate z is Pochhammer's notation for a rising factorial; $[z]_a = z(z-1)\cdots(z-a+1)$ is a falling factorial.

(x) $n_i = q-\Delta_i'' = q-\Delta_i+\Delta_i'$ is an alternative way of expressing the components of n (see the remark below).

(xi) the dimension of irreps $[m]$ and $[m^f]$, where $m^f_{i3} = m_{i3}+\Delta_i$, are denoted by

$$\begin{aligned} \mathrm{Dim}(x) &= -x_1x_2x_3/2, \\ \mathrm{Dim}(x^f) &= -x_1^f x_2^f x_3^f/2, \end{aligned} \quad (2.8f)$$

where

$$x_i^f = x_i+\Delta_j-\Delta_k, \quad (ijk) \text{ cyclic.} \quad (2.8g)$$

(xii) for each 3-tuple of nonnegative integers $\alpha=(\alpha_1,\alpha_2,\alpha_3)$ and each point $x\in\mathbb{M}$, the function $L(\alpha;x)$ is defined by

$$L(\alpha;x) = \prod_{ijk}\alpha_i!(x_i+1)_{\alpha_j}(-x_i+1)_{\alpha_k}, \quad (2.8h)$$

where the symbol $\Pi_{ijk}A_{ijk}$ denotes the cyclic product

$$\prod_{ijk}A_{ijk} = A_{123}A_{231}A_{312}; \quad (2.8i)$$

it is also useful to note that

$$\mathrm{Dim}(x)L(\alpha;x) = \frac{1}{2}\prod_{ijk}(-1)^{\alpha_i+1}\alpha_i!(x_i-\alpha_k)_{\alpha_j+\alpha_k+1}. \quad (2.8j)$$

Remarks: (a) One finds that for specified $(\Delta_1,\Delta_2,\Delta_3)$ satisfying Eq. (2.8a) and all $(\Delta_1',\Delta_2',\Delta_3')$ satisfying $\Delta_1'+\Delta_2'+\Delta_3' = p-q$ and $0\leqslant\Delta_i'\leqslant p-q$, that the domain of definition of the 3-tuple $n=(n_1,n_2,n_3)$ is that given by Eq. (2.8d). These considerations are important in enumerating the components of certain vectors (m-tuples) that appear below.

(b) The product $L(\alpha;x)$ of linear factors occurs repeatedly in the subsequent developments; its introduction here serves to simplify that presentation.

Let us now express the left-hand side of Eq. (2.7) in terms of the notations introduced above. Here the first relevant result is the denominator function, which may be written concisely in terms of the function $L(\alpha;x)$ defined in Eqs. (2.8h) and (2.8i) above:

$$\left(\left[D\begin{pmatrix}[\Delta_1' & \Delta_2' & \Delta_3'] \\ [p-q & 0 & 0]\end{pmatrix}\otimes D\begin{pmatrix}[\Delta_1'' & \Delta_2'' & \Delta_3''] \\ [q & q & 0]\end{pmatrix}\right]([m])\right)^2$$
$$= D_n^2(\Delta;x) = \frac{(-1)^{\Delta_2-q}\,\mathrm{Dim}(x)}{(p-q)!q!\,\mathrm{Dim}(x^f)}L(n';x')L(n;-x), \quad (2.9a)$$

where n' and x' are given in terms of n and x by

$$n_i' = n_i+\Delta_i-q, \quad i=1,2,3 \quad (2.9b)$$

$$x_i' = x_i-n_j+n_k, \quad (ijk)\text{cyclic.} \quad (2.9c)$$

In Eqs. (2.9), we regard $\Delta=(\Delta_1,\Delta_2,\Delta_3)$ as a specified shift pattern for an operator with irrep labels $[p\ q\ 0]$. Accordingly, (n_1,n_2,n_3) has the domain $\mathbb{D}(p,q,\Delta)$ as discussed above. The phase factor $(-1)^{\Delta_2-q}$ in Eq. (2.9a) is required in order to ensure that $D_n^2(\Delta;x)\geqslant 0$ for $x\in\mathbb{L}^+$.

Remark: The factor Dim (x^f) in $D_n^2(\Delta;x)$ divides the numerator; that is, $D_n^2(\Delta;x)$ is a polynomial in (x_1,x_2,x_3). Its expression is somewhat simpler if we do not effect this cancellation.

It is also useful for the subsequent discussion to write the denominator $D_n^2(\Delta;x)$ in the following form:

$$\frac{1}{D_n^2(\Delta;x)} = (-1)^{q+1-\Delta_2}\frac{\mathrm{Dim}(x^f)}{[\mathrm{Dim}(x)]^2}\,\frac{(q!)^4(p-q)!N_n(\Delta;x)}{L(\Delta;x)L(qqq;x)}, \quad (2.10a)$$

where N_n is defined by

$$N_n(\Delta;x) = \frac{(x_1-n_2+n_3)(x_2-n_3+n_1)(x_3-n_1+n_2)}{2n_1!n_2!n_3!}$$
$$\times\prod_{ijk}(-\Delta_i)_{q-n_i}(x_i-\Delta_k)_{q-n_j}$$
$$\times(-x_i-\Delta_j)_{q-n_k}(x_i-q)_{q-n_j}(-x_i-q)_{q-n_k}. \quad (2.10b)$$

This result is obtained by straightforward cancellations of common factors in $L(\Delta;x)\,L(qqq;x)/L(n';x')\,L(n;-x)$ [see Eqs. (2.8h), (2.8j), and (2.9)]. The term $L(\Delta;x)\;L(qqq;x)$ thus contains all common factors in the set of terms $L(n';x')L(n;-x)$, $n\in\mathbb{D}(p,q,\Delta)$; it is not, however, *necessarily* the least product of common factors.

The second step we take in simplifying the left-hand side of Eq. (2.7) is to introduce the variable y defined by

$$y = p_{22}-\tfrac{1}{3}(p_{13}+p_{23}+p_{33}). \quad (2.11)$$

Observing that the μ_i defined in Eq. (2.6d) are related to the n_i and λ_i defined in Eqs. (2.8d) and (2.8b), respectively, by

$$\mu_i = n_i-\lambda_i, \quad (2.12)$$

we find that the rising factorial terms in the left-hand side of Eq. (2.7) assume the following form:

$$(-1)^{\mu_1+\mu_2+\mu_3}\prod_{i=1}^{3}(p_{i3}+\Delta_i'-p_{22}-q)_{\mu_i}$$
$$= \prod_{ijk}(y+1+\tfrac{1}{3}(x_j-x_k)+\lambda_i)_{n_i-\lambda_i}$$
$$= \sum_{t=1}^{\mathscr{M}} y^{\mathscr{M}-t}F_n^t(\lambda;x), \quad (2.13)$$

where the expansion of the rising factorial into a polynomial in y defines fully the polynomials denoted by $F^t_n(\lambda;x)$ for each $t = 1,2,\ldots,\mathscr{M}$. In obtaining the right-hand side of relation (2.13), we have used [see Eqs. (2.8b), (2.8c), and (2.12)]

$$\mu_1 + \mu_2 + \mu_3 + 1 = \mathscr{M}. \tag{2.14}$$

We next use the expansion (2.13) and the denominator in the form of Eq. (2.10) to rewrite the left-hand side of Eq. (2.7) as

$$(-1)^{\Phi}\sum_{t=1}^{\mathscr{M}} y^{\mathscr{M}-t} V^t_n(\Delta,\lambda;x), \tag{2.15a}$$

where we have defined

$$V^t_n(\Delta,\lambda;x) \equiv F^t_n(\lambda;x)/D_n(\Delta;x). \tag{2.15b}$$

The phase factor Φ in Eq. (2.15a) is obtained from Eqs. (2.7), (2.9a), and (2.12)–(2.14) to be

$$\Phi = \Delta_2 + \Delta_3 + \lambda_3 + \mathscr{M} - 1. \tag{2.15c}$$

We next introduce the following notations for the quantities occurring in the right-hand side of Eq. (2.7), where we anticipate in the new notation that the various quantities depend, in fact, on only the variables Δ, λ, n, x, y, and t; this result is justified by inspection of the explicit equations for each of the quantities in terms of factors arising from the left-hand side of Eq. (2.7) [see Eqs. (2.23)–(2.25) below]:

$$D(\Gamma_t;x) = D_{\Gamma_t}([m]) = D\begin{pmatrix} & \Gamma_t & \\ p & q & 0\end{pmatrix}([m]), \tag{2.16a}$$

$$P_t(x,y) = \mathscr{P}_t(m_{22},[m]), \tag{2.16b}$$

$$R^t_n(\Delta,\lambda;x) = (-1)^{\Phi}\left\{\begin{pmatrix} p & q & 0 \\ & \Gamma_t & \end{pmatrix}\begin{pmatrix} & \max & \\ p-q & 0 & 0 \\ & \Gamma' & \end{pmatrix}\begin{pmatrix} q & q & 0 \\ & \Gamma'' & \end{pmatrix}\right\}([m]+[\Delta]). \tag{2.16c}$$

Using these notations, Eq. (2.7) becomes

$$\prod_{ijk}\frac{(y+1+\frac{1}{3}(x_j-x_k)+\lambda_i)_{n_i-\lambda_i}}{D_n(\Delta;x)} = \sum_{t=1}^{\mathscr{M}} y^{\mathscr{M}-t}V^t_n(\Delta,\lambda;x) = \sum_{t=1}^{\mathscr{M}} R^t_n(\Delta,\lambda;x)\frac{P_t(x,y)}{D(\Gamma_t;x)}. \tag{2.17}$$

As we shall see, this set of relations (2.17) (one for each triple n in its domain) plays a primary role in the subsequent developments.

In order to solve Eqs. (2.17) for the various quantities on the right-hand side, it is convenient to introduce the vectors (m-tuples for some finite integer m)

$$\mathbf{V}^t(\Delta,\lambda;x) \quad\text{and}\quad \mathbf{R}^t(\Delta,\lambda;x), \tag{2.18}$$

with components enumerated by $n \in \mathbb{D}(p,q,\Delta)$ [see Eq. (2.8d)]. Equation (2.17) then assumes the vector form

$$\sum_{t=1}^{\mathscr{M}} y^{\mathscr{M}-t}\mathbf{V}^t(\Delta,\lambda;x) = \sum_{t=1}^{\mathscr{M}}\mathbf{R}^t(\Delta,\lambda;x)\frac{P_t(x,y)}{D(\Gamma_t;x)}. \tag{2.19}$$

Let us next define the scalar product of arbitrary (real) vectors $\mathbf{X}$ and $\mathbf{Y}$ with components $\{X_n\}$ and $\{Y_n\}$ in the usual way:

$$(\mathbf{X},\mathbf{Y}) = \sum_n X_n Y_n.$$

In terms of this notation, the vectors $\mathbf{R}^t(\Delta,\lambda;x)$ [with components that are Racah coefficients—see Eq. (2.16c)] are orthonormal:

$$(\mathbf{R}^s(\Delta,\lambda;x),\mathbf{R}^t(\Delta,\lambda;x)) = \delta_{st}. \tag{2.20}$$

Using this fact, one can see that the equating of equal powers of the variable y in each side of relation (2.19) is *exactly* the process of expressing the set of linearly independent vectors (for x in region $\mathbf{L}_0$, see Fig. 2)

$$\mathbf{V}^1(\Delta,\lambda;x),\mathbf{V}^2(\Delta,\lambda;x),\ldots,\mathbf{V}^{\mathscr{M}}(\Delta,\lambda;x) \tag{2.21a}$$

in terms of the orthonormal vectors

$$\mathbf{R}^1(\Delta,\lambda;x),\mathbf{R}^2(\Delta,\lambda;x),\ldots,\mathbf{R}^{\mathscr{M}}(\Delta,\lambda;x) \tag{2.21b}$$

by applying the Gram–Schmidt orthonormalization process to the set of vectors (2.21a) *in the order displayed* (first $\mathbf{V}^1$, then $\mathbf{V}^2$,...).

Before implementing the preceding observation to obtain the explicit results given below, we emphasize that the use of the Gram–Schmidt procedure is merely a *device* for solving Eqs. (2.19) in explicit form—the solution itself is unique, including all phases [fixed by relation (2.19) itself]; the equations themselves originate uniquely (up to phase conventions) from the canonical split of the multiplicity, as reviewed briefly in the Introduction. The Gram–Schmidt procedure as used here is not *ad hoc*.

The fact that Eqs. (2.19) are uniquely solved by application on the Gram–Schmidt procedure to the ordered set of vectors (2.21a) allows us to obtain an explicit expression for each of the quantities on the right-hand side of Eq. (2.19). In order to express these results in a more concise form, we will sometimes suppress in certain symbols their dependence on Δ, λ, and x. Thus, we write

$$V^t_n = V^t_n(\Delta,\lambda;x), \tag{2.22a}$$

$$(\mathbf{V}^s,\mathbf{V}^t) = (\mathbf{V}^s(\Delta,\lambda;x),\mathbf{V}^t(\Delta,\lambda;x)) = \sum_n \frac{F^s_n(\lambda;x)F^t_n(\lambda;x)}{D^2_n(\Delta;x)}. \tag{2.22b}$$

The Gram determinant of the first t vectors in the set (2.21a) is similarly denoted A_t:

$$A_t = A_t(\Delta,\lambda;x) = \det\begin{bmatrix}(\mathbf{V}^1,\mathbf{V}^1) & \cdots & (\mathbf{V}^1,\mathbf{V}^t)\\ \vdots & & \vdots \\ (\mathbf{V}^t,\mathbf{V}^1) & \cdots & (\mathbf{V}^t,\mathbf{V}^t)\end{bmatrix}, \tag{2.22c}$$

with $A_0 = 1$.

Using these abbreviated notations, we may express the results obtained from the Gram–Schmidt process described above by the following equations:

$$R^t_n = R^t_n(\Delta\lambda;x) = \det\begin{bmatrix} (\mathbf{V}^1,\mathbf{V}^1) & \cdots & (\mathbf{V}^1,\mathbf{V}^{t-1}) & V^1_n \\ \vdots & & \vdots & \vdots \\ (\mathbf{V}^t,\mathbf{V}^1) & \cdots & (\mathbf{V}^t,\mathbf{V}^{t-1}) & V^t_n \end{bmatrix}(A_{t-1}A_t)^{-1/2}, \tag{2.23}$$

$$D(\Gamma_t;x) = (A_{t-1}/A_t)^{1/2}, \tag{2.24}$$

$$P_t(x,y) = \frac{1}{A_t}\det\begin{bmatrix} (\mathbf{V}^1,\mathbf{V}^1) & \cdots & (\mathbf{V}^1,\mathbf{V}^{t-1}) & Q_1(x,y) \\ \vdots & & \vdots & \vdots \\ (\mathbf{V}^t,\mathbf{V}^1) & \cdots & (\mathbf{V}^t,\mathbf{V}^{t-1}) & Q_t(x,y) \end{bmatrix}, \tag{2.25a}$$

where

$$Q_t(x,y) = \sum_{s=t}^{\mathscr{M}} y^{\mathscr{M}-s}(\mathbf{V}^s,\mathbf{V}^t), \quad \text{each } t = 1,\ldots,\mathscr{M}. \tag{2.25b}$$

Remark: If $\mathbf{X}$ is a vector written as a sum $\mathbf{X} = \Sigma^m_{t=1} a_t\mathbf{X}_t$ of m linearly independent vectors $\mathbf{X}_1, \ldots, \mathbf{X}_m$, then the Gram–Schmidt procedure applied to this ordered set of vectors yields the expression for $\mathbf{X}$ in terms of the ordered orthonormal vectors $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ as

$$\mathbf{X} = \sum_t b_t\mathbf{Y}_t, \quad \text{with } b_t = \sum_{s=t}^{m} a_s(\mathbf{X}_s,\mathbf{Y}_t).$$

Observe that if $\mathbf{X}$ is a prescribed vector, then the freedom of reversing the sign of any $\mathbf{Y}_t$ is reflected by a reversal of the sign of b_t. In the application above, we have the correspondences ($m = \mathscr{M}$)

$$a_t \rightarrow y^{\mathscr{M}-t}, \quad \mathbf{X}_t \rightarrow \mathbf{V}^t(\Delta\lambda;x),$$
$$b_t \rightarrow P_t(x,y)/D(\Gamma_t;x), \quad \mathbf{Y}_t \rightarrow \mathbf{R}^t(\Delta\lambda;x).$$

Invoking also the fact that the highest-degree term in $P_t(x,y)$ is $y^{\mathscr{M}-t}$ yields $D(\Gamma_t;x) = (A_{t-1}/A_t)^{1/2}$; hence, this implies Eqs. (2.24) and (2.25) above. Since, by convention, we have specified the sign of $P_t(x,y)/D(\Gamma_t;x)$, there is no further freedom in choosing the signs in Eqs. (2.23) and (2.25).

Equations (2.23)–(2.25) are quite important for they give the explicit expressions for the invariant (Racah) coefficients of the type

$$\left\{\begin{pmatrix} p & q & 0 \\ & \Gamma_t & \end{pmatrix}\begin{pmatrix} & \max & \\ p-q & 0 & 0 \\ & \Gamma' & \end{pmatrix}\begin{pmatrix} q & q & 0 \\ & \Gamma'' & \end{pmatrix}\right\}([m^f])$$

and for all projective functions having maximal lower pattern [see Eq. (2.5a)].

Our purpose in the present paper is to focus on the denominator function, which is uniquely defined by the above procedure to have the form

$$D^2(\Gamma_t;x) = A_{t-1}/A_t. \tag{2.26}$$

The determinantal form of this result is, to be sure, quite complicated; it is the goal of the present paper to show that remarkable simplifications can be effected.

For explicitness, let us give the elements of the determinant A_t in terms of the quantities introduced in Eqs. (2.10):

$$(\mathbf{V}^r,\mathbf{V}^s) = \frac{\mathrm{Dim}(x^f)}{[\mathrm{Dim}(x)]^2}\,\frac{(-1)^{q+1-\Delta_2}(q!)^4(p-q)!}{L(\Delta;x)L(qqq;x)} \times\sum_n F^r_n(\lambda;x)F^s_n(\lambda;x)N_n(\Delta;x), \tag{2.27}$$

each r, $s = 1, 2, \ldots, t$. The factors $L(\Delta;x)$, $L(qqq;x)$, and $N_n(\Delta;x)$ are defined in Eqs. (2.8h) and (2.10b); the factors F^s_n by the expansion (2.13); and the summation in (2.27) is over all $n = (n_1,n_2,n_3) \in \mathbb{D}(p,q,\Delta)$ [see Eq. (2.8d)].

For future reference, it is convenient also to define the function $H_n(\Delta\lambda;x,y)$ by

$$H_n(\Delta\lambda;x,y) = \frac{\Pi_{ijk}(y+1+\frac{1}{3}(x_j-x_k)+\lambda_i)_{n_i-\lambda_i}}{D_n(\Delta;x)}; \tag{2.28a}$$

the components of the vectors $\mathbf{V}^t$ are then defined by the expansion

$$H_n(\Delta\lambda;x,y) = \sum_{t=1}^{\mathscr{M}} y^{\mathscr{M}-t}V^t_n(\Delta\lambda;x). \tag{2.28b}$$

The preceding results complete the definition of the denominator function $D(\Gamma_t;x)$ in its determinantal form, Eq. (2.26). Before considering the simplification of this result, let us note that an important property of this function is its symmetries. These are summarized and proved in Sec. V. A preliminary identity required in that analysis stems from the following key idea: The construction given in Eq. (2.17) can be effected in two distinct ways, which must give identical answers for the projective operator Γ_t. The two distinct ways to calculate are (i) couple in the "normal order" $[p-q\,0\,0]\otimes[q\,q\,0]$ as in Eq. (2.3); or (ii) couple in the "opposite order," that is, according to $[q\,q\,0]\otimes[p-q\,0\,0]$. For this second coupling, we have a relation, which is similar to Eq. (2.3):

$$\begin{bmatrix} & \Gamma'' & \\ q & q & 0 \\ q & q & \\ & q & \end{bmatrix}\begin{bmatrix} & \Gamma' & \\ p-q & 0 & 0 \\ p-q & 0 & \\ & p-q & \end{bmatrix} = \sum_{t=1}^{\mathscr{M}}\left\{\begin{pmatrix} p & q & 0 \\ & \Gamma_t & \end{pmatrix}\begin{pmatrix} & \max & \\ q & q & 0 \\ & \Gamma'' & \end{pmatrix}\begin{pmatrix} p-q & 0 & 0 \\ & \Gamma' & \end{pmatrix}\right\}\begin{bmatrix} & \Gamma_t & \\ p & q & 0 \\ p & q & \\ & p & \end{bmatrix}. \tag{2.29}$$

It is to be noted that although the invariant factor $\{\cdots\}$ in this relation differs from that in Eq. (2.3), the projective operator is, and must be, identical.

Clearly, starting from Eq. (2.29), we can repeat the entire calculation, which starts with Eq. (2.3) and leads to the results in Eqs. (2.22)–(2.25); this will lead us then to relationships between the sets of results calculated in the two ways. Indeed, the entire calculation already given can be taken over immediately upon identifying the form that replaces H_n given by Eq. (2.28a). This we now do (in outline).

Let us use script letters $\mathscr{D}_n$, $\mathscr{H}_n$, $\mathscr{V}^t_n$, ... to denote quantities analogous to D_n, H_n, V^t_n, ..., but originating now from the opposite coupling order, Eq. (2.29). Thus, we evaluate the operator relation (2.29) on the initial $U(3)\supset U(2)$ states $[m]=[m_{13},m_{23},m_{33}]$, $[m']=[m_{12},m_{22}]$, so that the final states are $[m]+[\Delta]$, $[m']+[q,q]$, *exactly as in the normal coupling order*. We find that the factor analogous to $H_n(\Delta,\lambda;x)$ defined by Eq. (2.28a) is

$$\mathscr{H}_n(\Delta,\lambda;x)=\prod_{ijk}(y'+1+\tfrac{1}{3}(x_j-x_k)-2\Delta_i+\Delta_j+\Delta_k+\lambda_i)_{n_i-\lambda_i}\times[\mathscr{D}_n(\Delta;x)]^{-1}. \tag{2.30}$$

where y' is defined in terms of the variable y in Eq. (2.11) by

$$y'=-y+1-\tfrac{1}{3}(p-2q). \tag{2.31}$$

Thus, the numerator term in this result is (up to sign) just the opposing arrow factor, analogous to (2.6b) and (2.13), but calculated from the coupling order in Eq. (2.29). Similarly, $\mathscr{D}^2_n(\Delta;x)$ is the denominator term analogous to that given by Eqs. (2.9), but with the denominators in the left-hand side of that result composed in the opposite order. From this direct calculation, one finds the following relation between the two functions $\mathscr{D}^2_n(\Delta;x)$ and $D^2_n(\Delta;x)$:

$$\mathscr{D}^2_n(\Delta;x)=\frac{\mathrm{Dim}(x)}{\mathrm{Dim}(x^f)}D^2_n(\Delta;-x^f), \tag{2.32a}$$

where the variables x^f (f for final) are defined in terms of x by

$$x_i^f=x_i+\Delta_j-\Delta_k,\quad (ijk)\text{ cyclic}. \tag{2.32b}$$

Thus, the barycentric variables (x_1^f,x_2^f,x_3^f) are defined in terms of the final state vector labels $p_{i3}^f=p_{i3}+\Delta_i$ $(i=1,2,3)$ in exactly the same way that the (x_1,x_2,x_3) are defined in terms of the initial labels p_{i3} $(i=1,2,3)$. But now one sees from this result and the numerator in Eq. (2.30) that the functions $\mathscr{H}_n$ and H_n are related by

$$\mathscr{H}_n(\Delta,\lambda;x,y)=\left[\frac{\mathrm{Dim}(x^f)}{\mathrm{Dim}(x)}\right]^{1/2}H_n(\Delta,\lambda;-x^f,y'). \tag{2.33}$$

This is the key relation for obtaining the consequences implied by the normal and opposite coupling orders in Eqs. (2.3) and (2.29), respectively.

In order to obtain the results implied by Eq. (2.33), let us next write out the analog of Eq. (2.17) for the coupling (2.29):

$$\mathscr{H}_n(\Delta,\lambda;x,y)=\sum_{t=1}^{\mathscr{M}}y^{\mathscr{M}-t}\mathscr{V}^t_n(\Delta,\lambda;x)$$
$$=\sum_{t=1}^{\mathscr{M}}\mathscr{R}^t_n(\Delta,\lambda;x)\frac{P_t(x,y)}{D(\Gamma_t;x)}, \tag{2.34a}$$

where we have defined

$$\mathscr{R}^t_n(\Delta,\lambda;x)=(-1)^{\Phi}\times\left\{\begin{pmatrix} p & q & 0\\ & \Gamma_t & \end{pmatrix}\begin{pmatrix} & \max & \\ q & q & 0\\ & \Gamma'' & \end{pmatrix}\begin{pmatrix} p-q & 0 & 0\\ & \Gamma' & \end{pmatrix}\right\}([m^f]) \tag{2.34b}$$

in which the phase Φ turns out to be exactly that given by Eq. (2.15c).

The first result implied by relation (2.33) is that the components of the vectors $\mathscr{V}^t(\Delta,\lambda;x)$ are related to those of $\mathbf{V}^t(\Delta,\lambda;x)$ by

$$\mathscr{V}^t_n(\Delta,\lambda;x)=\left[\frac{\mathrm{Dim}(x^f)}{\mathrm{Dim}(x)}\right]^{1/2}\sum_{s=1}^{t}a_{st}V^t_n(\Delta,\lambda;-x^f), \tag{2.35}$$

where the coefficients a_{st} may be determined from the binomial expansion of $(y')^{\mathscr{M}-t}$ into a sum of powers of y; in particular, $a_{tt}=(-1)^{\mathscr{M}-t}$.

Let us now recall that the Gram–Schmidt process is invariant under triangular transformations of the initial set of (ordered) linearly independent vectors in the sense that one obtains the same final set (up to normalization conventions for signs); in particular, the Gram determinant is *invariant* (no sign changes) to transformations having ± 1 on the diagonal. Thus, we find

$$\mathscr{A}_t(\Delta,\lambda;x)=\left[\frac{\mathrm{Dim}(x^f)}{\mathrm{Dim}(x)}\right]^tA_t(\Delta,\lambda;-x^f), \tag{2.36}$$

where $\mathscr{A}_t$ denotes the $t\times t$ Gram determinant of the vectors $\mathscr{V}^t$ [Eq. (2.22c) with script letters replacing the Latin ones].

Since Eqs. (2.23)–(2.25) are valid under the substitutions $V^t_n\to\mathscr{V}^t_n$, $A_t\to\mathscr{A}_t$, $R^t_n\to\mathscr{R}^t_n$, $y\to y'$, Eqs. (2.35) and (2.36) imply relations between the invariant functions for the two coupling orders, between the sets of numerator factors $\{P_t(x,y)\}$ and $\{P_t(-x^f,y)\}$ and between the denominator functions $D(\Gamma_t;x)$ and $D(\Gamma_t;-x^f)$. Of all these, the one of interest here is that for the denominator function, which must be the same in either of the forms, so that

$$D(\Gamma_t;x)=[A_{t-1}(\Delta,\lambda;x)/A_t(\Delta,\lambda;x)]^{1/2}$$
$$=[\mathscr{A}_{t-1}(\Delta,\lambda;x)/\mathscr{A}_t(\Delta,\lambda;x)]^{1/2};$$

that is, the following symmetry is true:

$$D(\Gamma_t;x)=\left[\frac{\mathrm{Dim}(x)}{\mathrm{Dim}(x^f)}\right]^{1/2}D(\Gamma_t;-x^f). \tag{2.37}$$

III. DEFINITION OF THE FUNCTION G^t_q

In the present section we rewrite the denominator function in a new form that will prove useful for determining its properties. While we could introduce this form [Eq. (3.3) below] in an *ad hoc* fashion, we attempt now to motivate this step. This requires bringing in some general aspects of the characteristic null space of a canonical unit tensor operator, which, as shown in the Introduction, has a definitive role in determining the SU(3) invariants $D(\Gamma_t;x)$.

In our ealier work (Refs. 5,6, and 8), we were able to show that for $t=1$, that is, for the function $D^2(\Gamma_1;x)$, one has

$$\frac{1}{D^2(\Gamma_1;x)}=A_1=(\mathbf{V}^1,\mathbf{V}^1)$$
$$=\frac{(-1)^{\Delta_2}}{C^1_{p,q}}\frac{\mathrm{Dim}(x^f)}{\mathrm{Dim}(x)}\frac{G^1_q(\Delta;x)}{L(\Delta;x)}. \tag{3.1}$$

The quantities appearing in this expression have the following definitions.

(i) The constant $C^1_{p,q}$ is given by

$$C^1_{p,q}=1/(p-q)!. \tag{3.2a}$$

(ii) The factor Dim(x) is the dimension of irrep $[m]$ (Weyl dimension formula) given by

$$\mathrm{Dim}(x) = -x_1x_2x_3/2, \tag{3.2b}$$

and $\mathrm{Dim}(x^f)$ is the same function defined in terms of the final irrep labels $[m]+[\Delta]$, so that

$$x_i^f = x_i + \Delta_j - \Delta_k, \quad (ijk)\text{cyclic}. \tag{3.2c}$$

(iii) The factor $L(\Delta;x)$ is the product of linear factors defined by

$$L(\Delta;x) = \prod_{ijk} \Delta_i!(x_i+1)_{\Delta_j}(-x_i+1)_{\Delta_k}. \tag{3.2d}$$

(iv) The factor $G_q^1(\Delta;x)$ [denoted $G_q(\Delta;x)$ in Refs. 6 and 8] is a polynomial in the variables $\Delta_1,\Delta_2,\Delta_3,\Delta_2+x_1$, Δ_3+x_2, Δ_1+x_3, Δ_3-x_1, Δ_1-x_2, Δ_2-x_3. This polynomial is given explicitly in Ref.6, where its symmetries and zeros are discussed in detail. (These properties of G_q motivated much of the present generalization).

Remark: It is nontrivial to prove directly that the 1×1 determinant $(\mathbf{V}^1, \mathbf{V}^1)$ in Eq. (3.1), as given by Eq. (2.27), defines precisely the *same* polynomial $G_q^1(\Delta;x)\equiv G(\Delta;x)$ given in Ref. 6. The validity of this identification is based on the occurrence of one and the same denominator function $D^2(\Gamma_t;x)$ in the present development and in the previous one (Ref. 6). [Equation (3.1) is expression (1.2) of Ref. 6, rewritten in a slightly modified form to suit the present discussion].

One of the striking features of the right-hand side of Eq. (3.1) is that each of the factors $L(\Delta;x)$ and $G_q^1(\Delta;x)$ has a direct interpretation in terms of the characteristic null space of the Wigner operator with which it is associated (the dimension factors are related to the normalization convention for Wigner operators): for the case being discussed ($t=1$) the operator is the "stretched operator" denoted by Γ_1, which transforms as irrep $[p\,q\,0]$, effects the shift $(\Delta_1,\Delta_2,\Delta_3)$ with $\lambda=(\lambda_1,\lambda_2,\lambda_3)=(0,0,0)$, and has maximal null space as compared to the other operators $\Gamma_2,\Gamma_3,\ldots,\Gamma_{\mathscr{M}}$ in the same multiplicity set. For this operator, $L(\Delta;x)$ is a product [the product (3.2d)] of linear factors corresponding to known lines of lattice points of zeros (and their symmetries) in the null space diagram; the polynomial factor $G_q^1(\Delta;x)$ has the property that it is the polynomial of lowest degree (q in this case) that can possess known symmetries and vanish on the lattice points of an equilateral triangle with side q [$q(q+1)/2$ lattice points in all], where the necessity of occurrence of this particular triangle of zeros can be uniquely associated with the maximal null space operator in question (see Ref. 6).

Guided by these results for the Γ_1 denominator function, the known results for all $\langle 4\,2\,0\rangle$ operators,[15] and the fact that the generalization of Eq. (3.1) should be at most to rational polynomials (ratio of two polynomials), and the form A_{t-1}/A_t of Eq. (2.26), we now write the denominator function as

$$\frac{1}{D^2(\Gamma_t;x)} = \frac{(-1)^{\Delta_2-t+1}}{C_{p,q}^t}\frac{\mathrm{Dim}(x^f)}{\mathrm{Dim}(x)}\frac{1}{L_t(\Delta;x)}\frac{G_q^t(\Delta;x)}{G_q^{t-1}(\Delta;x)}, \tag{3.3}$$

where $t=1,2,\ldots,\mathscr{M}$ with $G_q^0(\Delta;x)=1$. Since we know for $t=1$ that Eq. (3.3) is correct even when $\lambda\neq(0,0,0)$ in consequence of a reduction formula (to be discussed in the next section), we follow this guideline (known to work for the $\langle 4\,2\,0\rangle$ operators) and define the factor $L_t(\Delta;x)$ from the known lines of lattice points of zeros in the null space diagram for the operator Γ_t having $\lambda=(0,0,0)$ (Ref. 5). In analogy with the $G_q^1(\Delta;x)$ result, we expect the form (3.3) to be valid even for operators Γ_t for which $\lambda\neq(0,0,0)$. This is validated in the next section, where it is demonstrated that for general λ Eq. (3.3) reduces in exactly the correct way to supply a "final" set of linear factors in exact agreement with the known lines of lattice points of zeros for the general operator Γ_t with $\lambda=(\lambda_1,\lambda_2,\lambda_3)$ general. Taking these results into account, we define the factor $L_t(\Delta;x)$ by

$$\begin{aligned} L_t(\Delta;x) &= L(\Delta_1-t+1,\Delta_2-t+1,\Delta_3-t+1;x)\\ &= \prod_{ijk}(\Delta_i-t+1)!(x_i+1)_{\Delta_j-t+1}\\ &\quad\times(-x_i+1)_{\Delta_k-t+1}, \end{aligned} \tag{3.4a}$$

where we have put the factors in the product in symmetrically to have agreement for $t=1$ with Eq. (3.2d). We make no assumptions regarding the functions $G_q^t(\Delta;x)$, except that $G_q^1(\Delta;x)$ be the known result.

The constant factor (independent of x) in Eq. (3.3) is defined by

$$C_{p,q}^t = \frac{(t-1)!(p-t+2)!(q-t+1)!}{(p-q)!(p+1)!q!}. \tag{3.4b}$$

This definition serves to define fully the function G_q^t [see Eq. (3.5a) below]; in making this definition we anticipate certain conventions for the function G_q^t [including agreement with Eq. (3.2a) for $t=1$] to be discussed in Sec. VII.

We now obtain the fully explicit definition of G_q^t by iterating Eq. (3.3) in $t=1,2,\ldots$ and comparing the resulting iterate at each step with those of $1/D^2(\Gamma_t;x)=A_t/A_{t-1}$. The result is

$$G_q^t(\Delta;x) = A_t(\Delta,\lambda;x)\prod_{s=1}^{t} L_s'(\Delta;x), \tag{3.5a}$$

where we have defined, for convenience of expression,

$$L_s'(\Delta;x) = C_{p,q}^s\frac{\mathrm{Dim}(x)}{\mathrm{Dim}(x^f)}L_s(\Delta;x). \tag{3.5b}$$

In Eq. (3.5a), $A_t=A_t(\Delta,\lambda;x)$ is the Gram determinant defined earlier in Eqs. (2.22) and (2.27).

Finally, it is convenient to express G_q^t as given by Eqs. (3.5) in terms of the "rationalized" elements of the determinant A_t as given by Eq. (2.27). Thus, defining N^{rs} by

$$N^{rs} = N^{rs}(\Delta,\lambda;x) = \sum_n F_n^r(\lambda;x)F_n^s(\lambda;x)N_n(\Delta;x) \tag{3.6a}$$

and the $t\times t$ determinant $\overline{A}_t$ by

$$\overline{A}_t = \overline{A}_t(\Delta,\lambda;x) = \det(N^{rs}), \tag{3.6b}$$

we find

$$\begin{aligned} &G_q^t(\Delta;x)\\ &\quad= (-1)^{t(q+1)}\prod_{s=1}^{t}\frac{(s-1)!(p-s+2)!(q-s+1)!}{(p+1)!}\\ &\qquad\times\prod_{s=1}^{t}\frac{1}{\Pi_{ijk}(-\Delta_i)_{s-1}(-x_i-\Delta_j)_{s-1}(x_i-\Delta_k)_{s-1}}\\ &\qquad\times\left[\frac{(q!)^3}{\mathrm{Dim}(x)L(qqq;x)}\right]^t\overline{A}_t(\Delta,\lambda;x). \end{aligned} \tag{3.6c}$$

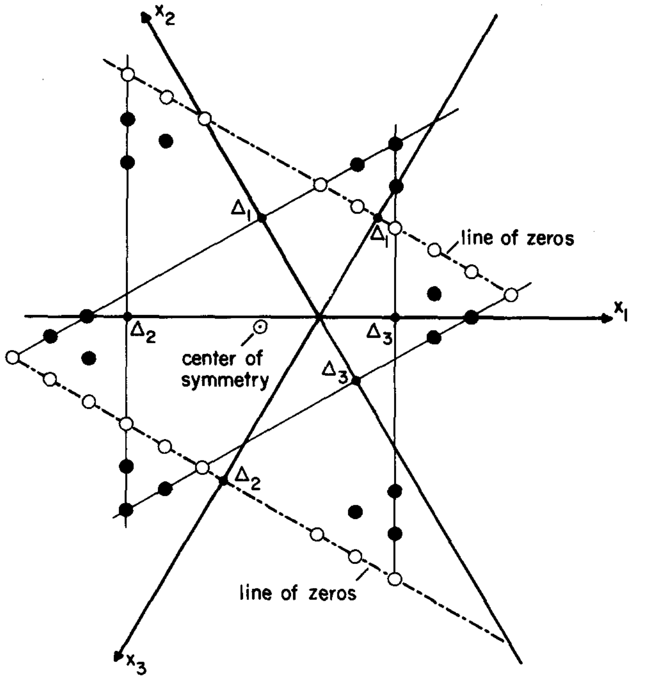


FIG. 4. Zeros of the polynomial $G_3(3\ 5\ 2;x_1\ x_2\ x_3)$. This polynomial vanishes at each of the six points (three large open circles and three large solid circles) of each of the six equilateral triangles symmetrically placed about the center of symmetry at the point $(-\frac{3}{2},\frac{1}{2},1)$. The linear factors of the polynomial are $(x_3+3)(x_3-5)$. Hence, the polynomial also vanishes on the lines $x_3=-3$ and $x_3=5$ (the dash–dot lines). Removing these linear factors from the polynomial leaves the new polynomial $G_2(2\ 4\ 3;x_1+1,x_2-1,x_3)$, which still vanishes at each of three points (the large solid circles) of each of six equilateral triangles still symmetrically placed about the center of symmetry.

We emphasize again that no assumptions concerning the properties of G_q^t have entered into the derivation of Eq. (3.6c): We could have introduced Eq. (3.3) at the outset as a purely mathematical step and arrived at Eq. (3.6c). We have attempted to motivate this step by pointing out its relationship to other results, thereby providing at least the suggestion that each G_q^t may be polynomial, despite the formidable determinantal formulation given it at this point. Working directly from Eqs. (3.5) and (3.6a) and the properties of the Gram determinant A_t, we seek to determine the significant properties of these functions G_q^t.

IV. REDUCTION FORMULA FOR G_q^t

One of the more remarkable properties of the polynomial $G_q^1(\Delta;x)=G_q(\Delta;x)$ developed and discussed in Ref. 5 is the unexpected manner in which the *implicit* dependence of this function on the step-function parameters $\lambda=(\lambda_1,\lambda_2,\lambda_3)$ account for the further factoring of linear terms from $G_q(\Delta;x)$ in just the right way to accommodate lines of lattice points of zeros that must develop because the sides of the six triangle of zeros (one in the lexical region of the Möbius plane) and five other triangles of zeros (occurring because of symmetry) become aligned. Because of the importance of this geometrical phenomenon in understanding this function, a null space diagram illustrating it (see Fig. 4) in a special case is reproduced from Ref. 5. It is significant that the polynomial remaining after all linear terms are factored out is again a polynomial of the G_q^t type. In the example, it is $G_2(243;x_1+1,x_2-1,x_3)$.

The property reviewed briefly above is a key one for understanding and developing the properties of the general G_q^t functions defined by Eq. (3.6c) of the last section. Despite the fact that we have not yet shown G_q^t to be polynomial, we can still show that it reduces in just the desired way, as will now be demonstrated.

The mathematical transformation that underlies this behavior was already given in Ref. 5 [see Eqs. (2.13) and (2.15) there]. It is a transformation of the form

$$(p,q,\Delta,\lambda,x)\to(\hat{p},\hat{q},\hat{\Delta},\hat{\lambda},\hat{x}), \tag{4.1a}$$

where

$$\hat{x}_i=x_i-\lambda_j+\lambda_k, \tag{4.1b}$$

$$\hat{\Delta}_i=\Delta_i+\lambda_i-\lambda_j-\lambda_k,\quad (ijk)\text{ cyclic;} \tag{4.1c}$$

$$\hat{p}=p,\hat{q}=q-(\lambda_1+\lambda_2+\lambda_3)=\mathscr{M}-1. \tag{4.1d}$$

The following results are then obtained from the above:

$$\hat{\Delta}_1+\hat{\Delta}_2+\hat{\Delta}_3=\hat{p}+\hat{q}, \tag{4.2a}$$

$$\hat{\lambda}=(\hat{\lambda}_1,\hat{\lambda}_2,\hat{\lambda}_3)=(0,0,0). \tag{4.2b}$$

In this last result $\hat{\lambda}_1$ is defined by

$$\hat{\lambda}_i=\max(0,\hat{q}-\hat{\Delta}_i), \tag{4.2c}$$

in accord with Eq. (2.8b).

The key relationship for obtaining the reduction formula for the general G_q^t [Eq. (4.15) below] is Eq. (2.17) as formulated in terms of H_n given by Eq. (2.28a). As we now show, it is the transformation properties of the function H_n that underlies the reduction formula (as well as some of the symmetries discussed in the next section).

Consider then the properties of the function H_n under the transformation (4.1) above. For this purpose, it is convenient to define the function $h_m(\lambda;x,y)$ by

$$h_m(\lambda;x,y)=\prod_{ijk}\Big(y+1+\frac{1}{3}(x_j-x_k)+\lambda_i\Big)_{m_i}, \tag{4.3a}$$

where $m=(m_1,m_3,m_3)$ is an arbitrary 3-tuple with non-negative integer components m_i. The numerator factor in H_n in Eq. (2.28a) is then given by [see Eq. (2.8e)]

$$h_{n-\lambda}(\lambda;x,y)=\prod_{ijk}\Big(y+1+\frac{1}{3}(x_j-x_k)+\lambda_i\Big)_{n_i-\lambda_i}. \tag{4.3b}$$

In terms of the notation (4.3a), we can also unambiguously define

$$\begin{aligned}h_{n-\lambda}(0;x,y)&=h_m(0,0,0;x,y)|_{m=n-\lambda}\\&=\prod_{ijk}\Big(y+1+\frac{1}{3}(x_j-x_k)\Big)_{n_i-\lambda_i}.\end{aligned} \tag{4.3c}$$

The first identity that can now be proved by direct substitution of the new variables (4.1) is

$$h_{n-\lambda}(\lambda;x,y)=h_{n-\lambda}(0;\hat{x},\hat{y}), \tag{4.4a}$$

where $\hat{y}$ is defined by

$$\hat{y}=y+\tfrac{1}{3}(q+1-\mathscr{M}). \tag{4.4b}$$

The second identity we require is somewhat more difficult to prove, but also follows by direct substitution:

$$[a_{p,q}\ \mathrm{Dim}(x^f)]^{1/2}D_n(\Delta;x) = [a_{\hat{p},\hat{q}}\ \mathrm{Dim}(\hat{x}^f)]^{1/2}D_{n-\lambda}(\hat{\Delta};\hat{x}), \tag{4.5a}$$

where we have defined the constant $a_{p,q}$ by

$$a_{p,q} = (p-q)!q!. \tag{4.5b}$$

The only subtlety in the derivation of Eq. (4.5a) is that one must recognize identities such as

$$\frac{[x_i + n_k + \Delta_j - q + \lambda_j]_{\lambda_j}[x_i - n_j + q - \Delta_k]_{\lambda_k}}{[x_i + n_k]_{\lambda_j}[x_i - n_j + \lambda_k]_{\lambda_k}} = 1,$$

which is an unconventional falling factorial relation that results from the fact that the λ_i assume only two values, 0 or $q - \Delta_i$.

In terms of the notation above, we write the function H_n defined by Eq. (2.28a) as

$$H_n(\Delta,\lambda;x,y) = h_{n-\lambda}(\lambda;x,y)/D_n(\Delta;x), \tag{4.6a}$$

and similarly,

$$H_{n-\lambda}(\hat{\Delta},0;\hat{x},\hat{y}) = h_{n-\lambda}(0;\hat{x},\hat{y})/D_{n-\lambda}(\hat{\Delta};\hat{x}). \tag{4.6b}$$

Using Eqs. (4.4a) and (4.5a) proved above, we thus obtain

$$\frac{H_n(\Delta,\lambda;x,y)}{[a_{p,q}\ \mathrm{Dim}(x^f)]^{1/2}} = \frac{H_{n-\lambda}(\hat{\Delta},0;\hat{x},\hat{y})}{[a_{\hat{p},\hat{q}}\ \mathrm{Dim}(\hat{x}^f)]^{1/2}}, \tag{4.6c}$$

where this relation is an identity under the transformation (4.1).

Let us next use relation (4.6c) to determine the desired reduction formula. For this, we consider the expansion of H_n into the vectors V_n^t as given by Eq. (2.28b), and the similar expansion for $H_{n-\lambda}$ in the right-hand side of Eq. (4.6c), which is

$$H_{n-\lambda}(\hat{\Delta},0;\hat{x},\hat{y}) = \prod_{t=1}^{\mathscr{M}} \hat{y}^{\mathscr{M}-t}V_{n-\lambda}^t(\hat{\Delta},0;\hat{x}). \tag{4.7}$$

Using next the identity (4.6c), we find the following relation between the components of the vector $\mathbf{V}^t(\Delta,\lambda;x)$ and those of $\mathscr{V}^t(\hat{\Delta},0;\hat{x})$:

$$V_n^t(\Delta,\lambda;x) = \left[\frac{a_{p,q}\ \mathrm{Dim}(x^f)}{a_{\hat{p},\hat{q}}\ \mathrm{Dim}(\hat{x}^f)}\right]^{1/2} \sum_{s=1}^{t} A_{ts}V_{n-\lambda}^s(\hat{\Delta},0;\hat{x}), \tag{4.8}$$

where the coefficients A_{ts} are determined completely from the binomial expansion of $(\hat{y})^{\mathscr{M}-t}$ as a sum of powers in y; in particular, $A_{tt} = 1$.

Recalling again that the Gram–Schmidt process is invariant under triangular transformation of the type (4.8) so that the Gram determinant is also *invariant* (no sign changes), we find that the Gram determinants $A_t(\Delta,\lambda;x)$ and $A_t(\hat{\Delta},0;\hat{x})$ are related by

$$A_t(\Delta,\lambda;x) = \left[\frac{a_{p,q}\ \mathrm{Dim}(x^f)}{a_{\hat{p},\hat{q}}\ \mathrm{Dim}(\hat{x}^f)}\right]^t A_t(\hat{\Delta},0;\hat{x}), \tag{4.9}$$

each $t = 1,2,\ldots,\mathscr{M}$.

We can now use Eq. (4.9) to obtain a relation between two different denominator functions and two different G_q^t functions. First, let us rewrite Eq. (2.26) in the fully explicit form

$$D^2(\Gamma_t;x) = A_{t-1}(\Delta,\lambda;x)/A_t(\Delta,\lambda;x). \tag{4.10a}$$

In particular, this relation applies also to the family of operators having irrep labels $[\hat{p}\ \hat{q}\ 0]$, shift pattern $(\hat{\Delta}_1,\hat{\Delta}_2,\hat{\Delta}_3)$, and operator patterns $\hat{\Gamma}_1,\hat{\Gamma}_2,\ldots,\hat{\Gamma}_{\mathscr{M}}$. The denominator function for each of these operators is

$$D^2(\hat{\Gamma}_t;\hat{x}) = A_{t-1}(\hat{\Delta},0;\hat{x})/A_t(\hat{\Delta};0;\hat{x}), \tag{4.10b}$$

each $t = 1,2,\ldots,\mathscr{M}$. Equation (4.9) yields the following relation between the denominator functions (4.10):

$$[a_{p,q}\ \mathrm{Dim}(x^f)]^{1/2}D(\Gamma_t;x) = [a_{\hat{p},\hat{q}}\ \mathrm{Dim}(\hat{x}^f)]^{1/2}D(\hat{\Gamma}_t;\hat{x}). \tag{4.11}$$

This result expresses the invariance of the form $[a_{p,q}\ \mathrm{Dim}(x^f)]^{1/2}D(\Gamma_t;x)$ to the transformation (4.1).

Using either Eq. (3.3) and the above result (4.11) or Eq. (3.5a) and Eq. (4.9), we now find a relation between G_q^t-type functions. First, let us rewrite Eq. (3.5a) in terms of the present notation:

$$G_q^t(\Delta,\lambda;x) = A_t(\Delta,\lambda;x)\prod_{s=1}^{t} L_s^t(\Delta;x), \tag{4.12a}$$

where we introduce the implicit dependence of G_q^t on λ. Specializing this result to $\lambda = (0,0,0)$ and renaming variables then gives

$$G_{\hat{q}}^t(\hat{\Delta},0;\hat{x}) = A_t(\hat{\Delta},0;\hat{x})\prod_{s=1}^{t} L_s^t(\hat{\Delta};\hat{x}). \tag{4.12b}$$

Combining Eqs. (4.12) and (4.9) yields the desired reduction formula in the form

$$G_q^t(\Delta,\lambda;x) = G_{\hat{q}}^t(\hat{\Delta},0;\hat{x})\prod_{s=1}^{t}\frac{a_{p,q}\ \mathrm{Dim}(x^f)L_s^t(\Delta;x)}{a_{\hat{p},\hat{q}}\ \mathrm{Dim}(\hat{x}^f)L_s^t(\hat{\Delta};\hat{x})}. \tag{4.13a}$$

The product in this result may be simplified using Eqs. (3.4), (3.5), and (4.5b):

$$\begin{aligned}\frac{a_{p,q}\ \mathrm{Dim}(x^f)L_s^t(\Delta;x)}{a_{\hat{p},\hat{q}}\ \mathrm{Dim}(\hat{x}_f)L_s^t(\hat{\Delta};\hat{x})} &= (-1)^{q+1-\mathscr{M}}\frac{(q-s+1)!}{(\mathscr{M}-s)!}\prod_{i=1}^{3}\frac{(\Delta_i - s+1)!}{(\hat{\Delta}_i - s+1)!}\\ &\quad\times\prod_{ijk}(x_i - \Delta_k + s - 1)_{\lambda_i}(-x_i - \Delta_j + s - 1)_{\lambda_i}.\end{aligned} \tag{4.13b}$$

Finally, writing (by suppressing the implicit λ again)

$$G_q^t(\Delta;x) = G_q^t(\Delta,\lambda;x), \tag{4.14a}$$

$$G_{\hat{q}}^t(\hat{\Delta};\hat{x}) = G_{\hat{q}}^t(\hat{\Delta},0,\hat{x}), \tag{4.14b}$$

we obtain the following theorem.

Theorem 4.1: The function G_q^t satisfies the reduction formula

$$\begin{aligned}G_q^t(\Delta;x) &= (-1)^{t(q+1-\mathscr{M})}\prod_{s=1}^{t}\frac{(q-s+1)!}{(\mathscr{M}-s)!}\\ &\quad\times\prod_{i=1}^{3}\frac{(\Delta_i - s+1)!}{(\hat{\Delta}_i - s+1)!}\\ &\quad\times\prod_{s=1}^{t}\prod_{ijk}(x_i - \Delta_k + s - 1)_{\lambda_i}(-x_i - \Delta_j + s - 1)_{\lambda_i}\\ &\quad\times G_{\hat{q}}^t(\hat{\Delta};\hat{x}),\end{aligned} \tag{4.15}$$

each $t = 1,2,\ldots,\hat{q}+1 = \mathscr{M}$. ■

For future reference, let us also note the identity

$$\prod_{i=1}^{3}\frac{(\Delta_i - s+1)!}{(\hat{\Delta}_i - s+1)!} = \prod_{ijk}\frac{[\Delta_i + \lambda_i - s + 1]_{\lambda_j+\lambda_k}}{[q-s+1]_{\lambda_i}}. \tag{4.16}$$

This relation together with Eq. (4.15) shows that the terms that factor from $G_q^t(\Delta;x)$ for $\lambda=(\lambda_1,\lambda_2,\lambda_3)\neq(0,0,0)$ are linear factors in $\Delta_1,\Delta_2,\Delta_3,\Delta_2+x_1,\Delta_3+x_2,\Delta_1+x_3,\Delta_3-x_1,$ $\Delta_1-x_2,\Delta_2-x_3$.

V. SYMMETRIES OF THE DENOMINATOR FUNCTION $D(\Gamma_t;x)$ AND THE FUNCTION $G_q^t(\Delta;x)$

For the function $G_q^1(\Delta;x)$, which occurs in the denominator function for the stretched operator [see Eq. (3.1)], we established in Ref. 6 the existence of some remarkable symmetries. In the present section, we extend these symmetries to the general function $G_q^t(\Delta;x)$, $t=1,\dots,\mathscr{M}-1$.

In order to describe these symmetries, it is convenient to replace the variables (Δ,x) by a 3×3 array:

$$A=\begin{bmatrix}\Delta_1-t+1 & \Delta_2-t+1+x_1 & \Delta_3-t+1-x_1\\ \Delta_2-t+1 & \Delta_3-t+1+x_2 & \Delta_1-t+1-x_2\\ \Delta_3-t+1 & \Delta_1-t+1+x_3 & \Delta_2-t+1-x_3\end{bmatrix}. \tag{5.1}$$

Note that not all of the nine variables in the array A are independent since $\Delta_1+\Delta_2+\Delta_3=p+q$ and $x_1+x_2+x_3=0$. These constraints suffice to make the sum of entries in each row and the sum of entries in each column equal to $p+q-3t+3$. The matrix array A is therefore a 3×3 magic square. [The array introduced in Ref. 6 was the special case of (5.1) having $t=1$.]

In terms of this array, we write the denominator function and G function as

$$D(\Gamma_t;x)=D_q^t(A),\quad G_q^t(\Delta;x)=G_q^t(A). \tag{5.2}$$

We shall demonstrate in this section that the function $G_q^t(A)$, for fixed q and t, is invariant under the $72=6\times6\times2$ transformations corresponding to row permutations, column permutations, and transposition of the array A. We call such a transformation of the variables $(\Delta_1,\Delta_2,\Delta_3,x_1,x_2,x_3)$ a determinantal transformation and refer to the invariance of a function $f(A)$ under these transformations as determinantal symmetry.

For example, the transformation T of the variables (Δ,x) corresponding to transposition of the array A is

$$\begin{aligned}T:\quad &\Delta_1\to\Delta_1,\quad &&\Delta_2\to\Delta_2+x_1,\quad &&\Delta_3\to\Delta_3-x_1,\\ &x_1\to-x_1,\quad &&x_2\to-x_3,\quad &&x_3\to-x_2,\end{aligned} \tag{5.3}$$

and conversely.

The existence of determinantal symmetry for the functions $G_q^1(\Delta;x)$ symmetry was found earlier. In generalizing this special case, two methods are available: The first uses properties of the generating relation, Eq. (2.17); the second uses the definition (3.6) directly.

The idea behind the first method of proof (using the properties of the generating function H_n) is the following: Since Eq. (2.28b) is an identity, we may regard the variables y,x_1,x_2,x_3 in the function H_n as indeterminates: the procedure of identifying first the vectors $\mathbf{V}^t(\Delta,\lambda;x)$ and then effecting the Gram–Schmidt procedure on the ordered set of vectors $\mathbf{V}^1,\mathbf{V}^2,\dots,\mathbf{V}^{\mathscr{M}}$ uniquely solves those equations and yields unambiguously relations (2.23)–(2.25) for the respective quantities appearing in the original set of relations—any symmetries obeyed by the quantity $H_n(\Delta,\lambda;x,y)$ [see Eq. (4.6a)], hence, by the vectors $\mathbf{V}^t$ are necessarily reflected also in the Gram determinant and then also in the denominator function and G_q^t itself. [The fact that the procedure is invariant to the transformation $y\to\pm y+$ const allows one to interpret the significance of the transformations given here in terms of the fundamental operations (permutations of indices 1,2,3 conjugation, coupling order, etc.) that exist in the algebra of Wigner operators. This will be done in a future paper.]

The first symmetry we consider is called *index symmetry*. It refers specifically to the subscript, 1,2,3 indexing in the Hilbert space irrep label (m_{13},m_{23},m_{33}) and the shift label $(\Delta_1,\Delta_2,\Delta_3)$. More precisely, with each $P\in S_3$ (symmetric group of order 3) we associate an *index permutation*

$$P:(1,2,3)\mapsto(i_1,i_2,i_3), \tag{5.4a}$$

where (i_1,i_2,i_3) is a rearrangement of (1,2,3). The action of P is defined here in a very specific way: Namely, it is defined to be the simultaneous reindexing of the partial hooks $p=(p_{13},p_{23},p_{33})$ and of the shift pattern $\Delta=(\Delta_1,\Delta_2,\Delta_3)$ given by

$$\begin{aligned}P:\quad &p\to Pp=(p_{i_13},p_{i_23},p_{i_33}),\\ &\Delta\to P\Delta=(\Delta_{i_1},\Delta_{i_2},\Delta_{i_3}),\end{aligned} \tag{5.4b}$$

for P given by (5.4a). As used here, all other actions of P are to be derived from (5.4b). For example, the action of P on the 3-tuple $x=(x_1,x_2,x_3)=(p_{23}-p_{33},p_{33}-p_{13},p_{13}-p_{23})$ is

$$P:x\to Px=\delta_P x, \tag{5.4c}$$

where δ_P is the signature of P.

The first result for G_q^t refers to index permutations.

Lemma 5.1: The function $G_q^t(\Delta;x)$ has index symmetry:

$$G_q^t(P\Delta;Px)=G_q^t(\Delta;x), \tag{5.5}$$

each $P\in S_3$.

Proof: One easily verifies from the definition of $H_n(\Delta,\lambda;x,y)$ given by Eq. (4.6a) and the definition of $D_n(\Delta;x)$ given by Eq. (2.10) that

$$H_{Pn}(P\Delta,P\lambda;Px,y)=H_n(\Delta,\lambda;x,y). \tag{5.6a}$$

Observe that $\Delta\to P\Delta$ *implies* $n\to Pn$ and $\lambda\to P\lambda$. The invariance relation (5.6a) for H_n implies this same relation for the vector components $V_n^s(\Delta,\lambda;x)$, hence, for the Gram determinant (2.22c) as well:

$$A_t(P\Delta,P\lambda;Px)=A_t(\Delta,\lambda;x). \tag{5.6b}$$

We verify directly from the definitions (3.4a) and (3.5b) that also

$$L_s^t(P\Delta;Px)=L_s^t(\Delta;x),$$

which together with the definition of $G_q^t(\Delta;x)=G_q^t(\Delta,\lambda;x)$ given by Eq. (3.5a) and property (5.6b) imply the property of G_q^t stated in the lemma. ■

The index symmetry proved in Lemma 5.1 corresponds to cyclic row interchange in the array (5.1) for P an even permutation and to interchange of two rows followed by column 2 and column 3 interchange for P an odd permutation.

The next symmetry $G_q^t(\Delta;x)$ we consider corresponds to column 2 and column 3 interchange in array A, which we denote by C:

$$C: \Delta \rightarrow C\Delta = \Delta,$$
$$x \rightarrow Cx = (-x_1 - \Delta_2 + \Delta_3, \quad -x_2 - \Delta_3 + \Delta_1, \quad -x_3 - \Delta_1 + \Delta_2). \tag{5.7a}$$

We then have the following lemma.

Lemma 5.2: The function $G_q^t(\Delta;x)$ has the symmetry

$$G_q^t(C\Delta;Cx) = G_q^t(\Delta;x). \tag{5.7b}$$

Proof: The principal result required for the proof of this lemma has already been given in Sec. II—this symmetry originates from the possibility of coupling in two different orders to obtain the same projective function. The result proved in Sec. II [see Eq. (2.37)] is

$$\frac{1}{D^2(\Gamma_t; -x^f)} = \frac{\mathrm{Dim}(x)}{\mathrm{Dim}(x^f)} \frac{1}{D^2(\Gamma_t;x)}. \tag{5.8}$$

The additional relation we require is

$$L_t(\Delta; -x^f) = \frac{\mathrm{Dim}(x)}{\mathrm{Dim}(x^f)} L_t(\Delta;x), \tag{5.9}$$

which is verified directly from the definition of $L_t(\Delta;x)$ given by Eq. (3.4). Combining Eq. (5.8) and (5.9) with Eq. (3.3), we find

$$\frac{G_q^t(\Delta; -x^f)}{G_q^{t-1}(\Delta; -x^f)} = \frac{G_q^t(\Delta;x)}{G_q^{t-1}(\Delta;x)}. \tag{5.10}$$

Iteration of this result (in t) now yields the relation

$$G_q^t(\Delta; -x^f) = G_q^t(\Delta;x), \tag{5.11}$$

which proves the lemma. ■

The symmetries of $G_q^t(\Delta;x)$ proved in Lemmas (5.1) and (5.2) correspond to the transformation of variables (Δ,x) induced by row permutations and column 2–column 3 interchange in the array A. We still require the transpositional symmetry T [Eq. (5.3b)] of $G_q^t(\Delta;x)$ in order to prove the full determinantal symmetry. (The index transformations $P \in S_3$ together with C and T generate the full 72 determinantal transformations of A.)

Before turning to the proof of the invariance of $G_q^t(\Delta;x)$ under the transformation T, let us discuss several features of the proof.

Determinantal transformations of the nine "variables"

$$(\Delta_1, \Delta_2, \Delta_3, \Delta_2 + x_1, \Delta_3 + x_2, \Delta_1 + x_3, \Delta_3 - x_1, \Delta_1 - x_2, \Delta_2 - x_3) \tag{5.12}$$

places these variables on *equal footing*. The "operator variables" $(\Delta_1, \Delta_2, \Delta_3)$ are, however, clearly not treated on the same footing as the "Hilbert space variables" (x_1, x_2, x_3) in the formulation of Wigner coefficients as matrix elements of unit tensor operators. This is evident in the denominator function $D_n^2(\Delta;x)$ [see Eqs. (2.10a) and (2.8h)], where the Δ_i occur as *factorials* and not as polynomial forms [in contrast to the $(\Delta \pm x)$-type variables]. Thus, any discussion of the transpositional symmetry, which uses the function $H_n(\Delta,\lambda;x,y)$, in which (x_1,x_2,x_3) are regarded as indeterminates, necessarily involves extending the domain of all the variables (5.12) to nonintegral values. It is natural here to use the gamma function and the extension

$$z! \rightarrow \Gamma(z+1). \tag{5.13}$$

But a difficulty is encountered immediately with this procedure—it is highly nonunique! For example, if we use $(z)_a \rightarrow \Gamma(z+a)/\Gamma(z)$ to extend the function $L(\Delta;x)$ as defined first by Eq. (2.8h) and second by Eq. (2.8j), we find two results:

$$T: L(\Delta;x) \rightarrow L(\Delta;x), \tag{5.14a}$$

$$T: \mathrm{Dim}(x) L(\Delta;x) \rightarrow \mathrm{Dim}(x) L(\Delta;x) \times \frac{\sin \pi(\Delta_3 + 1) \sin \pi(x_3 + \Delta_1 + 1)}{\sin \pi(x_1 - \Delta_3) \sin \pi(x_2 - \Delta_1)}, \tag{5.14b}$$

where we have used $\Gamma(z)\Gamma(1-z) = \pi/\sin \pi z$ in obtaining the second result.

In consequence of the foregoing results, it might appear that the transpositional symmetry of the function $G_q^t(\Delta;x)$ is at best ambiguous. This, however, is not the case as the final form of $G_q^t(\Delta;x)$ given by Eqs. (3.6) and (2.10b) shows: Each term in that result—the determinant $\overline{A}_t(\Delta,\lambda;x)$, the factor $\mathrm{Dim}(x) L(qqq;x)$, and the factor $\Pi_{ijk}(-\Delta_i)_{s-1} \times (-x_i - \Delta_j)_{s-1}(x_i - \Delta_k)_{s-1}$—is a polynomial in the variables (5.12). [For factors depending only on (x_1, x_2, x_3), we write $x_i = (x_i + \Delta_j) - \Delta_j$ to place this property in evidence.] Since $G_q^t(\Delta;x)$ has this "polynomial structure" in the variables (5.12), it follows that each possible extension of the results of Sec. II, using the gamma function map (5.13), must, in fact, lead to the same final answer for $G_q^t(\Delta;x)$.

The preceding results serve to show that the extension, using the gamma function, of the results of Sec. II to nonintegral values of the nine variables in the matrix array A exists and leads to an unambiguous expression for $G_q^t(\Delta;x)$. The final form of the result must, however, be that given by Eqs. (3.6) and (2.10b)—which is the result obtained naturally without the necessity of extension.

The main point of the preceding discussion is this: We should examine the final form of $G_q^t(\Delta;x)$ as given by Eqs. (3.6) and (2.10b) in ascertaining the symmetry under the transformation T. But clearly both the polynomial $N_n(\Delta;x)$, hence, $\overline{A}_t(\Delta,\lambda;x)$, and the denominator terms in $G_q^t(\Delta;x)$ [Eq. (3.6c)] are invariant under T. We have thus proved the following lemma.

Lemma 5.3: The function $G_q^t(\Delta;x)$ has the transpositional symmetry T:

$$G_q^t(T\Delta,Tx) = G_q^t(\Delta;x). \tag{5.15}$$

The result implied by Lemmas (5.1)–(5.3) (as discussed earlier in this section) is summarized by Theorem 5.1.

Theorem 5.1: The function $G_q^t(\Delta;x)$ written in the form $G_q^t(A)$ is invariant under all determinantal transformations of the array A.

VI. POLYNOMIAL PROPERTIES OF G_q^t

The main result of this section is Theorem 6.1, which asserts that $G_q^t(\Delta;x)$ for each $t = 1,2,\ldots,\mathscr{M}$ is a polynomial in the nine variables

$$(\Delta_1, \Delta_2, \Delta_3, \Delta_2 + x_1, \Delta_3 + x_2, \Delta_1 + x_3, \Delta_3 - x_1, \Delta_1 - x_2, \Delta_2 - x_3) \tag{6.1}$$

up to an overall factor invariant under determinantal symmetry. The reduction formula given by Eqs. (4.15) and (4.16) plays a key role here: It shows that for $\lambda \neq (0,0,0)$ the function $G_q^t(\Delta;x)$ is a product of linear factors in the variables (6.1) times a function $G_{\hat{q}}^t(\hat{\Delta},\hat{x})$ in which $\hat{\lambda} = (0,0,0)$. Accordingly, it is no restriction to prove the asserted polynomial property for $G_q^t(\Delta;x)$ under the assumption that $\lambda = (0,0,0)$.

While making the above specialization to $\lambda = (0,0,0)$ in $G_q^t(\Delta;x)$ it is at the same time convenient to extend the domain of the parameters $\Delta = (\Delta_1, \Delta_2, \Delta_3)$ and to take $x = (x_1, x_2, x_3)$ to be an arbitrary point $x \in \mathbb{M}$: Thus, one notices in Eqs. (3.6) and (2.10b), which fully define $G_q^t(\Delta;x)$, that all quantities are well defined if we replace Δ_i by ξ_i, p by $\xi_1 + \xi_2 + \xi_3 - q$, and take $\xi = (\xi_1, \xi_2, \xi_3)$ to be an arbitrary point $\xi \in \mathbb{R}^3$.

For clarity, we repeat the resulting definition of $G_q^t(\xi;x)$, each $\xi \in \mathbb{R}^3$, each $x \in \mathbb{M}$:

$$G_q^t(\xi;x) = (-1)^{t(q+1)} \times \prod_{s=1}^{t} \frac{(s-1)!(q-s+1)!}{(\xi_1+\xi_2+\xi_3-q-s+3)_{s-1}} \times \prod_{s=1}^{t} \frac{1}{\Pi_{ijk}(-\xi_i)_{s-1}(-x_i-\xi_j)_{s-1}(x_i-\xi_k)_{s-1}} \times \frac{A_t(\xi;x)}{[\mathrm{Dim}(x)\Pi_{i=1}^3(x_i+1)_q(-x_i+1)_q]^t}, \tag{6.2a}$$

where (i) q is an arbitrary integer $0,1,\ldots,$ and for each q the values of t are $1,2,\ldots,q$; (ii) $A_t(\xi;x)$ is the $t \times t$ determinant with element in row r and column s given by

$$A^{rs}(\xi;x) = \sum_m F_m^r(x) F_m^s(x) N_m(\xi;x), \tag{6.2b}$$

$$N_m(\xi;x) = \frac{(x_1-m_2+m_3)(x_2-m_3+m_1)(x_3-m_1+m_2)}{2m_1!m_2!m_3!} \times \prod_{ijk}(-\xi_i)_{q-m_i}(-x_i-\xi_j)_{q-m_k}(x_i-\xi_k)_{q-m_j} \times (-x_i-q)_{q-m_k}(x_i-q)_{q-m_j}, \tag{6.2c}$$

where $m = (m_1, m_2, m_3)$ is any 3-tuple of non-negative integers that sum to q, and the summation in Eq. (6.2b) is over all such 3-tuples [see Remark (a) below]; and (iii) F_m^s is defined by the expansion

$$\prod_{ijk}\left(y+1+\frac{1}{3}(x_i-x_j)\right)_{m_k} = \sum_{s=1}^{q} y^{q+1-s} F_m^s(x). \tag{6.2d}$$

Remarks: (a) The summation over m in Eq. (6.2b) replaces the earlier summation over all $n \in \mathbb{D}(p,q,\Delta)$ (with $\lambda = 0$). This replacement of n by m is quite natural and moreover assures that in the specialization of (ξ_1,ξ_2,ξ_3) back to $(\Delta_1,\Delta_2,\Delta_3)$ we recover from Eqs. (6.2) exactly the earlier definition of $G_q^t(\Delta;x)$. This is true because $N_m(\Delta;x) = 0$ [in consequence of $(-\Delta_i)_{q-m_i} = 0$ for $m_i \leqslant q - \Delta_i - 1$] unless $m_i \geqslant q - \Delta_i$, each $i = 1,2,3$, which implies also that $m_j + m_k \geqslant 2q - \Delta_j - \Delta_k$ $(j \neq k)$, and therefore $m_i \leqslant p - \Delta_i$, each $i = 1,2,3$. Since also $0 \leqslant m_i \leqslant q$, the only nonzero terms in (6.2b) occur for $m \in \mathbb{D}(p,q,\Delta)$ (with $\lambda = 0$) when $\xi = \Delta$.

(b) We emphasize again that the functions $G_q^t(\xi;x)$ defined by Eqs. (6.2) determine fully the general $G_q^t(\Delta;x)$ function with $\lambda \neq (0,0,0)$, since we recover the general result by setting $q = \hat{q}$, $\xi = \hat{\Delta}$, $x = \hat{x}$ in $G_q^t(\xi;x)$ and then multiplying by the factors in the reduction formula (4.15).

(c) If we change the notation and write

$$G_q^t(\xi;x) = G_q^t\begin{pmatrix} \xi_1-t+1 & \xi_2-t+1+x_1 & \xi_3-t+1-x_1 \\ \xi_2-t+1 & \xi_3-t+1+x_2 & \xi_1-t+1-x_2 \\ \xi_3-t+1 & \xi_1-t+1+x_3 & \xi_2-t+1-x_3 \end{pmatrix}, \tag{6.3}$$

then it follows from the results of Sec. V that $G_q^t(\xi;x)$ has determinantal symmetry in the matrix array.

Let us turn next to the proof of the polynomial properties of $G_q^t(\xi;x)$. The principal result we wish to prove is the next theorem.

Theorem 6.1: The function $G_q^t(\xi;x)$ has the form

$$G_q^t(\xi;x) = \left[\prod_{s=1}^{t} \frac{(s-1!(q-s+1)!}{(\xi_1+\xi_2+\xi_3-q-s+3)_{s-1}}\right] g_q^t(\xi;x), \tag{6.4a}$$

where $g_q^t(\xi;x)$ is a polynomial in (x_1, x_2, x_3) of total degree

$$2t(q-t+1), \tag{6.4b}$$

and in (ξ_1, ξ_2, ξ_3) of total degree

$$2t(q-t+1) + \tfrac{1}{2}t(t-1). \tag{6.4c}$$

Remarks: (a) The total degree of a monomial $x^a y^b z^c$ is defined to be $a+b+c$. Note that the total degree of a single monomial term is preserved even when the variables are *linearly* dependent (over $\mathbb{R}$); for example, the total degree in (x,y) of $x^a y^b (x+y)^c$ is still $a+b+c$. This result need, of course, not be true for a sum of monomials. Accordingly, in asserting that the total degree in (x_1, x_2, x_3) of $g_q^t(\xi;x)$ is $2t(q-t+1)$, we mean that this is the total degree in any two of the independent variables in (x_1, x_2, x_3) after eliminating the third dependent one.

(b) We have not attempted to write $G_q^t(\xi;x)$ in a form that exhibits explicitly the (proved) determinantal symmetry (this is nontrivial), since our main point here is to establish the polynomial properties themselves.

(c) The factor $\sigma = \xi_1 + \xi_2 + \xi_3$ (and, hence, any function of σ) is invariant under all determinantal transformations; accordingly, it is more difficult to give a general argument that the factor $\Pi_{s=1}^t (\sigma - q - s + 3)_{s-1}$, which is of total degree $t(t-1)/2$ in (ξ_1, ξ_2, ξ_3), should also divide $g_q^t(\xi;x)$, leaving then a function $G_q^t(\xi;x)$ that is polynomial of total degree $2t(q-t+1)$ in both (ξ_1, ξ_2, ξ_3) and (x_1, x_2, x_3). We have been unable to prove this result, but believe it to be true.

We demonstrate the above theorem in three steps, using directly the definition of $G_q^t(\xi;x)$ given by Eqs. (6.2): The first step of the proof shows that certain zeros of the (polynomial) determinant $A_t(\xi;x)$ originate from each element of the determinant (Lemma 6.1); the second step establishes the existence of further zeros of $A_t(\xi;x)$ by appealing to the finiteness of the functions $G_q^t(\xi;x)$ implied by null space properties

(Lemma 6.2); finally, in the third step we deduce the asserted degree properties (Lemma 6.3). Together these three lemmas imply Theorem 6.1.

Referring to the definition of $G^t_q(\xi;x)$ given by Eqs. (6.2), we now prove Lemma 6.1.

Lemma 6.1: The factor

$$\left[\mathrm{Dim}(x)\prod_i (x_i+1)_q(-x_i+1)_q\right]^t \tag{6.5}$$

divides the determinant $A_t(\xi;x)$.

Proof: We give the proof of this lemma by showing that the factor

$$\mathrm{Dim}(x)\prod_i (x_i+1)_q(-x_i+1)_q \tag{6.6}$$

divides each element $A^{rs}(\xi;x)$ $(r,s=1,2,...,t)$ of the determinant $A_t(\xi;x)$. Since the roots of the factor (6.6) are those $x=(x_1,x_2,x_3)\in\mathbb{L}$ such that also

$$x_i\in\mathbb{Z}_q\equiv\{-q,-q+1,...,q\}, \tag{6.7}$$

the desired result is true if it can be proven that [see Eq. (6.2b)]

$$\sum_m F^r_m(x)F^s_m(x)N_m(\xi;x)=0, \tag{6.8}$$

each $x\in\mathbb{L}$, $x_i\in\mathbb{Z}_q$.

To prove Eq. (6.8), we first observe from the definition (6.2c) of $N_m(\xi;x)$ that

$$N_m(\xi;x)=0, \tag{6.9a}$$

for each x_i in the set

$$\{-q,-q+1,...,-m_k-1;m_j+1,m_j+2,...,q\},$$
$$(ijk)\text{ cyclic.} \tag{6.9b}$$

[These zeros come from the factors $(-x_i-q)_{q-m_j}$ and $(x_i-q)_{q-m_j}$ in $N_m(\xi;x)$.] It follows from this result that a given term (m_1,m_2,m_3) in the sum (6.8) is zero *unless*

$$x_i\in\{-m_k,-m_k+1,...,m_j\},\quad (ijk)\text{ cyclic.} \tag{6.9c}$$

We next show that the sum (6.8) is also zero for each x_i in the set (6.9c) because of *pairwise* cancellation of terms.

Thus, for $x_3=a_3$ and

$$a_3\in\{-m_2,-m_2+1,...,m_1\}, \tag{6.10}$$

we now demonstrate that the (m_1,m_2,m_3) and (m_2+a_3, m_1-a_3,m_3) terms in the sum (6.8) cancel exactly. The general result for x_i then follows because of index symmetry:

$$A^{rs}(P\xi;Px)=A^{rs}(\xi;x).$$

The two results required for this proof are

$$F^r_{(m_1,m_2,m_3)}(x_1,x_2,a_3)=F^r_{(m_2+a_3,m_1-a_3,m_3)}(x_1,x_2,a_3), \tag{6.11a}$$

$$N_{(m_1,m_2,m_3)}(x_1,x_2,a_3)=-N_{(m_2+a_3,m_1-a_3,m_3)}(x_1,x_2,a_3), \tag{6.11b}$$

each $a_3\in\{-m_2,-m_2+1,...,m_1\}$. The first property (6.11a) is easily proved from Eq. (6.2d) (use of the relation $x_1+x_2+a_3=0$ is required). The second relation (6.11b) is somewhat more tedious to prove. The factorial relations needed are

$$\begin{aligned}&(x_1-\xi_3)_{q-m_2}(-x_2-\xi_3)_{q-m_1}\\&\quad=(x_1-\xi_3)_{q-m_1+a_3}(-x_2-\xi_3)_{q-m_2-a_3},\end{aligned} \tag{6.12a}$$

$$\begin{aligned}&(-\xi_1)_{q-m_1}(-a_3-\xi_1)_{q-m_2}(-\xi_2)_{q-m_2}(a_3-\xi_2)_{q-m_1}\\&\quad=(-\xi_1)_{q-m_2-a_3}(-a_3-\xi_1)_{q-m_1+a_3}\\&\quad\times(-\xi_2)_{q-m_1+a_3}(a_3-\xi_2)_{q-m_2-a_3}\end{aligned} \tag{6.12b}$$

$$\begin{aligned}&(-a_3-q)_{q-m_2}(a_3-q)_{q-m_1}/m_1!m_2!\\&\quad=(-a_3-q)_{q-m_1+a_3}(a_3-q)_{q-m_2-a_3}\\&\quad\times 1/(m_2+a_3)!(m_1-a_3)!.\end{aligned} \tag{6.12c}$$

[Relation (6.12a) is an application of the general identity $(z)_\alpha(z+a)_\beta=(z)_{\beta+a}(z+a)_{\alpha-a}$ for $\alpha,\beta,\beta+a,\alpha-a$ nonnegative integers (using also $-x_2=x_1+a_3$); relation (6.12b) uses the preceding general identity twice; and relation (6.12c) is the special case $\xi_1=\xi_2=q$ of (6.12b).] Finally, the dimension factor in $N_m(\xi;x)$ reverses sign under the transformation $m_1\mapsto m_2+x_3$, $m_2\mapsto m_1-x_3$, $m_3\mapsto m_3$. These results prove Eq. (6.11b) and complete the proof of Lemma 6.1. ■

The result of Lemma 6.1 is that P^t_q, defined by

$$P^t_q(\xi;x)=\frac{A_t(\xi;x)}{[\mathrm{Dim}(x)\Pi^3_{i=1}(x_i+1)_q(-x_i+1)_q]^t}, \tag{6.13}$$

is a polynomial in the variables $\xi\in\mathbb{R}^3$, $x\in\mathbb{M}$. (It clearly also has determinantal symmetry.)

The second result we require for the proof of Theorem 6.1 is Lemma 6.2.

Lemma 6.2: The factor

$$\prod_{s=1}^t\prod_{ijk}(-\xi_i)_{s-1}(-x_i-\xi_j)_{s-1}(x_i-\xi_k)_{s-1} \tag{6.14}$$

divides the polynomial $P^t_q(\xi;x)$.

Proof: We have been unable to give a direct proof of this result, based on Eqs. (6.2). Its validity may, however, be established as follows: Referring to the null space diagram (Fig. 3) for the operator Γ_t, we see that the denominator function $D^2(\Gamma_t;x)$ $(\lambda=0)$ can have no lines of zeros of the form $x_1=\Delta_3$, $x_1=\Delta_3-1,...,x_1=\Delta_3-t+2$, since $1/D^2(\Gamma_t;x)$ must be finite in this region of the Möbius plane, indeed, positive if we also take $x_3\geqslant\Delta_2-t+2$, and $x_2\leqslant-\Delta_2-\Delta_3-2t+2$. It follows by induction on t [starting with $t=1$ and the fact that $G^1_q(\Delta;x)$ is polynomial, hence, finite for all finite Δ and x], using the form of $D^2(\Gamma_t;x)$ given by Eq. (3.3), that $G^t_q(\Delta;x)$ must be finite for each $x_1=\Delta_3$, $\Delta_3-1,...,\Delta_3-t+2$. This result implies that $G^t_q(\xi;x)$ given by Eqs. (6.2) must be finite for each

$$x_1\in\{\xi_3-t+2,\xi_3-t+3,...,\xi_3\}. \tag{6.15}$$

But the factor (6.14) is zero for each x_1 in the set (6.15) in consequence of the factors $(x_1-\xi_3)_{s-1}$, $s=2,3,...,t$. Thus, $G^t_q(\xi;x)$ is undefined (infinite) at each point in the set (6.15) unless $\Pi^t_{s=1}(x_1-\xi_3)_{s-1}$ divides $P^t_q(\xi;x)$. This result and the determinantal symmetry of the factor (6.14) and of $P^t_q(\xi;x)$ imply the statement in the lemma. ■

Together Lemmas (6.1) and (6.2) imply that the factor $g^t_q(\xi;x)$ in Eq. (6.4a) is a polynomial in (x_1,x_2,x_3) and in (ξ_1,ξ_2,ξ_3). The final result required for the proof of Theorem 6.1 is Lemma 6.3.

Lemma 6.3: The total degree of $g^t_q(\xi;x)$ in (x_1,x_2,x_3) and in (ξ_1,ξ_2,ξ_3) is $2t(q-t+1)$ and $2t(q-t+1)+\frac{1}{2}t(t-1)$, respectively.

Proof: In determining the total degree in x, we may put

$$x_1 = u, \quad x_2 = -2u, \quad x_3 = u. \qquad (6.16)$$

The total degree of $g_q^t(\xi;x)$ is then $2t(q-t+1)$ if we can deomonstrate that

$$g_q^t(\xi;u,-2u,u) = H_q^t(\xi)u^{2t(q-t+1)} + \text{(lowest-order terms in } u\text{)}, \qquad (6.17)$$

where $H_q^t(\xi)$ is nonzero. We prove Eq. (6.17) by direct examination of the various factors appearing in Eqs. (6.2). We list below the term of highest degree in u originating from these various factors (indicated by the arrow):

$$N_m(\xi;x) \mapsto -u^{8q+3}2^{2q+2m_2}\prod_{i=1}^{3}\frac{(-\xi_i)_{q-m_i}}{m_i!}, \qquad (6.18a)$$

$$F_m^r(x) \to a_m^r u^{r-1}, \qquad (6.18b)$$

with

$$a_m^r = \sum_{\substack{\alpha,\beta \\ \alpha+\beta=r-1}} (-1)^\alpha \binom{m_1}{\alpha}\binom{m_3}{\beta}, \qquad (6.18c)$$

$$\mathrm{Dim}(x)\prod_{i=1}^{3}(x_i+1)_q(-x_i+1)_q \to (-1)^q 2^{2q} u^{6q+3}, \qquad (6.18d)$$

$$\frac{A^{rs}(\xi;x)}{[\mathrm{Dim}(x)\Pi_i(x_i+1)_q(-x_i+1)_q]} \to (-1)^{q+1}u^{2q+r+s-2}h^{rs}(\xi), \qquad (6.18e)$$

with

$$h^{rs}(\xi) = \sum_m a_m^r a_m^s 2^{2m_2}\prod_{i=1}^{3}\frac{(-\xi_i)_{q-m_i}}{m_i!}, \qquad (6.18f)$$

$$P_q^t(\xi;x) \to (-1)^{t(q+1)}u^{2qt+t(t-1)}h_t(\xi), \qquad (6.18g)$$

where $h_t(\xi)$ is the $t\times t$ determinant with elements $h^{rs}(\xi)$ [see Eq. (6.13) for the definition of P_q^t];

$$\prod_{s=1}^{t}\prod_{ijk}(-x_i-\xi_j)_{s-1}(x_i-\xi_k)_{s-1} \to (-1)^{(t/2)(t-1)}2^{t(t-1)}u^{3t(t-1)}. \qquad (6.18h)$$

Combining these relations with Eqs. (6.2) yields the result (6.17) with

$$H_q^t(\xi) = \frac{(-1)^{(1/2)t(t-1)}2^{-t(t-1)}}{\Pi_{s=1}^t \Pi_{i=1}^3(-\xi_i)_{s-1}} h_t(\xi). \qquad (6.19)$$

The only point requiring further discussion in the derivation of Eq. (6.19) is the determination of the coefficient a_m^r in Eq. (6.18c). This may be obtained from Eq. (6.2d) by the following procedure: Set

$$f_m^r(u) = F_m^r(u,-2u,u) = a_m^r u^{r-1} + \text{(lower-order terms)}. \qquad (6.20a)$$

Then

$$(y+1-u)_{m_1}(y+1)_{m_2}(y+1+u)_{m_3} = \sum_{r=1}^{q+1} y^{q+1-r}f_m^r(u). \qquad (6.20b)$$

We next set $u = yw$ in this relation and find that the coefficients of y^q (recall that $m_1+m_2+m_3=q$) on the left- and right-hand sides of Eq. (6.20b) are $(1-w)^{m_1}(1+w)^{m_3}$ and $\Sigma_r a_m^r w^{r-1}$, respectively; that is,

$$(1-w)^{m_1}(1+w)^{m_3} = \sum_{r=1}^{q+1} a_m^r w^{r-1}, \qquad (6.20c)$$

which yields

$$a_m^r = \sum_{\substack{\alpha,\beta \\ \alpha+\beta=r-1}} (-1)^\alpha\binom{m_1}{\alpha}\binom{m_3}{\beta}.$$

It remains still to show that the function (of ξ) $H_q^t(\xi)$ defined by Eq. (6.19) is not identically zero. For this it is sufficient to take

$$\xi_1 = \xi_2 = \xi_3 = v \qquad (6.21)$$

and show that the highest degree term in v in $h_t(v,v,v)$ is nonzero. Using

$$h^{rs}(v,v,v) = v^{2q}\beta^{rs} + \text{(lower-order terms in } v\text{)}, \qquad (6.22a)$$

we obtain

$$h_t(v,v,v) = v^{2tq}B_t + \text{(lower-order terms in } v\text{)}, \qquad (6.22b)$$

where B_t is the $t\times t$ determinant with elements

$$\beta^{rs} = \sum_m \frac{a_m^r a_m^s 2^{2m_2}}{m_1!m_2!m_3!}. \qquad (6.23a)$$

Thus, we must show that

$$B_t \neq 0. \qquad (6.23b)$$

Since we can write β^{rs} as the inner product

$$\beta^{rs} = (b^r, b^s) \qquad (6.23c)$$

with

$$b_m^r \equiv 2^{m_2}a_m^r/[m_1!m_2!m_3!]^{1/2}, \qquad (6.23d)$$

it follows that B_t is a *Gram determinant*. It is necessarily nonzero if the vectors

$$\mathbf{b}^1, \mathbf{b}^2, \ldots \mathbf{b}^t \qquad (6.24a)$$

are linearly independent; that is, if

$$\sum_{r=1}^{t} c_r \mathbf{b}^r = \mathbf{0} \qquad (6.24b)$$

implies $c_1 = c_2 = \cdots = c_t = 0$, then $B_t \neq 0$. Equation (6.24b) may be rewritten in terms of components as

$$\sum_{r=1}^{t} c_r a_m^r = 0. \qquad (6.24c)$$

Choosing the $t+1$ components $m = (0,q,0), (0,q-1,1),\ldots, (0,q-t+1,t-1)$ of the vector $\mathbf{a}^r$, this last equation reduces to

$$\sum_{r=1}^{t} c_r\binom{s-1}{r-1} = 0, \qquad (6.24d)$$

each $s = 1,2,\ldots,t$, which, in turn, implies the desired result, $c_1 = c_2 = \cdots = c_t = 0$; that is, the t vectors (6.24a) are linearly independent; hence, $B_t \neq 0$.

These results prove that $H_q^t(\xi)\neq 0$, and complete the proof that the total degree of $g_q^t(\xi;x)$ in (x_1,x_2,x_3) is exactly $2t(q-t+1)$.

We can go one step further with Eq. (6.19), since we have proved that $H_q^t(\xi)$ is a polynomial in ξ. Using the highest degree term,

$$\prod_{s=1}^{t}\prod_{i=1}^{3}(-\xi_i)_{s-1} \to (-1)^{(1/2)t(t-1)}v^{(3/2)t(t-1)},$$

from the denominator of (6.19) for the special variables (6.21) and combining this result with Eq. (6.22b) yields

$$H_q^t(v,v,v) = \frac{(-1)^{(1/2)t(t-1)}}{2^{t(t-1)}} B_t v^{2t(q-t+1)+(1/2)t(t-1)}$$
$$+ \text{(lower-order terms in } v).$$

This result proves that the total degree in ξ of $H_q^t(\xi)$, hence, of $g_q^t(\xi;x)$ is at *least* $2t(q-t+1)+\frac{1}{2}t(t-1)$.

Starting again with Eqs. (6.2), but this time leaving x general and choosing the special values (6.21) for ξ everywhere in the numerator and denominator factors, we can by selecting the various highest-degree terms in v and adding algebraically (plus for numerator factors, minus for denominator factors) easily establish that the total degree of $g_q^t(\xi;x)$ in ξ is at *most* $2t(q-t+1)+\frac{1}{2}t(t-1)$. From this result and the one above, we conclude that the total degree of $g_q^t(\xi;x)$ in ξ is *exactly* $2t(q-t+1)+\frac{1}{2}t(t-1)$; this result completes the proof of the lemma. ■

Remark: If the factor $\Pi_{s=1}^t(\xi_1+\xi_2+\xi_3-q-s+3)_{s-1}$ divides $g_q^t(\xi;x)$ [see Eq. (6.4a)], then the preceding degree considerations (Lemma 6.3) show that the total degree of $G_q^t(\xi;x)$ in each of ξ and x is $2t(q-t+1)$.

VII. CONVENTIONS IN DEFINING G_q^t

The denominator functions $D^2(\Gamma_t;x)$ determine the functions $G_q^t(\Delta;x)$ only up to an arbitrary multiplicative constant [see the paragraph containing Eq. (3.4b)]. Thus, if we were to change the definitions of $G_q^t(\Delta;x)$ and $L_t(\Delta;x)$ in Eq. (3.3) to

$$G_q^t(\Delta;x) \to \overline{G}_q^t(\Delta;x) = a_t(p,q,\Delta)G_q^t(\Delta;x), \qquad (7.1a)$$

$$L_t(\Delta;x) \to \overline{L}_q^t(\Delta;x) = \frac{a_t(p,q,\Delta)}{a_{t-1}(p,q,\Delta)} L_t(\Delta;x), \qquad (7.1b)$$

then we would obtain the same denominator function $D^2(\Gamma_t;x)$. Moreover, this is the most general transformation we can make that preserves the linear factors in the x_i occurring in $L_t(\Delta,x)$—a condition we require from null space.

The freedom to choose the constant $a_t(p,q,\Delta)$ in Eq. (7.1) is quite important: First of all, this freedom allows us to actually realize the transpositional symmetry for the functions $G_q^t(\Delta;x)$. We have proved that $G_q^t(\Delta;x)$, as already defined, does have transpositional symmetry, but it will now be recognized that this fact depended on the choice of the multiplicative constants $\alpha_1!\alpha_2!\alpha_3!$ appearing in the definition of $L(\alpha;x)$ [see Eqs. (2.8h), (3.2d), and (3.4a)]. A different choice would lead to "determinantal symmetry of $G_q^t(\Delta;x)$ up to a multiplicative factor" except for row permutations and column 2–column 3 interchange. In order not to destroy the full transpositional symmetry, we accordingly require that $a_t(p,q,\Delta)$ should be independent of the Δ_i for each specified p. The remaining properties of $a_t(p,q)$ (dropping now Δ) are a matter of *convention*. The particular conventions we choose to fix uniquely $a_t(p,q)$ are the following: (i) $a_0(p,q)=1$; (ii) the reduction formula as given by Eqs. (4.15) and (4.16) should hold for all p,q,Δ; and (iii) $\overline{G}_q^{q+1}(\Delta;x)=1$ for all q and Δ (note that by Theorem 6.1 this function is independent of the x_i). It is the purpose of the present section to prove that the two conditions (ii) and (iii) uniquely determine

$$a_t(p,q)=1, \qquad (7.2)$$

for all $t \geq 1, p,q$.

First, we observe that the reduction formula (4.15) applied both to $\overline{G}_q^t(\Delta;x)=a_t(p,q)G_q^t(\Delta;x)$ and $G_q^t(\Delta;x)$ itself yields

$$a_t(p,q)=a_t(p,q-\lambda_1-\lambda_2-\lambda_3). \qquad (7.3)$$

Since this relation is to hold for all t ($=1,2,\ldots,\mathscr{M}$), all p, all q, and all Δ_i [satisfying the conditions given in Eqs. (2.8a)], we conclude that $a_t(p,q)$ must be independent of q (otherwise it will depend in denumerably many instances on the Δ_i). Accordingly, we may choose q to be any convenient value that does not restrict t and p, say, $q=t-1$ and consider that $\Delta_i \leq t-1$ (since the constant does not depend on the Δ_i). Thus, we arrive at

$$a_t(p,q)=a_t(p,t-1), \qquad (7.4)$$

for all $t=1,2,\ldots$.

Second, by proving that $G_{t-1}^t(\Delta;x)=1$ [and therefore that $a_t(p,t-1)=1$ is implied by the convention $\overline{G}_{t-1}^t(\Delta;x)=1$], we obtain the desired result, Eq. (7.2), from relation (7.4) above. Equivalently (since $t=1,2,\ldots$), we establish relation (7.2) by proving the next lemma.

Lemma 7.1: The function $G_q^{q+1}(\xi;x)$ is unity; that is,

$$G_q^{q+1}(\xi;x)=1, \qquad (7.5)$$

each $q=0,1,\ldots$.

Proof: The total degree of $G_q^{q+1}(\xi;x)$ in x is zero; hence, from Eqs. (6.4a), (6.17), and (6.19), we find that the property (7.5) is implied by the identity

$$\det(h) \equiv \det\begin{bmatrix} h^{1,1} & \cdots & h^{1,q+1} \\ \vdots & & \vdots \\ h^{q+1,1} & \cdots & h^{q+1,q+1}\end{bmatrix}$$
$$=(1)^{(1/2)q(q+1)}2^{q(q+1)}$$
$$\times \prod_{s=1}^{q+1} \frac{(\xi_1+\xi_2+\xi_3-s-q+3)_{s-1}\Pi_{i=1}^3(-\xi_i)_{s-1}}{(s-1)!(s-1)!}, \qquad (7.6a)$$

where [see Eqs. (6.18c) and (6.18f)]

$$h^{r,s}=h^{rs}(\xi)=\sum_m a_m^r a_m^s 2^{2m_2}\prod_{i=1}^3 \frac{(-\xi_i)_{q-m_i}}{m_i!}, \qquad (7.6b)$$

$$a_m^r = \sum_{\substack{\alpha,\beta \\ \alpha+\beta=r-1}} (-1)^\alpha \binom{m_1}{\alpha}\binom{m_3}{\beta}. \qquad (7.6c)$$

A direct proof of Eq. (7.6a) is quite difficult; we will prove it indirectly by making use of the fact that $\Pi_{s=1}^{q+1}\Pi_{i=1}^3(-\xi_i)_{s-1}$ must divide the determinant (in consequence of the transpositional symmetry proved in Sec. V). Accordingly, we need only verify the numerical factors and the term $\Pi_s(\xi_1+\xi_2+\xi_3-s-q+3)_{s-1}$ in the right-hand side of (7.6a); for this, it is sufficient to consider a special choice of the variables (ξ_1,ξ_2,ξ_3) such as $\xi_1=\xi_1$, $\xi_2=-1$, $\xi_3=q$.

Before implementing this procedure let us first transform the determinant in Eqs. (7.6) by

$$a_m^r \to \alpha_m^r = \sum_{s=r}^{q+1} a_m^s \binom{s-1}{r-1}. \qquad (7.7a)$$

This transformation leaves the determinant invariant, since it corresponds to allowed row and column operations. The coefficients α^r_m and the elements of the transformed determinant may now be obtained in explicit form not involving summation. Thus, noticing that the transformation $w \to w+1$ in Eq. (6.20c) yields

$$(-w)^{m_1}(2+w)^{m_3} = \sum_{r=1}^{q+1} \alpha^r_m w^{r-1},$$

we find the coefficients α^r_m defined by Eq. (7.7a) may also be written as

$$\alpha^r_m = (-1)^{m_1} 2^{m_1+m_3-r+1} \binom{m_3}{r-m_1-1}. \qquad (7.7b)$$

[Thus, the rather complicated summation in the right-hand side of relation (7.7a) can actually be summed.]

In the next step we make the substitution $a^r_m \to \alpha^r_m$ of coefficients in Eq. (7.6b), set $\xi_3 = q$, $m_3 = q - m_1 - m_2$, $(-q)_{m_1+m_2} = (-1)^{m_1+m_2} q!/(m_1+m_2)!$, and rearrange factors thus obtaining the elements of the transformed determinant as

$$\begin{aligned} h^{rs} \to & (-1)^q 2^{2q+2-r-s} \binom{q}{r-1} (-\xi_1)_{q-s+1} (-\xi_2)_{s-1} \\ & \times \sum_{m_1} \binom{r-1}{m_1} \binom{\xi_1 - q + s - 1}{s - m_1 - 1} \\ & \times \sum_{m_2} \binom{q-r+1}{m_2} \binom{\xi_2 - s + 1}{q - m_2 - s + 1}. \end{aligned} \qquad (7.8a)$$

The two summations in this result may be summed using the binomial (function) sum rule. Carrying this out, we thus find that the determinant $h_t(\xi_1,\xi_2,q)$ is reduced to one with elements

$$\begin{aligned} h^{rs} \to \bar{h}^{rs} \equiv & \frac{2^{2q+2-r-s}}{q!} \binom{q}{r-1} \binom{q}{s-1} \\ & \times (\xi_1 - q + r)_{s-1} (\xi_1 - q + s)_{q-s+1} \\ & \times (\xi_2 - s + 2)_{s-1} (\xi_2 - r + 2)_{q-s+1}. \end{aligned} \qquad (7.8b)$$

We next remove the factor

$$2^{q-s+1} \binom{q}{s-1} (\xi_1 - q + s)_{q-s+1} (\xi_2 - s + 2)_{s-1}$$

from column s (each $s = 1,2,...,q+1$) and the factor

$$2^{q-r+2} \binom{q}{r-1} (q!)$$

from row r (each $r = 1,2,...,q+1$) thus bringing $\det(\bar{h})$ to the form

$$\begin{aligned} \det(\bar{h}) = & 2^{q(q+1)} \det(c) \prod_{s=1}^{q+1} \binom{q}{s-1} (-\xi_1)_{s-1} (-\xi_2)_{s-1} \\ & \times [(s-1)!(s-1)!]^{-1}, \end{aligned} \qquad (7.9a)$$

where $\det(c)$ is the $(q+1) \times (q+1)$ determinant with elements

$$c^{rs} = (\xi_1 - q + r)_{s-1} (\xi_2 - r + 2)_{q-s+1}. \qquad (7.9b)$$

For the proof of Eq. (7.6a), we must demonstrate equality between the right-hand sides of Eqs. (7.9a) and (7.6a), where we put $\xi_3 = q$ in the latter; that is, we must still prove

$$\det(c) = \prod_{s=1}^{q+1} (s-1)!(\xi_1 + \xi_2 - s + 3)_{s-1}. \qquad (7.10)$$

To prove Eq. (7.10) we again appeal to the fact that $\det(c)$ can only be a function of $\xi_1 + \xi_2$ [we are considering the special case $\xi_3 = q$ of Eq. (7.6a)]. Thus, we may set $\xi_2 = -1$ in Eqs. (7.9b) and (7.10). In this case $c^{rs} = 0$ for each $r = 1,2,...,q+1-s$ so that up to sign the determinant is the product of the elements along the minor diagonal:

$$\det(c) = (-1)^{(1/2)q(q+5)} \prod_{r=1}^{q+1} c^{r,q+2-r},$$

which yields Eq. (7.10) for $\xi_2 = -1$, thus completing the proof of the lemma. ■

The identity (7.6a) is an example of the type of relations that arise in the present investigation. It is itself a special case of a more general relation given below that we conjecture to be true. We present this conjecture not only because of its intrinsic interest, but also because of its importance in our program of investigating the properties of the polynomials G^t_q.

Conjecture: The identity,

$$h_t(\xi) = \det \begin{bmatrix} h^{1,1}(\xi) & \cdots & h^{1,t}(\xi) \\ \vdots & & \vdots \\ h^{t,1}(\xi) & \cdots & h^{t,t}(\xi) \end{bmatrix} \qquad (7.11)$$

$$\begin{aligned} = & (-4)^{t(t-1)/2} \prod_{s=1}^{t} \{ [(-\xi_i)_{s-1}/(s-1)!(s-1)!] \\ & \times (\xi_1 + \xi_2 + \xi_3 - q - s + 3)_{s-1} \} \\ & \times \sum_{\lambda\mu\nu} \left\{ \frac{4^{\mu_1 + \cdots + \mu_t} h(\lambda\mu\nu)}{M(\lambda) M(\mu) M(\nu)} \right. \\ & \times \prod_{s=1}^{t} (-\xi_1 + s - t)_{q-t+1-\lambda_s} \\ & \times (-\xi_2 + s - t)_{q-t+1-\mu_s} \\ & \left. \times (-\xi_3 + s - t)_{q-t+1-\nu_s} \right\}. \end{aligned}$$

is true, where the quantities in this relation have the following definitions.

(i) The elements $h^{rs}(\xi)$ of the determinant are those defined in Eqs. (6.18c) and (6.18f).

(ii) $\lambda = [\lambda_1, \lambda_2, ..., \lambda_t]$ denotes an irrep label of the unitary group U(t) with $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_t \geqslant 0$, each λ_i a nonnegative integer; λ may also be regarded as the shape of a Young frame, $Y(\lambda)$. The symbols μ, ν, ...denote irrep labels of the same type as λ.

(iii) $h(\lambda\mu\nu)$ denotes the number of times irrep $[q-t+1,...,q-t+1]$ ($q-t+1$ repeated t times) is contained in the direct product $\lambda \times \mu \times \nu$, and is defined to be zero if $[q-t+1,...,q-t+1] \notin \lambda \times \mu \times \nu$.

(iv) $M(\lambda)$ is the measure of the Young frame $Y(\lambda)$ and has the definition

$$M(\lambda) = \frac{\prod_{s=1}^{t} (\lambda_s - s + t)!}{\prod_{1, r<s}^{t} (\lambda_r - \lambda_s + s - r)},$$

and $M(\mu)$, $M(\nu)$,... have corresponding definitions. [For $t = 1$, we define $M(\lambda_1) = \lambda_1!$.]

The importance of the relation (7.11) for the present work is that its validity establishes the following result for the functions G^t_q [combine the above result with Eqs. (6.4a), (6.17), and (6.19)]:

$$G^t_q(\xi;u,-2u,u)=\prod_{s=1}^{t}\frac{(q-s+1)!}{(s-1)!}u^{2q(q-t+1)}$$
$$\times\sum_{\lambda\mu\nu}\frac{4^{\mu_1+\cdots+\mu_t}h(\lambda\mu\nu)}{M(\lambda)M(\mu)M(\nu)}$$
$$\times\prod_{s=1}^{t}(-\xi_1+s-t)_{q-t+1-\lambda_s}$$
$$\times(-\xi_2+s-t)_{q-t+1-\mu_s}$$
$$\times(-\xi_3+s-t)_{q-t+1-\nu_s}$$
$$+\text{(lower-order terms in } u). \tag{7.12}$$

Remarks: (a) We have proved relation (7.11) in two cases: $t=1$ and $t=q+1$. For $t=1$, it reduces to Eq. (7.6b) (since $a^1_m=1$); for $t=q+1$, to Eq. (7.6a).

(b) The proof of relation (7.11), hence, of Eq. (7.12) giving the leading term in $G^t_q(\xi;u,-2u,u)$ would support our belief that the $\xi_1+\xi_2+\xi_3$ factors in Eq. (6.4a) divide $g^t_q(\xi;x)$.

VIII. ZEROS OF G^t_q

We have proved in Sec. VI that the function $G^t_q(\xi;x)$ defined by Eqs. (6.2) is a polynomial of degree $2t(q-t+1)$ in (x_1,x_2,x_3) for $t=1,...,q+1$. The existence of the characteristic null space $\mathscr{N}_t$ of the canonical Wigner operator

$$\left\langle\begin{matrix} & \Gamma_t & \\ p & q & 0 \\ & \cdot & \end{matrix}\right\rangle$$

implies that the polynomial $G^t_q(\Delta;x)$ possesses zeros. In the present section we investigate the occurrence of these zeros from two viewpoints: In Subsection A we establish the existence and location of zeros of G^t_q using properties of null space and a recurrence relation that must be satisfied by the multiplicity function $m_t(x)$ (defined below) for a zero; that is, the results obtained in Subsection A make no explicit use of the properties of the polynomials themselves. In Subsection B we use the symmetries of G^t_q, the reduction formula (4.15), and the property $G^{q+1}_q=1$ to verify explicitly and refine various results obtained in Subsection A, which together yield our principal results for zeros as stated in Theorems 8.1–8.3.

A. Implication of null space for zeros of G^t_q

As discussed in the Introduction, the characteristic null space $\mathscr{N}_t$ is determined fully by the properties of the intertwining function $I_{[M],\Delta}$. The subset of lattice points $\mathbb{N}_t\subset\mathbb{L}^+$ corresponding to the characteristic null space $\mathscr{N}_t$ is described in Fig. 3.

We seek now the consequences for the polynomial $G^t_q(\Delta;x)$ implied by the fact that the denominator function must obey

$$D^2(\Gamma_t;x)=0,\quad \text{each } x\in\mathbb{N}_t, \tag{8.1a}$$

$$D^2(\Gamma_t;x)\neq0,\quad \text{each } x\in\mathbb{L}^+-\mathbb{N}_t, \tag{8.1b}$$

each $t=1,2,...,q+1$.

The key relation for this discussion is Eq. (3.3). Since the ratio of dimension factors in that result is a consequence of normalization conventions (see Ref. 1, p. 76), it is the zeros of the function $R_t(\Delta;x)$ defined by

$$R_t(\Delta;x)\equiv L_t(\Delta;x)G^{t-1}_q(\Delta;x)/G^t_q(\Delta;x) \tag{8.2}$$

that are relevant; that is, it is

$$R_t(\Delta;x)=0,\quad \text{each } x\in\mathbb{N}_t, \tag{8.3}$$

which is implied by Eq. (8.1). We emphasize again that relation (8.1), hence, (8.3), is a consequence of the *intrinsic structure* of the intertwining function and its level subspaces. This is invariant information provided by the group structure itself; accordingly, the requirement (8.3) is necessary in the definition of a canonical Wigner operator.

Because of the reduction formula, Eq. (4.15), we can, without loss of generality, restrict the present discussion to the case $\lambda=(0,0,0)$; that is, we may impose the conditions

$$\Delta_i\geqslant q,\quad i=1,2,3. \tag{8.4}$$

Let us next define what we mean by the multiplicity of a zero of a polynomial $p(x,y)$ of two (independent) variables $(x,y)\in\mathbb{R}^2$. This uses the concept of directional derivative. Suppose that $(a,b)\in\mathbb{R}^2$ is a zero of $p(x,y)$; that is, $p(a,b)=0$. If

$$\left(\alpha\frac{\partial}{\partial x}+\beta\frac{\partial}{\partial y}\right)^r p(x,y)\Big|_{(a,b)}=0, \tag{8.5}$$

for each $r=0,1,...,m-1$ for *arbitrary* $(\alpha,\beta)\in\mathbb{R}^2$, then the zero (a,b) of $p(x,y)$ is said to have multiplicity $m(a,b)=m$. For example, the multiplicity of the zero (1,1) of $(x-1)(y-1)(x+y-2)$ is $m(1,1)=m=3$.

If a point $(a,b)\in\mathbb{R}^2$ is not a zero of $p(x,y)$, we assign the value $m=0$ to the point. In this way, we associate with each polynomial p defined in $\mathbb{R}^2$ a new function m_p, also defined in $\mathbb{R}^2$, which is non-negative-integer-valued:

$$m_p(x,y)=\begin{cases}m(a,b), & \text{for } (x,y)=(a,b)\\ & \text{and } p(a,b)=0,\\ 0, & \text{for } p(x,y)\neq0.\end{cases} \tag{8.6}$$

If $a=(a_1,a_2,a_3)$ is a zero of a polynomial $P(x_1,x_2,x_3)$ defined in $\mathbb{M}$, that is, $x\in\mathbb{M}$, then the multiplicity $m(a)$ of the zero a is defined by $m(a_1,a_2,a_3)=m(a_1,a_3)$, where $m(a_1,a_3)$ is the multiplicity of the zero (a_1,a_3) of $p(x_1,x_3)\equiv P(x_1,-x_1-x_3,x_3)$. For example, the multiplicity of the zero $(1,-2,1)$ of $(x_1-1)(x_2+2)(x_3-1)$ is 3. In analogy to definition (8.6), we also define the function m_P in $\mathbb{M}$ by

$$m_P(x)=\begin{cases}m(a), & \text{for } x=a \text{ and } P(a)=0,\\ 0, & \text{for } P(x)\neq0.\end{cases} \tag{8.7}$$

Choosing $P(x)=G^t_q(\Delta;x)$ [resp. $P(x)=L_t(\Delta;x)$] in definition (8.7) and denoting the generic function $m_p(x)$ by $m_t(x)$ [resp. $l_t(x)$], we have explicitly

$$m_t(x)=\begin{cases}m(a), & \text{for } x=a \text{ and } G^t_q(\Delta;a)=0,\\ 0, & \text{for } G^t_q(\Delta;x)\neq0,\end{cases} \tag{8.8a}$$

$$l_t(x)=\begin{cases}m(a), & \text{for } x=a \text{ and } L_t(\Delta;a)=0,\\ 0, & \text{for } L_t(\Delta;x)\neq0.\end{cases} \tag{8.8b}$$

Here the index t assumes values $1,2,\ldots,q+1$; in addition, we allow $t=0$ in (8.8a) so that

$$m_0(x)=m_{q+1}(x)=0, \tag{8.8c}$$

since $G_q^0(\Delta;x)=G_q^{q+1}(\Delta;x)=1$.

Definition (8.6) of the multiplicity of a zero of a polynomial p may be extended to rational polynomials p/p' by $m_{p/p'}=m_p-m_{p'}$. Applying this definition (as amended above) to the function $R_t(\Delta;x)$ [Eq. (8.2)], we obtain the function r_t:

$$r_t(x)=l_t(x)+m_{t-1}(x)-m_t(x), \tag{8.9}$$

for each $t=1,2,\ldots,q+1$, and for each $x\in\mathbb{M}$. Thus, r_t is a function in $\mathbb{M}$ with values $r_t(x)\in\mathbb{Z}$ (set of integers). We refer to $r_t(x)$ as the rank of the denominator $D^2(\Gamma_t;x)$ at point x and call r_t a rank function.

Relation (8.9) is the principal result for investigating the properties of the functions m_t. By regarding the $l_t(x)$ and $r_t(x)$ as known for $t=1,\ldots,q+1$, relation (8.9) becomes a recurrence relation for the functions m_t. This is the viewpoint we generally adopt. The functions l_t are, of course, known explicitly from the zeros of the linear factors in $L_t(\Delta;x)$. The rank functions r_t are not known *a priori*, so that the principal task in implementing this viewpoint is one of determining the conditions imposed on the rank functions by null space, and possibly other considerations. Clearly, relation (8.9) determines the $m_t(x)$ explicitly only when the $r_t(x)$ are known. For this investigation, we restrict our considerations to points $x\in\mathbb{L}^+$ (the set of lattice points in $\mathbb{M}$ corresponding to irrep spaces $[m_{13},m_{23},m_{33}]$).

Implementing the viewpoint discussed above, we write Eq. (8.9) in the form

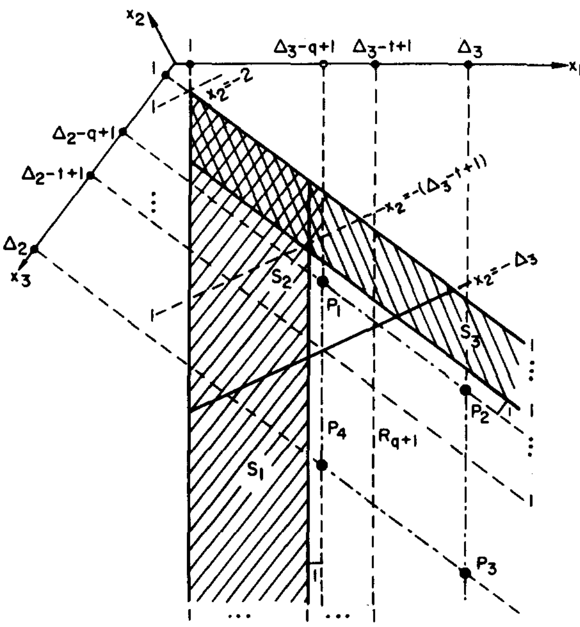


FIG. 5. Values of $l_t(x)$. The value of $l_t(x)$ at each $(x_1,x_2,x_3)\in\mathbb{L}^+$ is obtained by adding the following contributions from the x_1-, x_2-, x_3- lines, respectively: 1 if $1\leqslant x_1\leqslant\Delta_3-t+1$, 0 otherwise; 1 if $-2\geqslant x_2\geqslant-(\Delta_3-t+1)$, 0 otherwise; 1 if $1\leqslant x_3\leqslant\Delta_2-t+1$, 0 otherwise. The subdomains of $\mathbb{L}^+$ denoted by S_1,S_2,S_3 are useful for describing contributions to the multiplicity function $m_t(x)$ from various lines as described in Eq. (8.14). The rhombic-shaped domain $\mathbb{R}_{q+1}$ with vertices P_1,P_2,P_3,P_4 is described in Eq. (8.15); it plays an important role in this section.

$$m_t(x)-m_{t-1}(x)=l_t(x)-r_t(x) \tag{8.10}$$

and iterate in t starting first with $m_0(x)=0$ and then with $m_{q+1}(x)=0$ and derive

$$m_t(x)=\sum_{s=1}^{t}\left[l_s(x)-r_s(x)\right], \tag{8.11a}$$

$$m_t(x)=\sum_{s=t+1}^{q+1}\left[r_s(x)-l_s(x)\right],\quad t=1,\ldots,q. \tag{8.11b}$$

The equality of these two expressions gives the following condition that the rank functions must satisfy:

$$\sum_{s=1}^{q+1}r_s(x)=\sum_{s=1}^{q+1}l_s(x). \tag{8.12}$$

We also impose the following conditions on the rank functions in consequence of their role in determining where the denominator function $D(\Gamma_t;x)$ vanishes: (i) for each $x\in\mathbb{N}_t$, the value of r_t is a positive integer, that is,

$$r_t(x)\in\mathbb{Z}^+; \tag{8.13a}$$

(ii) for each $x\in\mathbb{L}^+$ and not in $\mathbb{N}_t$, the value of r_t is zero, that is,

$$r_t(x)=0,\quad\text{for }x\in\mathbb{L}^+-\mathbb{N}_t. \tag{8.13b}$$

The strategy now is to evaluate the right-hand side of Eq. (8.12) from explicit values of the $l_s(x)$ and then to use conditions (8.13) to determine the $r_s(x)$, when possible. The value $l_s(x)$ at $x=(a_1,a_2,a_3)\in\mathbb{L}^+$ is either 0,1,2, or 3; it is obtained by counting the number of linear factors in $(-x_1+1)_{\Delta_3-s+1}(x_2+2)_{\Delta_3-s}(-x_3+1)_{\Delta_2-s+1}$ that have value zero at $x=a$. To carry out this counting, it is convenient to use Fig. 5. The number 1 at the lower end of the vertical lines denotes that there is one linear factor in $(-x_1+1)_{\Delta_3-t+1}$ that vanishes on that line; the 1's at the right-hand margin of the figure have this same interpretation for the factors $(-x_3+1)_{\Delta_2-t+1}$, while those arranged vertically near the $x_1=1$ line refer to the factors $(x_2+2)_{\Delta_3-t}$. All other x_1-, x_2-, x_3- lines in $\mathbb{L}^+$ have 0 assigned to them, respectively. The number $l_t(x)$ is then obtained by summing the three integers corresponding to the x_1, x_2, and x_3 lines.

It is convenient at this point to define some subdomains (sets of lattice points) of $\mathbb{L}^+$ and identify them in Fig. 5.

(i) The "strips" S_i $(i=1,2,3)$ are defined by

$$\begin{aligned}S_1&=\{x\in\mathbb{L}^+\,|\,1\leqslant x_1\leqslant\Delta_3-q,\ x_3\geqslant1\},\\ S_2&=\{x\in\mathbb{L}^+\,|\,x_1\geqslant1,\ x_3\geqslant1,\ -\Delta_3\leqslant x_2\leqslant-2\},\\ S_3&=\{x\in\mathbb{L}^+\,|\,1\leqslant x_3\leqslant\Delta_2-q,\ x_1\geqslant1\}.\end{aligned} \tag{8.14}$$

Since we take $x\in\mathbb{L}^+$, the set S_2 actually has a triangular boundary. By definition, each of these sets is empty if the defining conditions are violated (for example, $\Delta_3=q$ in S_1).

(ii) The set $\mathbb{R}_{q+1}$ with the rhombic boundary and vertices P_1,P_2,P_3,P_4 is defined by

$$\mathbb{R}_{q+1}=\left\{x\in\mathbb{L}^+\left|\begin{matrix}\Delta_3-q+1\leqslant x_1\leqslant\Delta_3\\ \Delta_2-q+1\leqslant x_3\leqslant\Delta_2\end{matrix}\right.\right\}. \tag{8.15}$$

The set $\mathbb{R}_{q+1}$ is of particular significance for the present discussion and appears several times later in various contexts.

(iii) The line segments $\mathbb{I}_k^{(i)}$ $(i=1,2,3)$ are defined by

$$\mathbf{l}_k^{(1)} = \{x \in \mathbb{L}^+ | x_1 = \Delta_3 - k + 1,\ x_2 \leqslant -(\Delta_3 + 1),$$
$$x_3 \geqslant \Delta_2 + 1\},\ k = 1,...,\Delta_3; \tag{8.16a}$$
$$\mathbf{l}_k^{(2)} = \{x \in \mathbb{L}^+ | x_2 = -(\Delta_2 + \Delta_3 - k + 2),$$
$$x_1 \leqslant \Delta_3 + 1,\ x_3 \leqslant \Delta_2 + 1\},\ k = 1,...,2q\ (l_1^{(2)} = \varnothing); \tag{8.16b}$$
$$\mathbf{l}_k^{(3)} = \{x \in \mathbb{L}^+ | x_3 = \Delta_2 - k + 1,\ x_1 \geqslant \Delta_3 + 1\},$$
$$k = 1,2,...,\Delta_2. \tag{8.16c}$$

The condition $x_2 \leqslant -(\Delta_3 + 1)$ in the definition of the line segment $\mathbf{l}_k^{(1)}$ is made to keep the endpoint exterior to the region S_2. The line segments $\mathbf{l}_k^{(2)}$ cover $\mathbb{R}_{q+1}$, that is, $\mathbb{R}_{q+1} = \cup_{k=1}^{2}\mathbf{l}_k^{(2)}$. Each of these lines of lattice points is parallel to the *minor diagonal* of $\mathbb{R}_{q+1}$.

(iv) The broken line $\mathbb{L}_s$ is defined by

$$\mathbb{L}_s = \mathbf{l}_s^{(1)} \cup \mathbf{l}_s^{(2)} \cup \mathbf{l}_s^{(3)}, \quad s = 1,...,q+1. \tag{8.17}$$

This broken line is the boundary between the null space $\mathbb{N}_s$ and $\mathbb{L}^+ - \mathbb{N}_s$ (cf. Figs. 2 and 5). Because of the inclusion property $\mathbb{N}_{q+1} \subset \cdots \subset \mathbb{N}_1$, we can also write

$$\mathbb{L}_s = \mathbb{N}_s - \mathbb{N}_{s+1}, \quad s = 1,...,q. \tag{8.18}$$

Using the above notations and definitions, we can now prove the following lemma.

Lemma 8.1: The sum $\Sigma_{s=1}^{q+1} l_s(x)$ has the following values in the indicated (disjoint) subdomains of $\mathbb{L}^+$:

(i) $\displaystyle\sum_{s=1}^{q+1} l_s(x) = k,$

$$\text{for } x \in \mathbb{L}_k,\ k = 1,...,q+1; \tag{8.19a}$$

(ii) $\displaystyle\sum_{s=1}^{q+1} l_s(x) = q+1,$

for either $x \in \mathbf{l}_k^{(1)} \subset S_1$

with $k = q+1,...,\Delta_3$

or $x \in \mathbf{l}_k^{(3)} \subset S_3$

$$\text{with } k = q+1,...,\Delta_2; \tag{8.19b}$$

(iii) $\displaystyle\sum_{s=1}^{q+1} l_s(x) = q+k,$

for $x \in \mathbf{l}_{q+k}^{(2)}$

$$\text{with } k = 2,...,\min(q, \Delta_2 - q + 1); \tag{8.19c}$$

(iv) $\displaystyle\sum_{s=1}^{q+1} l_s(x) > q+1$

$$\text{for all points } x \text{ not included in (i)–(iii).} \tag{8.19d}$$

Proof: There is considerable detailed information in this lemma—it can all be obtained from Fig. 5. Let us consider each case briefly. For the proof of (i), we have that for each specified k and for either $x \in \mathbf{l}_k^{(1)}$ or $x \in \mathbf{l}_k^{(3)}$ the multiplicity $l_s(x) = 0$ for $s = k+1,...,q+1$ and $l_s(x) = 1$ for $s = 1,...,k$. Thus, $\Sigma_{s=1}^{q+1} l_s(x) = k$ for either $x \in \mathbf{l}_k^{(1)}$ or $x \in \mathbf{l}_k^{(3)}$. Since $\mathbb{L}_k = \mathbf{l}_k^{(1)} \cup \mathbf{l}_k^{(2)} \cup \mathbf{l}_k^{(3)}$, we still need to consider the line segment $\mathbf{l}_k^{(2)} \subset \mathbb{R}_{q+1}$. For point $x \in \mathbf{l}_k^{(2)}$, we argue as follows: We have $\Sigma_{s=1}^{q+1} l_s(x) = k$ for $x = (\Delta_3 - k + 1,\ -\Delta_2 - \Delta_3 + k - 2,\ \Delta_2 + 1)$, which is the *endpoint* of the line $\mathbf{l}_k^{(1)}$. As we move from this endpoint to the left endpoint of $\mathbf{l}_k^{(2)}$ (boundary point of $\mathbb{R}_{q+1}$), we lose a zero from the x_1 factors $\Pi_{s=1}^{q+1}(-x_1 + 1)_{\Delta_3 - s + 1}$, but gain one from the x_3 factors $\Pi_{s=1}^{q+1}(-x_3 + 1)_{\Delta_2 - s + 1}$, thus keeping the sum at value k. This property repeats at each lattice point of $\mathbf{l}_k^{(2)}$ as we move across this line from the left endpoint to the right endpoint (again a boundary point of $\mathbb{R}_{q+1}$). These results prove (i).

Property (ii) refers to those lines $\mathbf{l}_k^{(1)}$ (resp. $\mathbf{l}_k^{(3)}$) to the left of $x_1 = \Delta_3 - q + 1$ (resp. above $x_3 = \Delta_2 - q + 1$) for which each $l_s(x) = 1$, $s = 1,...,q+1$.

The proof of property (iii) is a repetition of the argument proving property (i).

Points not included in (i)–(iii) have $l_s(x) \geqslant 1$ for $s = 1,...,q+1$ with at least one point having $l_s(x) \geqslant 2$. This proves property (iv). ∎

Using the results given in Lemma 8.1, we can now determine the rank functions r_t in certain subdomains of $\mathbb{L}^+$.

Lemma 8.2: The rank functions $r_1,...,r_{q+1}$ have the following values in the indicated subdomains of $\mathbb{L}^+$:

(i) $r_1(x) = \cdots = r_k(x) = 0, \quad r_{k+1}(x) = \cdots = r_{q+1}(x) = 1,$

$$\text{for } x \in \mathbb{L}_k,\quad k = 1,...,q+1; \tag{8.20a}$$

(ii) $r_k(x) = l_k(x) = 1,$

for either $x \in \mathbf{l}_k^{(1)} \subset S_1$

with $k = q+1,...,\Delta_3$

$$\text{or } x \in \mathbf{l}_k^{(3)} \subset S_3 \text{ with } k = q+1,...,\Delta_2; \tag{8.20b}$$

(iii) $r_1(x),...,r_{q+1}(x)$ are not determined by Lemma 8.1 for all x not included in (i)–(ii).

Proof: Properties (i) and (ii) are immediate consequences of Eq. (8.12) conditions (8.13), and properties (i) and (ii) in Lemma 8.1. For all other points $x \in \mathbb{L}^+$ not included in (i) and (ii), we have that

$$\sum_{s=1}^{q+1} r_s(x) > q+1. \tag{8.21}$$

In this case, conditions (8.13) are insufficient for the determination of the individual rank functions. ∎

Remarks: (a) It is worth remarking that the set of lattice points for which the rank functions are not given by Lemma 8.2 is a subset of the null space $\mathbb{N}_{q+1}$, which itself is a subset of all other $\mathbb{N}_t$, $t = 1,...,q$; that is, the irrep spaces corresponding to points $x \in \mathbb{N}_{q+1}$ are in the characteristic null space $\mathscr{N}_t$ of each of the Wigner operators

$$\left\langle \begin{matrix} & \Gamma_t & \\ p & q & 0 \\ & \cdot & \end{matrix} \right\rangle, \quad t = 1,...,q+1.$$

It is useful to show in Fig. 6 the relation of $\mathbb{N}_{q+1}$ to the set $\mathbb{R}_{q+1}$ identified earlier in Fig. 5. The explicit set of points $\mathbb{N}'_{q+1} \subset \mathbb{N}_{q+1}$ for which relation (8.21) is valid [the points not included in parts (i) and (ii) of Lemma 8.2] is obtained by removing from $\mathbb{N}_{q+1}$ the boundary $\mathbb{L}_{q+1}$ and the line segments $\mathbf{l}_k^{(1)}$ $(k = q+1,...,\Delta_3)$ and $\mathbf{l}_k^{(3)}$ $(k = q+1,...,\Delta_2)$:

$$\mathbb{N}'_{q+1} \equiv \mathbb{N}_{q+1} - \mathbb{L}_{q+1} - \bigcup_{k=q+1}^{\Delta_3} \mathbf{l}_k^{(1)} \bigcup_{k=q+1}^{\Delta_2} \mathbf{l}_k^{(3)}. \tag{8.22}$$

(b) The explicit solution for $m_t(x)$ for points

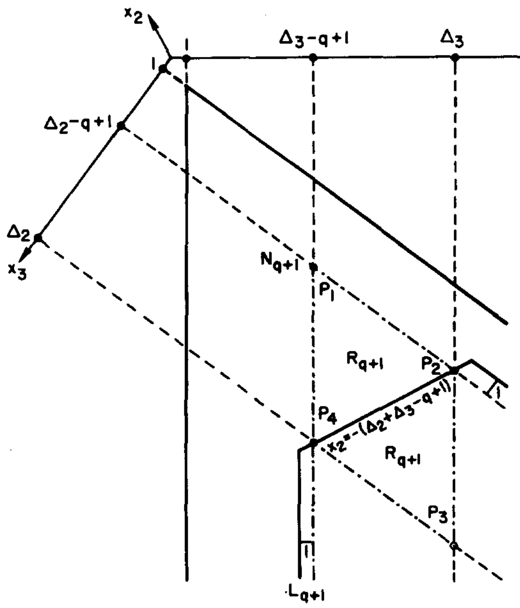


FIG. 6. The null space set $\mathbf{N}_{q+1}$. This is the set of lattice points interior to and on the bold solid boundary lines. All irrep labels $[m]$ such that $\phi([m]) \in \mathbf{N}_{q+1}$ correspond to irrep spaces $\mathscr{H}([m]) \subset \mathscr{N}_{q+1}$, which is a subspace of $\mathscr{H}$ contained in the null space of all unit tensor operators in the family $\{\langle\Gamma_t\rangle\}$.

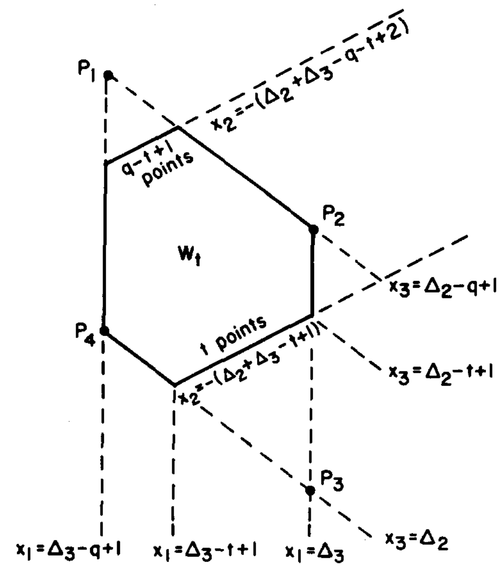


FIG. 7. The weight space $\mathbf{W}_t$. The set of lattice points (within the rhombic region R_{q+1}) interior to and on the bold solid boundary lines of the hexalateral defines the set of weight space points for irrep $[q-t, 0, -t+1]$ of U(3). It is positioned in the Möbius plane by the vertices P_1, P_2, P_3, P_4 (see Fig. 5).

$x \in \mathbb{L}^+ - \mathbf{N}'_{q+1}$ and each $t = 1,\ldots,q+1$ may be obtained from Lemmas 8.1 and 8.2:

$$m_t(x) = \sum_{s=1}^{t} l_s(x) - t. \tag{8.23}$$

We do not, however, use this expression directly for evaluating $m_t(x)$ in the subsequent discussion.

In order to make further progress in the determination of the rank functions $r_1,\ldots,r_{q+1}$ for all points $x \in \mathbb{L}^+$, it is convenient to summarize the results obtained thus far in a form that suggests generalization. For this purpose, we need to define additional classes of subsets of $\mathbb{L}^+$.

For each specified q and $(\Delta_1,\Delta_2,\Delta_3)$, the subset $\mathbf{W}_t \subset \mathbf{R}_{q+1}$ is defined by

$$\mathbf{W}_t = \{x \in \mathbf{R}_{q+1} \mid -(\Delta_2+\Delta_3-q-t+2) \geqslant x_2 \geqslant -(\Delta_2+\Delta_3-t+1)\}, \tag{8.24}$$

each $t = 1,\ldots,q$. The set $\mathbf{W}_t$ consists of those lattice points on the boundary and interior to the symmetric hexalateral (equilateral triangle for $t = 1,q$) shown in Fig. 7. We refer to the sets $\mathbf{W}_1$ and $\mathbf{W}_q$ as the lower and upper equilateral triangular subsets of $\mathbf{R}_{q+1}$, respectively.

For each specified q and $(\Delta_1,\Delta_2,\Delta_3)$, we also define subsets $\mathbf{T}_t \subset \mathbf{R}_{q+1}$ and $\mathbf{T}'_t \subset \mathbf{R}_{q+1}$, respectively, by

$$\mathbf{T}_t = \left\{ x \in \mathbf{R}_{q+1} \left| \begin{array}{l} x_1 > \Delta_3 - t + 1 \\ x_3 > \Delta_2 - t + 1 \\ -(\Delta_2+\Delta_3-t+2) \leqslant x_2 \leqslant -(\Delta_2+\Delta_3-2t+4) \end{array} \right. \right\}, \tag{8.25a}$$

$$\mathbf{T}'_t = \left\{ x \in \mathbf{R}_{q+1} \left| \begin{array}{l} x_1 < \Delta_3 - t + 1 \\ x_3 < \Delta_2 - t + 1 \\ -(\Delta_2+\Delta_3-2t+2) \leqslant x_2 \leqslant -(\Delta_2+\Delta_3-q-t+2) \end{array} \right. \right\}, \tag{8.25b}$$

each $t = 1,\ldots,q+1$, where $\mathbf{T}_1 = \mathbf{T}'_{q+1} = \varnothing$. These sets are the ones shown in Fig. 8 with the triangular boundaries (solid lines).

Finally, in Fig. 9, we give the explicit values $l_t(x)$ of the function l_t for each $x = \mathbf{R}_{q+1}$.

Using the sets introduced above, we can now summarize the principal results obtained thus far in this section in the following way.

Lemma 8.3: The following relations are valid for each $x \in \mathbb{L}^+ - \mathbf{N}'_{q+1}$ and for each $t = 1,\ldots,q+1$:

(i) $m_t(x) = 0$, for $x \notin \mathbf{W}_t$; (8.26a)

(ii) if $m_t(x) \neq 0$, then $r_t(x) = 1$; (8.26b)

(iii) $$m_t(x) - m_{t-1}(x) = l_t(x) - r_t(x) = \begin{cases} -1, & \text{if } x \in \mathbf{T}_t, \\ +1, & \text{if } x \in \mathbf{T}'_t, \\ 0, & \text{if } x \in \mathbb{L}^+ - \mathbf{T}_t - \mathbf{T}'_t. \end{cases} \tag{8.26c}$$

Proof: For each point $x \in \mathbb{L}^+ - \mathbf{N}'_{q+1}$, the value of $m_t(x)$ is given by Eq. (8.23). At each point $x \in \mathbb{L}^+ - \mathbf{N}'_{q+1} - \mathbf{W}_t$, the value of $\Sigma_{s=1}^{t} l_s(x)$ is t [see parts (i) and (iii) of Lemma 8.1 and Figs. 5, 7, and 9]. This proves part (i).

The only points $x \in \mathbb{L}^+ - \mathbf{N}'_{q+1}$ for which the relation $m_t(x) \neq 0$ is possibly true are those $x \in \mathbf{W}_t$. But for such

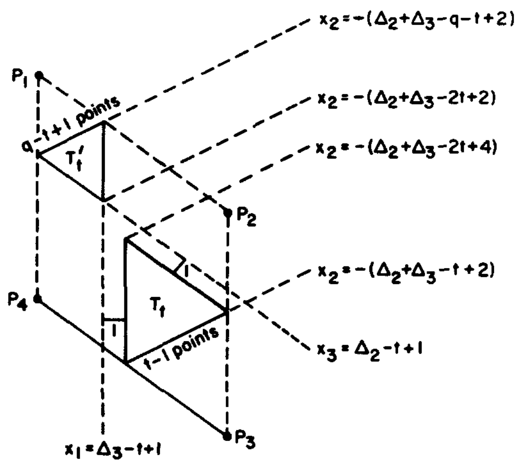


FIG. 8. Definition of domains $\mathbf{T}_t'$ and $\mathbf{T}_t$. The subdomains $\mathbf{T}_t'$ and $\mathbf{T}_t$ are the sets of lattice points interior to and on the solid boundary lines. These domains are significant for the study of the recurrence relation for $m_t(x) - m_{t-1}(x)$.

points, we have $r_t(x) = 1$ [part (ii) of Lemma 8.2]. This proves part (ii). [Observe, however, that $r_t(x) = 1$ does not imply $m_t(x) \neq 0$.]

Property (iii) is a restatement of recurrence relation (8.10) with the difference $l_t(x) - r_t(x)$ now identified explicitly from Lemmas 8.1 and 8.2 and the results given in Figs. 5–9. ■

We emphasize that Lemma 8.3 has *not* been established for points $x \in \mathbb{N}_{q+1}'$. Additional restrictions are required in order to determine the $r_t(x)$, hence, the $m_t(x)$, for $x \in \mathbb{N}_{q+1}'$. The next lemma gives such a result.

Lemma 8.4: If $m_t(x) = 0$ for $x \in \mathbb{N}_{q+1}'$ and $x \notin \mathbb{W}_t$, then Lemma 8.3 is valid for all $x \in \mathbb{L}^+$; hence, $m_t(x)$ is uniquely determined for all $x \in \mathbb{L}^+$ by the recurrence relation (8.26c).

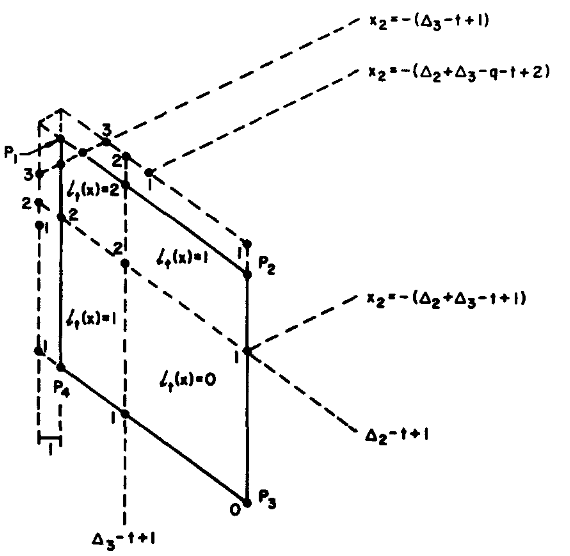


FIG. 9. Values of $l_t(x)$ in $\mathbf{R}_{q+1}$. For (x_1,x_2,x_3) in the rhombic-shaped domain $\mathbf{R}_{q+1}$, or one unit outside the left x_1- or the upper x_3-boundary edges, linear factors contribute the value $l_t(x)$ as shown to the rank function $r_t(x)$. The $l_t(x) = 2$ domain has as upper boundary edge the x_2 line $x_2 = -(\Delta_3 - t + 1)$ on which $l_t(x) = 3$, which either belongs to $\mathbf{R}_{q+1}$ or not according to $\Delta_2 + t \leqslant 2q - 1$ or $\Delta_2 + t > 2q - 1$. At any point on a boundary line between subdomains in $\mathbf{R}_{q+1}$ (dashed lines), $l_t(x)$ always takes the higher value. The fact that these values extend (at least) one unit off the boundary of $\mathbf{R}_{q+1}$ as shown is significant.

Proof: The hypothesis in Lemma 8.4 extends property (i) in Lemma 8.3 to all $x \in \mathbb{L}^+$, so that

$$m_t(x) = 0, \quad \text{for } x \in \mathbb{L}^+ - \mathbb{W}_t. \tag{8.27}$$

Accordingly, the recurrence relation (8.10) yields

$$r_t(x) = l_t(x), \quad \text{for } x \in \mathbb{L}^+ - \mathbb{W}_t \cup \mathbb{W}_{t-1}. \tag{8.28}$$

We have also determined earlier (Lemma 8.2) the value of $r_t(x)$ for all $x \in \mathbb{L}^+ - \mathbb{N}_{q+1}'$; this includes all points in $\mathbb{W}_t \cup \mathbb{W}_{t-1}$ having $x_2 \leqslant -(\Delta_2 + \Delta_3 - q + 1)$ (below and on the minor diagonal in $\mathbb{R}_{q+1}$):

$$r_t(x) = 1, \quad \text{for } x \in \mathbb{W}_t \text{ and } x_2 \leqslant -(\Delta_2 + \Delta_3 - q + 1), \tag{8.29a}$$

$$r_t(x) = 0, \quad \text{for } x \in \mathbb{W}_{t-1}, \quad x \notin \mathbb{W}_t. \tag{8.29b}$$

Equations (8.28) and (8.29) give the value of $r_t(x)$, $t = 1,...,q+1$ for all $x \in \mathbb{L}^+$ except those in $\mathbb{W}_t$ $(t \geqslant 2)$ above the minor diagonal of $\mathbb{R}_{q+1}$.

We next complete the determination of $r_t(x)$ for all $x \in \mathbb{L}^+$ by using relation (8.19c):

$$\sum_{s=1}^{q+1} r_s(x) = q + k, \tag{8.30a}$$

for

$$x_2 = -(\Delta_2 + \Delta_3 - q - k + 2), \quad k = 2,...,t, \tag{8.30b}$$

each $t = 2,...,q$. From Eq. (8.29a) and Fig. 9, we find

$$r_1(x) = l_1(x) = 2, \quad \text{for } x \in \mathbb{R}_{q+1},$$
$$-(\Delta_2 + \Delta_3 - q) \leqslant x_2 \leqslant -(\Delta_2 + \Delta_3 - q - t + 2). \tag{8.31a}$$

Using this result and condition (8.13a) in Eq. (8.30a) with $k = 2$ yields

$$r_2(x) = \cdots = r_{q+1}(x) = 1, \quad \text{for } x_2 = -(\Delta_2 + \Delta_3 - q). \tag{8.31b}$$

Using next Eq. (8.29a) and Fig. 9, we find

$$r_2(x) = l_2(x) = 2, \quad \text{for } x \in \mathbb{R}_{q+1},$$
$$-(\Delta_2 + \Delta_3 - q - 1) \leqslant x_2 \leqslant -(\Delta_2 + \Delta_3 - q - t + 2). \tag{8.32a}$$

Using this result, Eq. (8.31a), and condition (8.13a) in Eq. (8.30a) with $k = 3$ yields

$$r_3(x) = \cdots = r_{q+1}(x) = 1, \quad \text{for } x_2 = -(\Delta_2 + \Delta_3 - q - 1). \tag{8.32b}$$

Continuing in this way yields the following general result at step $s - 1$:

$$r_{s-1}(x) = l_{s-1}(x) = 2, \quad \text{for } x \in \mathbb{R}_{q+1},$$
$$-(\Delta_2 + \Delta_3 - q - s + 2) \leqslant x_2 \leqslant -(\Delta_2 + \Delta_3 - q - t + 2); \tag{8.33a}$$

$$r_s(x) = \cdots = r_{q+1}(x), \quad \text{for } x \in \mathbb{R}_{q+1}$$
$$\text{and } x_2 = -(\Delta_2 + \Delta_3 - q - s + 2), \tag{8.33b}$$

for $s = 2,...,t$. The sth relation in this set yields $r_t(x) = 1$ for $x_2 = -(\Delta_2 + \Delta_3 - q - s + 2)$. Thus, we find

$$r_t(x) = 1, \quad \text{for } x \in \mathbb{W}_t \text{ and } x_2 \geqslant -(\Delta_2 + \Delta_3 - q - 1). \tag{8.34}$$

We have now determined each $r_t(x)$ for all $x \in \mathbb{L}^+$: Equations (8.28), (8.29b), and (8.34) yield

$$r_t(x) = l_t(x), \quad \text{for } x \in \mathbb{L}^+ - \mathbb{W}_t, \tag{8.35a}$$

$$r_t(x) = 1, \quad \text{for } x \in \mathbb{W}_t, \tag{8.35b}$$

each $t = 1,...,q+1$. These two results and the values of $l_t(x)$ given in Fig. 9 yield the recurrence relation (8.26c), which is now valid for all $x \in \mathbb{L}^+$.

Finally, relation (8.26b) is also now valid for all $x \in \mathbb{L}^+$ in consequence of relation (8.34) and the fact that if $m_t(x) \neq 0$, then necessarily $x \in \mathbb{W}_t$. ■

Remark: Lemma 8.3 summarizes *proved results* for the functions $m_t(x)$, $l_t(x)$, and $r_t(x)$ for all $x \in \mathbb{L}^+ - \mathbb{N}'_{q+1}$. The hypothesis of Lemma 8.4, which is that $m_t(x) = 0$ for $x \in \mathbb{N}'_{q+1}$ and $x \notin \mathbb{W}_t$, allows us to extend the validity of Lemma 8.3 to all $x \in \mathbb{L}^+$. We have, however, not proved this hypothesis. Partial results in this direction are given in Theorems 8.2 and 8.3.

In order to proceed let us next give the solution to the recurrence relation (8.26c) *assuming its validity for all* $x \in \mathbb{L}^+$. We will do this, not by direct iteration, but by introducing a known set of functions M_t, $t = 1,...,q$, and showing that these functions satisfy the same recurrence relation. The key structural elements for this analysis are the weight spaces of the U(3) irreps $[q-t,0,-t+1]$, $t = 1,...,q$. (Alternatively, one may use the weight spaces of the equivalent SU(3) irreps $[q-1,t-1,0]$.) For this, we require an explicit realization of these weight spaces as sets of lattice points in the Möbius plane, as we next describe.

Let $w = (w_1,w_2,w_3)$ denote a weight of irrep $[q-t,0,-t+1]$ of U(3), where $t \in \{1,...,q\}$. The multiplicity of this weight, which we denote by $M_t(w)$, is given by

$$M_t(w) = \begin{cases} t - (\lambda_{w_1} + \lambda_{w_2} + \lambda_{w_3}), & \text{for } w_1 + w_2 + w_3 = q - 2t + 1 \text{ and } q - t \geqslant w_i \geqslant -t+1, \\ 0, & \text{otherwise,} \end{cases} \tag{8.36a}$$

where

$$\lambda_{w_i} = \max(0, -w_i). \tag{8.36b}$$

[This result is just the adaptation of Eq. (2.8c) to U(3) irreps.]

Let q,t, and $(\Delta_1,\Delta_2,\Delta_3)$ be specified (with $\Delta_i \geqslant q$). We now assign the weight $(w_1,w_2,w_3) \in [q-t,0,-t+1]$ and its multiplicity number $M_t(w)$ to the point $(x_1,x_2,x_3) \in \mathbb{L}^+$ given by the map

$$\begin{aligned} x_1 &= \Delta_3 - t + 1 - w_1, \\ x_2 &= -\Delta_2 - \Delta_3 + q - 1 - w_2, \\ x_3 &= \Delta_2 - t + 1 - w_3. \end{aligned} \tag{8.37a}$$

The set of points defined by

$$\mathbb{W}_t = \{(x_1,x_2,x_3) | (w_1,w_2,w_3) \in [q-t,0,-t+1]\} \tag{8.37b}$$

is then exactly the set $\mathbb{W}_t$ defined by Eq. (8.24) for each $t = 1,...,q$. Thus, the set of points $\mathbb{W}_t$ with the assigned multiplicity numbers $M_t(w)$ constitutes an explicit realization in the Möbius plane of the weight space of the U(3) irrep $[q-t,0,-t+1]$.

It is convenient now to introduce the function $\mathscr{M}_t$ on $\mathbb{L}^+$ defined by

$$\mathscr{M}_t(x) = \begin{cases} M_t(w), & \text{for } x \in \mathbb{W}_t, \\ 0, & \text{for } x \in \mathbb{L}^+ - \mathbb{W}_t. \end{cases} \tag{8.38}$$

Thus, $\mathscr{M}_t(x)$ is an extension of the multiplicity of a weight to all points in $\mathbb{L}^+$.

In order to derive a recurrence relation for $\mathscr{M}_t(x)$, it is convenient to use the geometrical realization of this function given by:

$$\mathscr{M}_t(x) = \min(t, q-t+1, 1 + d_t(x)), \quad x \in \mathbb{W}_t, \tag{8.39}$$

where $d_t(x)$ is the "distance" from lattice point $x \in \mathbb{W}_t$ to the nearest boundary of $\mathbb{W}_t$ as measured along the direction of the appropriate coordinate axis (1 lattice spacing = 1 unit of distance with $d_t = 0$ corresponding to boundary points).

Using Eq. (8.39), it is now straightforward to verify

$$\mathscr{M}_t(x) - \mathscr{M}_{t-1}(x) = \begin{cases} -1, & \text{if } x \in \mathbb{T}_t, \\ +1, & \text{if } x \in \mathbb{T}'_t, \\ 0, & \text{if } x \in \mathbb{L}^+ - \mathbb{T}_t - \mathbb{T}'_t, \end{cases} \tag{8.40}$$

for each $x \in \mathbb{L}^+$ and $t = 1,2,...,q+1$ with $\mathscr{M}_0(x) = \mathscr{M}_{q+1}(x) = 0$. Here $\mathbb{T}_t$ and $\mathbb{T}'_t$ are the sets defined by Eqs. (8.25) (see Fig. 8). We have thus proved that the function $\mathscr{M}_t(x)$ is the unique function satisfying for all $x \in \mathbb{L}^+$ the recurrence relation (8.26c) with boundary conditions $\mathscr{M}_0(x) = \mathscr{M}_{q+1}(x) = 0$. It is also clear that the rank functions $\{r_s(x) | s = 1,...,q+1\}$ that lead to this solution are

$$\begin{aligned} r_s(x) &= l_s(x), \quad \text{for } x \in \mathbb{L}^+ - \mathbb{W}_s, \\ r_s(x) &= 1, \quad \text{for } x \in \mathbb{W}_s. \end{aligned} \tag{8.41}$$

The restriction of $\mathscr{M}_t(x)$ to the domain $\mathbb{L}^+ - \mathbb{N}'_{q+1}$ gives (uniquely) the function $m_t(x)$ satisfying the properties of Lemma 8.3; that is, this proves Lemma 8.5.

Lemma 8.5: For each $x \in \mathbb{L}^+ - \mathbb{N}'_{q+1}$, the multiplicity function $m_t(x)$ is given by $m_t(x) = \mathscr{M}_t(x)$.

The following result is also true from Lemma 8.4.

Lemma 8.6: If $m_t(x) = 0$ for $x \in \mathbb{N}'_{q+1}$ and $x \notin \mathbb{W}_t$, then, for each $x \in \mathbb{L}^+$, the multiplicity function $m_t(x)$ is given by $m_t(x) = \mathscr{M}_t(x)$.

Using only the basic properties (8.13) of the rank functions, $r_1,...,r_{q+1}$ and the functions $\mathscr{M}_1,...,\mathscr{M}_q$ defined above, we can also prove the following theorem.

Theorem 8.1: The basic properties of the rank functions given by

$$r_t(x) \geqslant 1, \quad \text{for } x \in \mathbb{N}_t, \tag{8.42a}$$

$$r_t(x) = 0, \quad \text{for } x \in \mathbb{L}^+ - \mathbb{N}_t \tag{8.42b}$$

imply the following relations for the multiplicity functions m_t:

$$m_t(x) = \mathscr{M}_t(x), \quad \text{for } x \in \mathbb{L}^+ - \mathbb{N}'_{q+1}, \tag{8.43a}$$

$$m_t(x) \geqslant \mathscr{M}_t(x), \quad \text{for } x \in \mathbb{N}'_{q+1}. \tag{8.43b}$$

Proof: The identity (8.43a) is a restatement of Lemma 8.5. The new result here is relation (8.43b), which we now prove. For this, we use Eq. (8.11b):

$$m_t(x) = \sum_{s=t+1}^{q+1} [r_s(x) - l_s(x)]. \tag{8.44}$$

For $x \in \mathbb{W}_t$, the value $\mathscr{M}_t(x)$ obtains from Eq. (8.44) by setting $r_s(x) = 1$ for $s = t+1,...,q+1$ [see Eq. (8.41)]. This corresponds to the solution given by Eq. (8.40) for all $x \in \mathbb{L}^+$. From Eq. (8.42a) we now only have $r_s(x) \geqslant 1$ for $x \in \mathbb{W}_s \cup \mathbb{N}'_{q+1}$, $s = t+1,...,q+1$, so that relation (8.43b) follows from Eq. (8.44) for these x; for $x \in \mathbb{N}'_{q+1} - \mathbb{W}_t$, relation (8.43) is trivially true, since $\mathscr{M}_t(x) = 0$. ■

Theorem 8.1 gives the most complete result we have been able to derive for the multiplicity function $m_t(x)$ using only the recurrence relation and the properties (8.42) of the rank functions. The hypothesis of Lemma 8.6 gives additional (as yet unproved) properties of $m_t(x)$, hence, of the rank functions [see Eqs. (8.41)], under which the solution $\mathscr{M}_t(x)$ for $x \in \mathbb{L}^+ - \mathbb{N}'_{q+1}$ extends to all of $\mathbb{L}^+$. The next lemma is also of this type. For its statement, we require an additional property of the function $M_t(w)$.

It is well known that the multiplicity of the weight $\overline{w} = (\overline{w}_1,\overline{w}_2,\overline{w}_3) = (-w_1, -w_2, -w_3)$ in the conjugate irrep $[t-1,0,t-q]$ equals that of the weight (w_1,w_2,w_3) in irrep $[q-t,0,-t+1]$; that is,

$$M_{q-t+1}(-w_1,-w_2,-w_3) = M_t(w_1,w_2,w_3). \tag{8.45a}$$

Moreover, using the well-known symmetry of weights under permutations (Weyl group S_3), we can write

$$M_{q-t+1}(-w_3,-w_2,-w_1) = M_t(w_1,w_2,w_3). \tag{8.45b}$$

The property (8.45b) of conjugate weights corresponds, in the realization of weight spaces in the Möbius plane given by Eqs. (8.36)–(8.38), to the following: Reflect all points $x \in \mathbb{W}_t$ through the line $x_2 = -(\Delta_2 + \Delta_3 - q + 1)$ (the line containing the minor diagonal of $\mathbb{R}_{q+1}$) and equate $\mathscr{M}_{q-t+1}$ and $\mathscr{M}_t$ at corresponding points. This yields

$$\mathscr{M}_{q-t+1}(\overline{x}_1,\overline{x}_2,\overline{x}_3) = \mathscr{M}_t(x_1,x_2,x_3), \quad \text{for } x \in \mathbb{L}^+ \text{ and } \overline{x} \in \mathbb{L}^+, \tag{8.46a}$$

where

$$\begin{aligned} x_1 &\mapsto \overline{x}_1 = -x_3 + \Delta_2 + \Delta_3 - q + 1, \\ x_2 &\mapsto \overline{x}_2 = -x_2 - 2(\Delta_2 + \Delta_3 - q + 1), \\ x_3 &\mapsto \overline{x}_3 = -x_1 + \Delta_2 + \Delta_3 - q + 1. \end{aligned} \tag{8.46b}$$

[Reflection through $x_2 = -(\Delta_2 + \Delta_3 - q + 1)$ yields the same transformation $(x_1,x_2,x_3) \mapsto (\overline{x}_1,\overline{x}_2,\overline{x}_3,)$ as the substitution $t \mapsto q - t + 1$, $(w_1,w_2,w_3) \mapsto (-w_3,-w_2,-w_1)$ in Eqs. (8.37a).]

The existence of the symmetry (8.46) of the functions $\{\mathscr{M}_t | t = 1,...,q\}$ suggests we consider this symmetry for the functions $\{m_t | t = 1,...,q\}$. The result we obtain is the following lemma.

Lemma 8.7: The symmetry property

$$m_{q-t+1}(\overline{x}_1,\overline{x}_2,\overline{x}_3) = m_t(x_1,x_2,x_3), \quad \text{for } x \in \mathbb{L}^+ \text{ and } \overline{x} \in \mathbb{L}^+ \tag{8.47a}$$

is true for $t = 1,...,q$ if and only if

$$m_t(x) = 0, \quad \text{for all } x \in \mathbb{L}^+ - \mathbb{W}_t, \tag{8.47b}$$

for $t = 1,...,q$.

Proof: If $m_t(x) = 0$ for all $x \in \mathbb{L}^+ - \mathbb{W}_t$, the symmetry property is clearly true, since the assumption implies $m_t(x) = \mathscr{M}_t(x)$ all $x \in \mathbb{L}^+$ and each $\mathscr{M}_t(x)$ has the symmetry (8.47a).

To show that the symmetry relation (8.47a) implies relation (8.47b), we use Lemma 8.3, which yields

$$m_t(x) = 0, \quad \text{for } x \in \mathbb{L}^+ - \mathbb{W}_t, \quad x_2 \leqslant -(\Delta_2 + \Delta_3 - q + 1), \tag{8.48a}$$

each $t = 1,...,q$. We next apply relation (8.47a) to this relation and obtain

$$m_{q-t+1}(\overline{x}) = 0, \quad \text{for } x \in \mathbb{L}^+ - \mathbb{W}_t$$

with $x_2 \leqslant -(\Delta_2 + \Delta_3 - q + 1)$, each $t = 1,...,q$. Equivalently, this result may be written as

$$m_t(x) = 0, \quad \text{for } x \in \mathbb{L}^+ - \mathbb{W}_t, \quad x_2 \geqslant -(\Delta_2 + \Delta_3 - q + 1), \tag{8.48b}$$

each $t = 1,...,q$. Together, Eqs. (8.48) yield the "only if" part of the lemma. ■

A somewhat more restrictive result is the following [we show in Subsection B that the hypothesis of this lemma is valid (hence, the conclusion still holds) for all $x \in \mathbb{W}_t$].

Lemma 8.8: For each $t = 1,...,q+1$, assume that at each $x \in \mathbb{L}^+$ such that $m_t(x) \neq 0$ the relation $r_t(x) = 1$ is true. Then

$$m_t(x) = \mathscr{M}_t(x), \quad \text{for } x \in \mathbb{W}_t,$$

each $t = 1,...,q+1$.

Proof: We use the recurrence relation (8.10) and Theorem 8.1 repeatedly. Thus, for points x such that $m_1(x) \neq 0$, we have $m_1(x) = l_1(x) - 1 = \mathscr{M}_1(x)$ for $x \in \mathbb{W}_1$; then, for points x such that $m_2(x) \neq 0$, we have $m_2(x) = m_1(x) + l_2(x) - 1 = \mathscr{M}_2(x)$ for all $x \in \mathbb{W}_2$, etc. We thus establish that $m_t(x)$ is nonzero and equal to $\mathscr{M}_t(x)$ for all $x \in \mathbb{W}_t$. ■

B. Zeros of G^t_q implied by symmetries and the reduction formula

The determinantal symmetry of G^t_q has been proved in the earlier sections and the reduction formula (4.15) established. Although we choose $\Delta_i \geqslant q$ $(i = 1,2,3)$ in $G^t_q(\Delta;x)$, it may happen that a determinantal symmetry, which yields, say, the transformation

$$\Delta_i \to \Delta'_i, \quad x_i \to x'_i, \quad i = 1,2,3,$$

and correspondingly,

$$G^t_q(\Delta;x) = G^t_q(\Delta';x'),$$

yields also a set of $\lambda' = (\lambda'_1,\lambda'_2,\lambda'_3)$ that are not all zero. In this case, the reduction formula (4.15) may be applied to $G^t_q(\Delta';x')$ with zeros of the original $G^t_q(\Delta;x)$ function appearing as zeros of the linear factors (now in the x'_i) in Eq. (4.15).

Application of the preceding technique separately to (i) transposition followed by the permutation

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

of rows, and (ii) column 1–column 3 interchange of the array

A [Eq. (5.1)] yields the following results, where the conditions $\Delta_i \geqslant q$ are imposed.

(i) Transposition followed by the permutation

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

of rows:

$$\Delta_1' = \Delta_3 - x_1, \quad \Delta_2' = \Delta_1, \quad \Delta_3' = \Delta_2 + x_1;$$
$$x_1' = -x_2, \quad x_2' = -x_1, \quad x_3' = -x_3; \tag{8.49a}$$
$$\lambda_1' = q - \Delta_3 + x_1, \quad \lambda_2' = \lambda_3' = 0;$$

where x_1 is restricted to integer values such that

$$x_1 \geqslant \Delta_3 - q. \tag{8.49b}$$

In final form, we have

$$\begin{aligned} &G_q^t(\Delta_1,\Delta_2,\Delta_3;\Delta_3 - k, -x_3 - \Delta_3 - k, x_3) \\ &= (-1)^{t(q-k)} \prod_{s=1}^{t} (\Delta_1 - q + k - s + 2)_{q-k} \\ &\quad \times (\Delta_2 + \Delta_3 - q - s + 2)_{q-k} \\ &\quad \times \prod_{s=1}^{t} (x_3 - \Delta_2 + s - 1)_{q-k} \\ &\quad \times (-x_3 - \Delta_3 - \Delta_1 + k + s - 1)_{q-k} \\ &\quad \times G_k^t(\Delta_1 - q + k, \Delta_2, \Delta_3; \\ &\qquad \Delta_3 - q, -x_3 - \Delta_3 + k, x_3 + q - k), \end{aligned} \tag{8.50a}$$

where we have put $x_1 = \Delta_3 - k$, eliminated x_2, and used determinantal symmetry to replace $G_k^t(q,\Delta_1 - q + k, \Delta_2 + \Delta_3 - q; x_3 + \Delta_3 - k, q - \Delta_3, -x_3 - q + k)$, which appears in the direct application of Eq. (4.15), by the $G_k^t(\cdots)$ given. For specified $t \in \{1,2,\ldots,q+1\}$, the domain of k is

$$k = t-1, t, \ldots, q. \tag{8.50b}$$

(ii) Column 1–Column 3 interchange:

$$\Delta_1' = \Delta_3 - x_1, \quad \Delta_2' = \Delta_1 - x_2, \quad \Delta_3' = \Delta_2 - x_3;$$
$$x_i' = -x_k - \Delta_i + \Delta_j, \quad (ijk)\text{cyclic}; \tag{8.51a}$$
$$\lambda_1' = q - \Delta_3 + x_1, \quad \lambda_2' = 0, \quad \lambda_3' = q - \Delta_2 + x_3;$$

where x_1 and x_3 are restricted to integer values such that

$$x_1 \geqslant \Delta_3 - q, \quad x_3 \geqslant \Delta_2 - q. \tag{8.51b}$$

Then

$$\begin{aligned} &G_q^t(\Delta_1,\Delta_2,\Delta_3;\Delta_3 - q + j, -\Delta_2 - \Delta_3 + k + q, \Delta_2 - k - j) \\ &= (-1)^{t(q-k)} \\ &\quad \times \prod_{s=1}^{t} \frac{(q-s+1)!(\Delta_1 + \Delta_2 + \Delta_3 - 2q - s + 2)_{q-k}}{(k-s+1)!} \\ &\quad \times \prod_{s=1}^{t} (-\Delta_1 + s - 1)_j (-\Delta_3 + s - 1)_{q-j-k} \\ &\quad \times (-\Delta_2 - \Delta_3 + q - j + s - 1)_j (-\Delta_1 - \Delta_2 \\ &\quad + k + j + s - 1)_{q-j-k} \\ &\quad \times G_k^t(\Delta_1 - j, \Delta_2, \Delta_3 - q + j + k; \\ &\qquad \Delta_3 - q, -\Delta_2 - \Delta_3 + 2q - j, \Delta_2 - q + j). \end{aligned} \tag{8.52a}$$

Again, as in Eq. (8.50a), we have used determinantal symmetry to obtain $G_k^t(\cdots)$ in this form. For specified $t \in \{1,\ldots,q+1\}$, the domain of k is

$$k = t-1, \ldots, q, \tag{8.52b}$$

and, for each such k, the integer j has domain

$$j = 0, 1, \ldots, q-k. \tag{8.52c}$$

Let us next consider the implications of Eqs. (8.50) and (8.52).

Consider first the linear factors

$$\prod_{s=1}^{t} (x_3 - \Delta_2 + s - 1)_{q-k} \tag{8.53a}$$

from Eq. (8.50a). We find that

$$G_q^t(\Delta;x) = 0, \tag{8.53b}$$

for all points x given by

$$x = (\Delta_3 - k, -x_3 - \Delta_3 + k, x_3), \tag{8.53c}$$

with

$$x_3 = \Delta_2, \Delta_2 - 1, \ldots, \Delta_2 - (q + t - 2 - k), \tag{8.53d}$$

where k may be any of the values

$$k = t-1, t, \ldots, q-1. \tag{8.53e}$$

Observe that the points given by Eqs. (8.53c) and (8.53d) are just the points $x \in \mathbb{W}_t$ having $x_1 = \Delta_3 - k$. Moreover, the multiplicity of a zero at point x originating from the linear factors (8.53a) alone is *exactly* $\mathscr{M}_t(x)$. The second product of linear factors in x_3 occurring in Eq. (8.50a) has no zeros for any x in the domain given by Eqs. (8.53c)–(8.53e). Since $\Delta_i \geqslant q$, there are similarly no zeros arising from the linear factors in the Δ_i for any values of these parameters. However, the G_k^t term might possess zeros in the domain (8.53c)–(8.53e), so that we can only conclude

$$m_t(x) \geqslant \mathscr{M}_t(x), \quad \text{for } x \in \mathbb{W}_t \text{ and } x_1 = \Delta_3 - k, \tag{8.54a}$$

each $k = t-1, t, \ldots, q-1$. However, since $G_{t-1}^t = 1$, we have equality in Eq. (8.54a) for $k = t-1$:

$$m_t(x) = \mathscr{M}_t(x), \quad \text{for } x \in \mathbb{W}_t \text{ and } x_1 = \Delta_3 - t + 1. \tag{8.54b}$$

Indeed, for $k = t-1$, we obtain from Eqs. (8.50) the following fully explicit expression:

$$\begin{aligned} G_q^t(\Delta;x) &= (-1)^{t(q-t+1)} \prod_{s=1}^{t} (\Delta_1 - q + t - s + 1)_{q-t+1} \\ &\quad \times (\Delta_2 + \Delta_3 - q - s + 2)_{q-t+1} \\ &\quad \times \prod_{s=1}^{t} (x_3 - \Delta_2 + s - 1)_{q-t+1} \\ &\quad \times (-x_3 - \Delta_3 - \Delta_1 + t + s - 2)_{q-t+1}, \end{aligned} \tag{8.55a}$$

for all points x of the form

$$(x_1,x_2,x_3) = (\Delta_3 - t + 1, -x_3 - \Delta_3 + t - 1, x_3), \quad x_3 \in \mathbb{R}. \tag{8.55b}$$

We can, of course, apply index permutations $1 \to i$, $2 \to j$, $3 \to k$ to Eqs. (8.50) and (8.55) and obtain still other expressions for $G_q^t(\Delta;x)$. For example, applying the permutation $1 \to 3$, $2 \to 1$, $3 \to 2$ yields

$$m_t(x) \geqslant \mathscr{M}_t(x), \quad \text{for } x \in \mathbb{W}_t \text{ and } x_3 = \Delta_2 - k, \tag{8.56}$$

each $k = t-1, t, \ldots, q$, with equality in this result for

$k = t - 1$; that is, for $x_3 = \Delta_2 - t + 1$. The importance of the preceding results based on Eqs. (8.50) (and the equivalent results obtained by index permutations) is in showing that the explicit polynomials $G_q^t(\Delta;x)$ possess zeros which in position and multiplicity are in complete accord with results implied by null space properties alone.

Let us turn next to Eqs. (8.52). Here the result we obtain is quite different from that found above in that we obtain information as to where the polynomial is *nonzero*. Thus, choosing $k = t - 1$ in Eqs. (8.52) yields

$$G_q^t(\Delta;x) = (-1)^{t(q-t+1)} \prod_{s=1}^{t} \frac{(q-s+1)!(\Delta_1+\Delta_2+\Delta_3-2q-s+2)_{q-t+1}}{(t-s)!} \tag{8.57a}$$
$$\times \prod_{s=1}^{t} (-\Delta_1+s-1)_j(-\Delta_3+s-1)_{q-t+1-j}$$
$$\times(-\Delta_2-\Delta_3+q-j+s-1)_j(-\Delta_1-\Delta_2+t+j+s-2)_{q-t+1-j},$$

for all points x given by

$$(x_1,x_2,x_3) = (\Delta_3 - q + j, -\Delta_2 - \Delta_3 + q + t - 1,$$
$$\Delta_2 - t + 1 - j),\ j = 0,1,\dots,q-t+1. \tag{8.57b}$$

Since $G_q^t(\Delta;x) \neq 0$ for all such points x, we find

$$m_t(x) = 0, \tag{8.58}$$

for all points x given by Eq. (8.57b).

The points x given by Eq. (8.57b) for which we have proved $m_t(x) = 0$ are the lattice points on the x_2-line segment above and adjacent to the upper boundary of the region $\mathbb{W}_t$ with one point on each end lying *outside* the rhombic region $\mathbb{R}_{q+1}$. In order to account for these two points below, it is convenient to define the region $\mathbb{W}_t^*$ by

$$\mathbb{W}_t^* = \mathbb{W}_t \cup \left\{\begin{matrix}\text{set of lattice points one} \\ \text{unit from boundary of } \mathbb{W}_t\end{matrix}\right\}. \tag{8.59}$$

Combining the important property (8.58) with results from Subsection A, we can now prove the following theorem.

Theorem 8.2: The multiplicity $m_t(x)$ of a zero of $G_q^t(\Delta;x)$ at each point $x \in \mathbb{W}_t^*$ is given by

$$m_t(x) = \mathscr{M}_t(x). \tag{8.60}$$

Proof: Define the line segment $\mathbf{l}_t^*$ by

$$\mathbf{l}_t^* = \left\{ x \in \mathbb{L}^+ \middle| \begin{matrix} x_1 \leqslant \Delta_3 - t + 1,\ x_3 \leqslant \Delta_2 - t + 1, \\ x_2 = -(\Delta_2 + \Delta_3 - q - t + 1 \end{matrix} \right\}.$$

Equation (8.58) may then be written

$$m_t(x) = 0, \quad \text{for } x \in \mathbf{l}_t^*.$$

We also obtain the following relation from Eqs. (8.12) and (8.19c):

$$\sum_{s=1}^{q+1} r_s(x) = q + t + 1, \quad \text{for } x \in \mathbf{l}_t^*. \tag{8.61}$$

Next, we use the recursion relation (8.10) and Fig. 9 successively for $t = 1,2,\dots$ in the steps that follow.

For $t = 1$ and $x \in \mathbf{l}_1^*$, we obtain $m_1(x) = l_1(x) - r_1(x) = 0$; that is, $r_1(x) = l_1(x) = 2$ (Fig. 9 for $t = 1$), and from Eq. (8.61), $r_2(x) = \cdots = r_{q+1}(x) = 1$. Next, for $x \in \mathbf{l}_1^*$, we have $m_2(x) = l_2(x) - r_2(x) = l_2(x) - 1 = \mathscr{M}_2(x)$ (Fig. 9 for $t = 1$). Since we have already proved (Lemma 8.5) that $m_2(x) = \mathscr{M}_2(x)$ for $x_2 \leqslant -(\Delta_2 + \Delta_3 - q + 1)$, we conclude that $m_2(x) = \mathscr{M}_2(x)$ for all $x \in \mathbb{W}_2^*$.

For $t = 2$ and $x \in \mathbf{l}_2^*$, we obtain $m_2(x) = m_1(x) + l_2(x) - r_1(x) = 0$; that is, $r_1(x) + r_2(x) = l_1(x) + l_2(x) = 4$ (Fig. 9 for $t = 1,2$), and from Eq. (8.61), $r_3(x) = \cdots = r_{q+1}(x) = 1$. Next, for $x \in \mathbf{l}_2^*$, we have $m_3(x) = l_3(x) - r_3(x) = l_3(x) - 1 = \mathscr{M}_3(x)$ (Fig. 9 for $t = 3$). Moreover, for $x \in \mathbf{l}_1^*$, we have $m_3(x) = m_2(x) + l_3(x) - r_3(x) = \mathscr{M}_2(x) + l_3(x) - 1 = \mathscr{M}_3(x)$. Thus, we have proved $m_3(x) = \mathscr{M}_3(x)$ for either $x \in \mathbf{l}_1^*$ or $x \in \mathbf{l}_2^*$. Since we have already proved (Lemma 8.5) that $m_3(x) = \mathscr{M}_3(x)$ for $x_2 \leqslant -(\Delta_2 + \Delta_3 - q + 1)$, we conclude that $m_3(x) = \mathscr{M}_3(x)$ for all $x \in \mathbb{W}_3^*$.

Continuing in this manner (induction), we complete the proof. ∎

Remarks: (a) The general step in the above proof establishes the relations

$$\sum_{s=1}^{k} r_s(x) = 2k, \quad r_{k+1}(x) = \cdots = r_{q+1}(x) = 1, \tag{8.62}$$

for $x \in \mathbf{l}_k^*$, each $k = 1,2,\dots q$. In particular, selecting $r_t(x)$ from this set yields

$$r_t(x) = 1, \quad \text{for all } x \in \mathbb{W}_t^* \text{ and } x_2 \geqslant -(\Delta_2 + \Delta_3 - q). \tag{8.63}$$

This result and Eq. (8.26b) validate the hypothesis of Lemma 8.8 for points $x \in \mathbb{W}_t$; hence, this lemma also proves Theorem 8.2 for points $x \in \mathbb{W}_t$.

(b) Theorem 8.2 establishes that $m_t(x) = \mathscr{M}_t(x)$ only for points $x \in \mathbb{W}_t^*$, but does not prove $m_t(x) = \mathscr{M}_t(x) = 0$ for all $x \in \mathbb{L}^+ - \mathbb{W}_t^*$.

(c) The result given by Eqs. (8.55) is quite significant for it gives the polynomial explicitly at all points on the line $x_1 = \Delta_2 - t + 1$ and shows that the degree is $2t(q - t + 1)$ separately in each of the variables x_3, Δ_1, Δ_2, Δ_3.

Equations (8.50), a similar relation given below [Eqs. (8.64)], and Theorems 8.1 and 8.2 themselves may now be used to extend further the domain for which the relation $m_t(x) = \mathscr{M}_t(x)$ is valid. Before carrying this out, let us outline the proof of the following relation:

$$G_q^t(\Delta_1,\Delta_2,\Delta_3; x_1, -x_1 - \Delta_2 + k, \Delta_2 - k)$$
$$= (-1)^{t(q-k)} \prod_{s=1}^{t} (\Delta_3 - q + k - s + 2)_{q-k}$$
$$\times (\Delta_1 + \Delta_2 - q - s + 2)_{q-k}$$
$$\times \prod_{s=1}^{t} (x_1 - \Delta_3 + s - 1)_{q-k}$$
$$\times(-x_1 - \Delta_1 - \Delta_2 + k + s - 1)_{q-k}$$
$$\times G_k^t(\Delta_1,\Delta_2,\Delta_3 - q + k; x_1, -x_1 - \Delta_2 + k, \Delta_2 - q), \tag{8.64a}$$

for

$$k = t-1, t, \dots, q. \tag{8.64b}$$

This relation is proved by first applying the determinantal symmetry consisting of the row permutation

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

followed by transposition of the array A corresponding to the variables in $G_q^t(\cdots)$, and second by applying the reduction formula (4.15). This leads to Eqs. (8.64), except for the stated form of $G_k^t(\cdots)$, which is obtained by a further application of determinantal symmetry.

We can now extend the results given in Theorems 8.1 and 8.2 as follows.

Theorem 8.3: Define the (finite) subdomains $\mathbb{A}_{1t}$, $\mathbb{A}_{3t}$, and $\mathbb{A}_t = \mathbb{A}_{1t} \cup \mathbb{A}_{3t}$ of $\mathbb{L}$ by

$$\mathbb{A}_{1t} = \left\{ x \in \mathbb{L} \,\middle|\, \begin{matrix} \max(0, \Delta_3 - \Delta_2) \leqslant x_1 \leqslant \Delta_3 - q \\ x_2 \geqslant -(\Delta_2 - \Delta_3 - q),\ x_3 \geqslant \Delta_2 - t + 1 \end{matrix} \right\}, \tag{8.65a}$$

$$\mathbb{A}_{3t} = \left\{ x \in \mathbb{L} \,\middle|\, \begin{matrix} \max(0, \Delta_2 - \Delta_1) \leqslant x_3 \leqslant \Delta_2 - q \\ x_2 \geqslant -(\Delta_2 - \Delta_3 - q),\ x_1 \geqslant \Delta_3 - t + 1 \end{matrix} \right\}. \tag{8.65b}$$

Then

$$m_t(x) = \mathscr{M}_t(x), \quad \text{for } x \in (\mathbb{L}^+ - \mathbb{N}'_{q+1}) \cup \mathbb{W}_t^* \cup \mathbb{A}_t. \tag{8.65c}$$

Proof: The result given by Eq. (8.65c) adjoins the set of lattice points $\mathbb{A}_t$ to those points x for which the relation $m_t(x) = \mathscr{M}_t(x)$ has already been proved in Theorems 8.1 and 8.2. Thus, we need to show that $m_t(x) = \mathscr{M}_t(x)$ for $x \in \mathbb{A}_t$. [We define $\mathscr{M}_t(x) = 0$ should any points x having $x_1 = 0$ or $x_3 = 0$ belong to $\mathbb{A}_t$.]

Consider Eq. (8.50a). The linear factors multiplying G_k^t account exactly in position and multiplicity for the zeros of $G_q^t(\Delta;x)$ for all points $x \in \mathbb{W}_t$ given by

$$x = (\Delta_3 - k, -x_3 - \Delta_3 + k, x_3), \tag{8.66a}$$

$$x_3 = \Delta_2, \Delta_2 - 1, \dots, \Delta_2 - (q + t - k - 2). \tag{8.66b}$$

Since $m_t(x) = \mathscr{M}_t(x)$ for all $x \in \mathbb{W}_t$, it follows for all points given by Eqs. (8.65) that

$$G_k^t(\Delta_1 - q + k, \Delta_2, \Delta_3; \Delta_3 - q, \\ -x_3 + k - \Delta_3, x_3 + q - k) \neq 0. \tag{8.67a}$$

We observe next that this relation is valid for all integers $q,k,t,\Delta_1,\Delta_2,\Delta_3,x_3$ that obey the conditions

$$1 \leqslant t \leqslant k+1 \leqslant q,$$
$$\Delta_i \geqslant q \quad (i = 1,2,3), \tag{8.67b}$$
$$\Delta_2 - (q - t - k - 2) \leqslant x_3 \leqslant \Delta_2.$$

We next define Δ_1' and x_3' by $\Delta_1' = \Delta_1 - q + k$ and $x_3' = x_3 + q - k$. Using these definitions in Eq. (8.67a) and the validity of that relation for generic variables as described above, we now rename k to be q, q to be k, Δ_1' to be Δ_1, and x_3' to be x_3, thus proving

$$G_q^t(\Delta_1, \Delta_2, \Delta_3; \Delta_3 - k, \ -x_3 + k - \Delta_3, \ x_3) \neq 0. \tag{8.68a}$$

The domains of the parameters in this relation are obtained from Eqs. (8.66) and (8.67) and the definitions of Δ_3' and x_3' to be $1 \leqslant t \leqslant q+1 \leqslant k$, $\Delta_1 \geqslant q$, $\Delta_2 \geqslant k$, $\Delta_3 \geqslant k$, $\Delta_2 - t + 2 \leqslant x_3 \leqslant \Delta_2 + k - q$. We also verify relation (8.68a) for $k \geqslant q$, $x_3 = \Delta_2 - t + 1$ directly from Eq. (8.64a), and for $k = q$, $\Delta_2 - t + 1 \leqslant x_3 \leqslant \Delta_2$ from Theorem 8.2. Thus, for given $q \geqslant 1$ and $\Delta_i \geqslant q$ $(i = 1,2,3)$, relation (8.68a) is true for all k and x_3 such that

$$k = q, q+1, \dots, \min(\Delta_2, \Delta_3), \tag{8.68b}$$

$$\Delta_2 - t + 1 \leqslant x_3 \leqslant \Delta_2 + k - q. \tag{8.68c}$$

These relations and Eq. (8.68a) yield $G_q^t(\Delta;x) \neq 0$ for $x \in \mathbb{A}_{1t}$.

Similarly, beginning with Eqs. (8.64) and following the steps of the preceding argument, we obtain

$$G_q^t(\Delta_1, \Delta_2, \Delta_3; x_1, -x_1 - \Delta_2 + k, \Delta_2 - k) \neq 0, \tag{8.69a}$$

for

$$k = q, q+1, \dots, \min(\Delta_1, \Delta_2),$$
$$\Delta_3 - t + 1 \leqslant x_1 \leqslant \Delta_3 + k - q. \tag{8.69b}$$

Equivalently, we have $G_q^t(\Delta;x) \neq 0$ for $x \in \mathbb{A}_{3t}$. ■

Remark: Relations such as Eqs. (8.50), (8.52), and (8.64) may be used to show that $G_q^t(\Delta;x) \neq 0$ implies $G_k^t(\Delta';x') \neq 0$ (and conversely) for various $(\Delta;x)$ and $(\Delta';x')$, but so far a proof (if it is true) that $m_t(x) = \mathscr{M}_t(x)$ for all $x \in \mathbb{L}^+$ has not been possible.

IX. SUMMARY AND CONCLUDING REMARKS

In this concluding section, we summarize and discuss the main results obtained in this paper.

Beginning with a proof (Introduction) that certain U(3):U(2) projective operators must be the zero operator in consequence of the definition of a canonical unit tensor operator in terms of null space properties alone, we used these splitting conditions to construct explicitly (in Sec. II) the denominator function D_{Γ_t} in a determinantal form [Eqs. (2.22)–(2.24)]. [We also obtained in Eqs. (2.5a), (2.16), and (2.22)–(2.25) explicit expressions for certain Racah coefficients and the matrix elements of certain U(3):U(2) projective operators.] Since this denominator function is a U(3)-invariant characterization of the associated canonical unit tensor operator, we have focussed in the present paper on the delineation of its properties in order to bring it to a more comprehensible form. Significant results obtained here are (i) the expression [Sec. III, Eq. (3.3)] of the denominator functions $D^2_{\Gamma_t}$, $t = 1,2,\dots,\mathscr{M}$, in terms of successive ratios of a family of new functions $G_q^0, G_q^1, \dots, G_q^{\mathscr{M}}$, which turn out to be polynomials, and (ii) a reduction formula [Sec. IV, Eq. (4.15)] for the G_q^t functions, which is a key relation for establishing further properties of these functions.

Using the results obtained in Sec. II–IV, we then preceeded to demonstrate in Sec. V-VIII three significant properties of the functions $G_q^t(\Delta;x)$(taking $\Delta_i \geqslant q$, for simplicity).

(i) Symmetry: each function $G_q^t(\Delta;x)$ has symmetries in the parameters $(\Delta_1,\Delta_2,\Delta_3,x_1,x_2,x_3)$ that correspond to the 72 determinantal symmetries of a 3×3 array A associated with these parameters [see Eq. (5.1)].

(ii) Polynomial: each $G_q^t(\Delta;x)$ is a polynomial of total degree $2t(q - t + 1)$ in the variables (x_1,x_2,x_3) [and almost certainly also in the variables $(\Delta_1,\Delta_2,\Delta_3)$].

(iii) Zeros: each $G_q^t(\Delta,x)$ has zeros in the variables (x_1,x_2,x_3) which in position and multiplicity coincide exactly with that of the weight space of the U(3) irrep $[q-t,0,-t+1]$ (equivalently of the SU(3) irrep $[q-1,t-1,0]$).

As is evident from the lengthy developments given in the paper, the above properties of the denominator function have not obtained easily. Nonetheless, the simplicity and elegance of the final results justify, in our opinion, this effort. It is quite reasonable now to expect that the algebraic expressions for the general SU(3) Racah coefficients and the matrix elements of the general U(3):U(2) projective operators will show similar simplification. In addition to the many known applications of these coefficients to physical systems possessing SU(3) symmetry, we believe that these coefficients will possess, in analogy to SU(2) (see Ref. 1), a wealth of information for special functions and their generalizations (see Refs. 12 and 16).

In concluding, let us note that we have not given here the fully explicit polynomial form for $G_q^t(\Delta;x)$. This step is itself of considerable interest (and difficulty) and will be given in a second paper.

ACKNOWLEDGMENTS

Two of us (LCB and MAL) were supported in part by the National Science Foundation. The work of two of us (LCB) and (JDL) was performed under the auspices of the U. S. Department of Energy under Contract No. W-7405-ENG-36 and the Office of Basic Energy Sciences, Department of Applied Mathematics.

ENG-36 and the Office of Basic Energy Sciences, Department of Applied Mathematics.

[1]L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics*, Vol. 8, *The Racah-Wigner Algebra in Quantum Theory*, Vol. 9, *Encylopedia of Mathematics and Its Applications*, edited by G-C. Rota (Addison–Wesley, Reading, MA, 1981.)
[2]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[3]L. Michel, "Application of Group Theory to Quantum Physics: Algebraic Aspects," in *Lecture Notes in Physics*, Battelle Rencontres, edited by V. Bargmann (Springer, Berlin, 1970), pp. 36–143.
[4]J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970).
[5]J. D. Louck, M. A. Lohe, and L. C. Biedenharn, J. Math. Phys. **16**, 2408 (1975).
[6]L. C. Biedenharn and J. D. Louck, J. Math. Phys. **13**, 1985 (1972).
[7]L. C. Biedenharn, A. Giovannini, and J. D. Louck, J. Math. Phys. **8**, 691 (1967).
[8]L. C. Biedenharn, J. D. Louck, E. Chacón, and M. Ciftan, J. Math. Phys. **13**, 1957 (1972).
[9]D. E. Flath and L. C. Biedenharn, "Beyond the Enveloping Algebra of sl_3," preprint, 1982 (submitted for publication in Can. J. Math.).
[10]L. C. Biedenharn and D. E. Flath, Commun. Math. Phys. **93**, 143 (1984).
[11]L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. **8**, 89 (1968).
[12]R. A. Gustafson and S. C. Milne, Adv. Appl. Math. **4**, 422 (1983); see also cited references.
[13]L. C. Biedenharn, M. A. Lohe, and J. D. Louck, "The Canonical Resolution of the Multiplicity Problem for SU(3)," *in Proceedings of the Fourth International Colloquium on Group Theoretical Methods in Physics*, Nijmegen, The Netherlands, 1975, edited by A. Janner, T. Janssen, and M. Boon (Springer, New York, 1976), pp. 395–403.
[14]M. A. Lohe, L. C. Biedenharn, and J. D. Louck, J. Math. Phys. **18**, 1883 (1977).
[15]J. A. Castilho-Alcarás, L. C. Biedenharn, K. T. Hecht, and G. Neely, Ann. Phys. (N.Y.) **60**, 85 (1970).
[16]J. D. Louck and L. C. Biedenharn, J. Math. Anal. Appl. **59**, 423 (1977).

Algebraic formulas for some nontrivial U_n *6j* symbols and $U_{mn} \supset U_m \times U_n$ *3jm* symbols

Richard W. Haase
Institut für Theoretische Physik, Technische Universität Wien, Karlsplatz 13, A-1040 Vienna, Austria
Philip H. Butler
Physics Department, University of Canterbury, Christchurch, New Zealand

(Received 3 October 1984; accepted for publication 11 January 1985)

We give tables of algebraic formulas for some nontrivial *6j* symbols and *3jm* symbols of the unitary groups. The tables demonstrate that the building-up method can be used successfully to obtain the rank dependence of unitary group *j* and *jm* symbols. To emphasize the rank-dependent nature of this calculation, we have employed the composite Young tableaux notation (or back-to-back notation) to label the unitary group irreps. In using this notation, the transpose conjugate symmetry of the corresponding composite Young diagram leads to a new symmetry of the unitary group *6j* and *3jm* symbols. The transposition of the groups U_m and U_n gives rise to a further symmetry of the *3jm* symbols of $U_{mn} \supset U_m \times U_n$.

I. INTRODUCTION

The purpose of this paper is to present rank-dependent algebraic formulas for some *6j* symbols of U_n and some *3jm* symbols of $U_{mn} \supset U_m \times U_n$. The *6j* table includes all primitive *6j* symbols containing at most a power 3 irrep and all nonprimitive *6j* symbols containing power 2 irreps. All primitive *3jm* symbols of $U_{mn} \supset U_m \times U_n$ up to and including power 3 and power 2 nonprimitive *3jm* symbols are given. In addition, *6j* and *3jm* symbols with product multiplicity have been calculated. No branching multiplicity occurs in the reduction $U_{mn} \supset U_m \times U_n$ of the irreps considered.

The unitary groups have been used extensively since Jahn's[1] extension of Racah's's[2] fractional parentage work in atomic spectroscopy. From the various and wide applications,[3–7] it becomes apparent that one requires a method for calculating the *6j* and *3jm* symbols (symmetrized Racah and Clebsch–Gordan coefficients, respectively) of the unitary groups in a rank-independent manner. Our aim is to show that the building-up method, familiar to nuclear physicists, can be employed sucessfully to such a problem. Indeed, it could be applied to the series of Lie groups O_n,SO_n,Sp_n. Such a calculation would satisfy the needs of particle physics, for instance, where different group-subgroup schemes are always appearing. The method has previously been used to calculate tables of *6j* symbols for SU_6 and SU_3 (Ref. 8) and *3jm* symbols for certain subgroup bases of SU_6 and SU_3. However, it is impractical to produce tables for each unitary group that arises. In addition, *6j* and *3jm* symbols of the point groups[9] and algebraic formulas for SO_3 *6j* symbols and $SO_3 \supset SO_2$ *3jm* symbols[10] have been calculated by this method.

One of the advantages of the building-up method over the ladder operator techniques (familiar from angular momentum theory and used by many workers for U_n) is that only a knowledge of character theory, namely product and branching rules, is required. No representation matrices are needed. This aspect is useful for groups with large dimensional irreps and, as we shall show, in exploiting the rank independence of series of groups such as U_n. The general properties underlying the method are given elsewhere.[9–11] An outline is given in Sec. IV.

To emphasize the rank independence we have employed the composite Young tableaux notation[12–17] (or back-to-back notation) to label the unitary group irreps. This is described in Sec. II along with other U_n group information. A particular symmetry of Young tableaux is the transpose conjugate, which is generated by the alternating irrep $[1^l]$ of S_l. In defining this symmetry for composite Young diagrams, we arrive at a new symmetry of the unitary group *6j* and *3jm* symbols. This symmetry is discussed in Sec. V. In Sec. VI we take a look at the symmetry of the $U_{mn} \supset U_m \times U_n$ *3jm* symbols under transposition of the groups U_n and U_m.

II. U_n GROUP INFORMATION

In this section we give a brief outline of the properties of the irreps of U_n. The irreps of U_n and of all compact Lie groups may be labeled by partitions. A partition of an integer l into p parts $\lambda_1\lambda_2,...,\lambda_p$ with $\lambda_1+\lambda_2+\cdots+\lambda_p=l$, is denoted $(\lambda_1,\lambda_2,...,\lambda_p)$ or merely λ, and is said to be regular if the parts also satisfy

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p \geqslant 0. \tag{2.1}$$

For U_n a full set of standard labels is obtained by using n nonincreasing parts λ_i as in (2.1) but relaxing the positivity of the parts. By separately taking the positive and negative parts as two regular partitions μ and ν, we obtain the composite label for U_n irreps

$$\begin{aligned}\{\lambda\} &= \{\lambda_1,\lambda_2,...,\lambda_n\}\\ &= \{\nu_1,\nu_2,...,\nu_q,0,...,0,\\ &\qquad -\mu_p,...,-\mu_2,-\mu_1\} \qquad (2.2a)\\ &= \{\mu;\nu\}, \quad \text{where } p+q \leqslant n. \qquad (2.2b)\end{aligned}$$

(For U_n irreps we enclose the label in braces.) The usual association between regular partitions and Young diagrams may be extended to cover those with negative parts by forming the composite Young diagram. This is obtained by reflecting the Young diagram of μ about the vertical and pair-

 0022-2488/85/071493-21$02.50

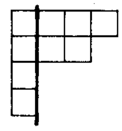
ing it "back-to-back" with that of ν. This construction has been used by Murnaghan[12] and Littlewood[13] but developed more fully by King and others.[14–17]

In (2.2) the composite labels $\{\mu;\nu\}$ satisfy the condition $p+q \leqslant n$. If the number of parts of $\{\mu;\nu\}$ exceeds n, it forms a nonstandard composite label. This may be standardized by the removal of continuous boundary hooks. The simpler partitions ρ and σ are obtained from μ and ν, respectively, by removing $(p+q-n-1)$ adjoining cells (called a hook) from the lower boundary of the corresponding Young diagrams. The hooks start at the foot of the first column of μ and ν and if they end in columns x and y, respectively, then

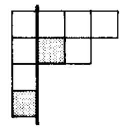
$$\{\rho;\sigma\} = (-)^{x+y+1}\{\mu;\nu\}. \tag{2.3}$$

The label vanishes (i.e., it labels a null irrep) unless the parts of ρ and σ are ordered as in definition (2.1). Repeated hook removals may be required to arrive at a standard composite label for U_n. As an example, the composite label $\{1^4;32\}$ is standard in U_n for $n>5$ since the hook $(h=p+q-n-1=4+2-n-1=5-n)$ is negative. However, for smaller values of n the label $(1^4;32)$ must be "modified"

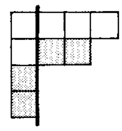
$n=5, \quad h=0, \quad \{1^4;32\} = -\{1^4;32\},$ i.e., $\{1^4;32\}=\varphi,$



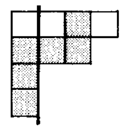
$n=4, \quad h=1, \quad \{1^4;32\}=\varphi,$



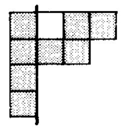
$n=3, \quad h=2, \quad \{1^4;32\} = +\{1^2;3\},$



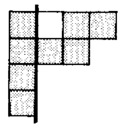
$n=2, \quad h=3, \quad \{1^4;32\}=\varphi,$



$n=1, \quad h=4, \quad \{1^4;32\} = -\{0;1\},$



$n=0, \quad h=5, \quad \{1^4;32\}=\varphi,$



(where φ is the null irrep). As can be seen from this example, the modification rule changes the irrep label in a nontrivial manner.

For more discussion the reader is referred to the literature.[15–17] By using the composite labels for U_n, many properties can be expressed in an n-independent manner. The complex conjugate pairs of irreps are labeled by composite labels with the partitions interchanged, $\{\mu;\nu\}=\{\nu;\mu\}^*$, and the power of the irrep $\{\mu;\nu\}$ is given as the sum $l+m$, where μ and ν are partitions of l and m, respectively. This simplicity is one of the principal reasons for using the composite labeling scheme.

Various formulas have been given for the dimensions of U_n irreps. Table I was derived using

$$\{\mu;\nu\} = N_n(\mu;\nu)/H(\mu)H(\nu), \tag{2.4}$$

$$N_n(\mu;\nu) = \prod_{\substack{i,j\\k,l}} (n-i-j+\mu_i+\nu_j+1) \times (n+k+l-\tilde{\mu}_k-\tilde{\nu}_l-1), \tag{2.5}$$

where the product ranges over all cells of the partitions of μ and ν. The denominator functions $H(\mu)$ and $H(\nu)$, which are Robinson's[18] hook formula, and $N_n(\mu;\nu)$ may be obtained diagrammatically.[19,20] The $N_n(\mu;\nu)$ gives explicitly the n dependence and takes the form of a factored polynomial of order $l+m$.

Under the transpose conjugate (or tilde) symmetry, where the partition $\tilde{\lambda}$ is obtained from λ by interchanging rows and columns of the Young diagram, we have

$$H(\tilde{\mu}) = H(\mu), \quad N_n(\tilde{\mu};\tilde{\nu}) = (-1)^{l+m}N_{-n}(\mu;\nu). \tag{2.6}$$

Hence the dimension formulas for $\{\tilde{\mu};\tilde{\nu}\}$ may be obtained from those of $\{\mu;\nu\}$ by replacing all factors $(n+a)$ by $(n-a)$ for positive or negative integers a. This symmetry of the dimension formulas, and, as we will see below, the n-independent and tilde invariance of the Kronecker product rule for U_n and branching rules $U_{mn} \supset U_m \times U_n$ (modulo modification rules) carry through into corresponding symmetries of the U_n $6j$ and $3jm$ symbols (see Secs. V and VI).

The Kronecker product rule for U_n is given as[17]

$$\{\mu;\nu\}\times\{\rho;\sigma\} = \sum_{\xi,\zeta}\{(\mu/\xi)\cdot(\rho/\zeta);(\nu/\zeta)\cdot(\sigma/\xi)\} = \sum m^{\tau;v}_{\mu;\nu\times\rho;\sigma}\{\tau;v\}, \tag{2.7}$$

where "/" and "·" are Schur function operations of division and outer multiplications and $m^{\tau;v}_{\mu;\nu\times\rho;\sigma}$ is the multiplicity of $\{\tau;v\}$ in the product of $\{\mu;\nu\}$ and $\{\rho;\sigma\}$. Tables of the Schur function operators are given in Ref. 20.

For a particular value of n, (2.7) may be subject to the modification rule. We illustrate this procedure by considering the Kronecker square of the adjoint (or generator) representation. Equation (2.7) gives

$$\begin{aligned}\{1;1\}\times\{1;1\} &= \sum_{\xi,\zeta}\{(1/\xi)\cdot(1/\zeta);(1/\zeta)\cdot(1/\xi)\}, \quad \xi,\zeta=0,1\\ &= \{2;2\}+\{2;1^2\}+\{1^2;2\}+\{1^2;1^2\}\\ &\quad +2\{1;1\}+\{0;0\}.\end{aligned} \tag{2.8}$$

For $n \geqslant 4$ the labels, as composite labels, are standard. For $n<4$, one applies the modification rule as appropriate to give

$$U_3{:}\{1;1\}\times\{1;1\} = \{2;2\}+\{2;1^2\}+\{1^2;2\}+\varphi + 2\{1;1\}+\{0;0\}, \tag{2.9}$$

$$\begin{aligned}U_2{:}\{1;1\}\times\{1;1\} &= \{2;2\}+\varphi+\varphi-\{1;1\} + 2\{1;1\}+\{0;0\}\\ &= \{2;2\}+\{1;1\}+\{0;0\},\end{aligned} \tag{2.10}$$

$$U_1{:}\varphi\times\varphi=\varphi-\varphi-\varphi-\{0\}+2\varphi+\{0\}$$
$$=\varphi, \tag{2.11}$$
$$U_0{:}(-\{0\})\times(-\{0\})=\varphi+\{0\}+\{0\}+\varphi$$
$$-2\varphi+\{0\}=\{0\}. \tag{2.12}$$

The U_2 result (2.10) illustrates that the modification rules give a natural n dependence to multiplicity separations.

We specify the Kronecker product rules by triads. Thus $(\{\mu;\nu\}\{\rho;\sigma\}\{\tau;\upsilon\}r)$ forms a triad if the identity irrep $\{0;0\}$ occurs at least $r+1$ times in the triple product $\{\mu;\nu\}\times\{\rho;\sigma\}\times\{\tau;\upsilon\}$. Note that the range of r is taken as $r+1=1,\ldots,m^{\tau;\upsilon}_{\mu;\nu\times\rho;\sigma}$; that is, r is initialized at zero. The transpose conjugate symmetry applied to the Kronecker product rule and to the Schur function operations allows one to prove that if $(\{\mu;\nu\}\{\rho;\sigma\}\{\tau;\upsilon\}r)$ is a triad then $(\{\mu;\nu\}\{\rho;\sigma\}\{\tau;\upsilon\}r)$ also exists as a triad for n large enough so that all partitions remain unmodified.

The branching rule for the reduction of composite labeled irreps of U_{mn} to $U_m\times U_n$ is given by the expression[17]

$$\{\mu;\nu\}\downarrow\sum_{\xi,\zeta,\eta,\sigma,\rho}(-1)^z\{((\mu/\sigma)\circ\xi)/\rho;((\nu/\tilde{\sigma})\circ\zeta)/\rho\}$$
$$\times\{\xi/\eta;\zeta/\eta\}$$
$$=\sum_{\rho,\sigma,\tau,\nu}m^{\mu;\nu}_{\rho;\sigma\times\tau;\upsilon}\{\rho;\sigma\}\times\{\tau;\upsilon\}, \tag{2.13}$$

where σ is a partition of z and "$\circ$" is the S-function operation of inner multiplication which defines the Kronecker products of irreps of the symmetric group. It is to be noted that the above expression gives a series of positive terms; however, it is not possible to demonstrate in general the cancellation of all negative terms.

In the special case when $\mu=0$ Eq. (2.13) simplifies to

$$\{0;\lambda\}=\sum_{\nu}\{0;\lambda\circ\nu\}\{0;\nu\}=\sum_{\mu,\nu}m^{\lambda}_{\mu\nu}\{0;\mu\}\{0;\nu\}. \tag{2.13'}$$

As an example of the use of Eq. (2.13) we reduce the adjoint representations

$$\{1;1\}\downarrow\sum(-1)^z\{((1/\sigma)\circ\xi)/\rho;((1/\sigma)\circ\zeta)/\rho\}$$
$$\times\{\xi/\eta;\zeta/\eta\}, \tag{2.14}$$

where σ is restricted to 0 and 1 by the division operation and ξ,ζ must be a partition of the same integer as $(1/\sigma)$ by the inner multiplication operators. Hence

$$\{1;1\}\downarrow\sum_{\rho,\eta}(+)\{(1\circ 1)/\rho;(1\circ 1)/\rho\}\times\{1/\eta;1/\eta\}$$
$$\times(-)\{(0\circ 0)/\rho;(0\circ 0)/\rho\}\times\{0/\eta;0/\eta\} \tag{2.15}$$
$$=[\{1/0;1/0\}+\{1/1;1/1\}]$$
$$\times[\{1/0;1/0\}+\{1/1;1/1\}]$$
$$+\{0/0;0/0\}\times\{0/0;0/0\}, \tag{2.16}$$

where ρ,η are also restricted to 0 and 1 by the division operation

$$\{1;1\}\downarrow[\{1;1\}+\{0;0\}]\times[\{1;1\}+\{0;0\}]$$
$$-\{0;0\}\times\{0;0\}$$
$$=\{1;1\}\times\{1;1\}+\{1;1\}\times\{0;0\}$$
$$+\{0;0\}\times\{1;1\}. \tag{2.17}$$

As with the Kronecker product rule, the branching rule also displays a similar transpose conjugate symmetry. If

$$\{\mu;\nu\}\supset m^{\mu;\nu}_{\rho;\sigma\times\tau;\upsilon}\{\rho;\sigma\}\times\{\tau;\upsilon\},$$
$$m^{\mu;\nu}_{\rho;\sigma\times\tau;\upsilon}>0 \tag{2.18}$$

(what we shall term a ket branching) then for m and n large enough the three following ket branchings also exist:

$$\{\mu;\nu\}\supset m^{\mu;\nu}_{\tilde{\rho};\tilde{\sigma}\times\tilde{\tau};\tilde{\upsilon}}\{\tilde{\rho};\tilde{\sigma}\}\times\{\tilde{\tau};\tilde{\upsilon}\}, \tag{2.19}$$
$$\{\tilde{\mu};\tilde{\nu}\}\supset m^{\tilde{\mu};\tilde{\nu}}_{\rho;\sigma\times\tilde{\tau};\tilde{\upsilon}}\{\rho;\sigma\}\times\{\tilde{\tau};\tilde{\upsilon}\}, \tag{2.20}$$
$$\{\tilde{\mu};\tilde{\nu}\}\supset m^{\tilde{\mu};\tilde{\nu}}_{\tilde{\rho};\tilde{\sigma}\times\tau;\upsilon}\{\tilde{\rho};\tilde{\sigma}\}\times\{\tau;\upsilon\}, \tag{2.21}$$

where all the branching multiplicities are equal.

A further symmetry of the branching rule is the symmetry arising from the transposition of the groups U_m and U_n. Again for large enough m and n, if the ket branching of (2.18) exists then we also have

$$\{\mu;\nu\}\supset m^{\mu;\nu}_{\tau;\upsilon\times\rho;\sigma}\{\tau;\upsilon\}\times\{\rho;\sigma\}, \tag{2.22}$$

with the same branching multiplicity. We call this symmetry the m–n transposition symmetry. For both this and the transpose conjugate symmetry the condition m and n large enough is to imply that the modification rule for irrep labels is not invoked. In the case when m and n are small, some of the ket branchings may have to be modified or may cancel with others, or may even be inadmissible. For example, in U_{mn} $(m>2,n>2)$

$$\{1;1^2\}\downarrow\{0;1\}\times\{0;1\}+\{0;1\}\times\{1;1^2\}$$
$$+\{0;1\}\times\{1;2\}+\{1;1^2\}\times\{0;1\}$$
$$+\{1;1^2\}\times\{1;2\}+\{1;2\}\times\{0;1\}$$
$$+\{1;2\}\times\{1;1^2\}, \tag{2.23}$$

while with $m=2$, $n>2$

$$\{1;1^2\}\downarrow\{0;1\}\times\{0;1\}+\{0;1\}\times\{1;1^2\}$$
$$+\{0;1\}\times\{1;2\}+\varphi\times\{0;1\}+\varphi\times\{1;2\}$$
$$+\{1;2\}\times\{0;1\}+\{1;2\}\times\{1;1^2\}, \tag{2.24}$$

and $m=1,n>2$

$$\{1;1^2\}\downarrow\{0;1\}\times\{0;1\}+\{0;1\}\times\{1;1^2\}$$
$$+\{0;1\}\times\{1;2\}-\{0;1\}\times\{0;1\}-\{0;1\}\times\{1;2\}$$
$$+\varphi\times\{0;1\}+\varphi\times\{1;1^2\}$$
$$\downarrow\{0;1\}\times\{1;1^2\}. \tag{2.25}$$

Note the two symmetries of Eqs. (2.18), (2.19), and (2.22) in Eq. (2.23). In later sections we will discuss these symmetries in relation to the algebraic formulas of the $3jm$ symbols of $U_{mn}\supset U_m\times U_n$. In addition, we shall look at the effects the modification rule has in determining the form of the algebraic formula of $3jm$ symbols.

Our final remarks on the unitary group concern the infinitely many one-dimensional irreps labeled by $\{a^n\}$ with

a taking positive and negative integer values. Such irreps are generated by the Kronecker powers of the one-dimensional determinantal irrep $\{1^n\}$ or its complex conjugate irrep $\{1^n\}^* = \{(-1)^n\}$. [Note we are using the notational form of (2.2a) to label the irreps.] The Kronecker product of any U_n irrep with the infinite series of one-dimensional irreps $\{a^n\}$ gives a series of "associated irreps"

$$\begin{aligned}\{\mu;\nu\}\times\{a^n\} &= \{\nu_1,\ldots,\nu_q,0,\ldots,0-\mu_p,\ldots,-\mu_1\}\times\{a^n\}\\ &= \{\nu_1+a,\ldots,\nu_q+a,a,\ldots,a,-\mu_p\\ &\quad +a,\ldots,-\mu_1+a\}\\ &= \{\mu';\nu'\}.\end{aligned} \tag{2.26}$$

Clearly the property of association is n dependent. We also note that from a series of associated irrep labels there is a unique label of the form $\{0;\lambda\}\equiv\{\lambda\}$ for which the nth part of the partition is zero and (λ) is a regular partition into $n-1$ parts. From (2.26) it follows that all associated irreps can be written as

$$\{\mu;\nu\} = \{a^n\}\times\{0;\lambda\}. \tag{2.27}$$

This decomposition is important when considering the reduction of U_n irreps with respect to the subgroup SU_n. Although we will be concerned with SU_n only briefly, we mention here that all associated irreps are equivalent under reduction to SU_n and can be labeled by $\{0;\lambda\} = \{\lambda\}$. This is a consequence of the result that all one-dimensional irreps subduce to the scalar irrep of SU_n.

In applying (2.27) to U_n Kronecker products we can rewrite

$$\{\mu_1;\nu_1\}\times\{\mu_2;\nu_2\}\supset m\{\mu_3;\nu_3\} \tag{2.28}$$

as

$$\begin{aligned}&(\{a_1^n\}\times\{\lambda_1\})\times(\{a_2^n\}\times\{\lambda_2\})\\ &\quad\supset(\{a_1^n\}\times a_2^n\})\times(\{\lambda_1\}\times\{\lambda_2\})\\ &\quad\supset m\{a_3^n\}\times\{\lambda_3\},\end{aligned} \tag{2.29}$$

where $a_3 = a_1 + a_2$ and λ_3 is a regular partition into at most n parts. In triad form this is written

$$\begin{aligned}&(\{\mu_1;\nu_1\}\{\mu_2;\nu_2\}\{\mu_3;\nu_3\}r)\\ &\quad= (\{a_1^n\}\{a_2^n\}\{a_3^n\})(\{\lambda_1\}\{\lambda_2\}\{\lambda_3\}r).\end{aligned} \tag{2.30}$$

Similarly, since under the reduction $\mathrm{U}_{mn}\supset\mathrm{U}_m\times\mathrm{U}_n$

$$\{a^{mn}\}\supset\{(an)^m\}\times\{(am)^n\}, \tag{2.31}$$

a ket branching (2.18) can be put into the form

$$\begin{aligned}&\{a^{mn}\}\times\{\lambda\}\supset m'\{b^m\}\times\{\mu\}\times\{c^n\}\times\{\nu\}\\ &\qquad\supset m'(\{b^m\}\times\{c^n\})\times(\{\mu\}\times\{\nu\}),\end{aligned} \tag{2.32}$$

where $amn = bm = cn$ and λ,μ, and ν are regular partitions of the same integer and λ has at most $mn-1$ parts.

III. A GUIDE TO THE TABLES

Table I lists the irreps of U_n up to power 3, giving the complex conjugation properties and algebraic formulas for the dimension of the irreps. In Table II we have given the list of $3j$ phases associated with each triad. The $3j$ phase, denoted $\{\gamma_1\gamma_2\gamma_3 r\}$, gives the symmetry on reordering coupled products. We have used the fact that the irreps of U_n listed in Table I are simple phase irreps; the first nonsimple phase irrep is the power 6 irrep $\{21;21\}$ of U_n $(n\geqslant 4)$. The permutation matrix appropriate to reordering can be chosen diagonal[10] and the diagonal elements are the $3j$ phases. If one of the irreps in the triad is the trivial (identity) irrep, the $3j$ phase reduces to the $1j$ phase

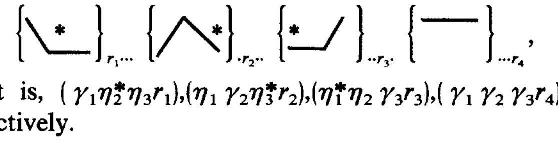
$$\{\gamma^*\gamma 00\} = \{\gamma\}.$$

The quasiambivalent choice,[10] $\{\gamma_1\}\{\gamma_2\}\{\gamma_3\} = +1$, where $(\gamma_1\gamma_2\gamma_3 r)$ forms a triad, can be used for U_n and implies that the A matrix appropriate to the complex conjugation symmetry can be chosen unity.

The $6j$ symbol is related to recouplings between a set of six irreps by mean of four triad couplings. The triads occur in the $6j$ symbol

$$\begin{Bmatrix}\gamma_1 & \gamma_2 & \gamma_3\\ \eta_1 & \eta_2 & \eta_3\end{Bmatrix}_{r_1r_2r_3r_4}, \tag{3.1}$$

in the order



(3.2)

that is, $(\gamma_1\eta_2^*\eta_3 r_1)$,$(\eta_1\gamma_2\eta_3^* r_2)$,$(\eta_1^*\eta_2\gamma_3 r_3)$,$(\gamma_1\gamma_2\gamma_3 r_4)$, respectively.

Symmetries are used to reduce the size of the tables. The full symmetries are given elsewhere[9–11] but to find a $6j$ in the table one needs the following. The $6j$ symbols are invariant under even permutations of the columns; there is the complex conjugation symmetry,

$$\begin{Bmatrix}\gamma_1 & \gamma_2 & \gamma_3\\ \eta_1 & \eta_2 & \eta_3\end{Bmatrix}_{r_1r_2r_3r_4} = \begin{Bmatrix}\gamma_1^* & \gamma_2^* & \gamma_3^*\\ \eta_1^* & \eta_2^* & \eta_3^*\end{Bmatrix}^*_{r_1r_2r_3r_4}, \tag{3.3}$$

the row flip symmetries, the (23) flip being

$$= \begin{Bmatrix}\gamma_1^* & \eta_2 & \eta_3^*\\ \eta_1^* & \gamma_2 & \gamma_3^*\end{Bmatrix}_{r_4r_3r_2r_1}, \tag{3.4}$$

and the column interchange symmetries, the (12) operation being

$$\begin{aligned}&= \begin{Bmatrix}\gamma_2 & \gamma_1 & \gamma_3\\ \eta_2^* & \eta_1^* & \eta_3^*\end{Bmatrix}_{r_2r_1r_3r_4}\\ &\quad\times\{\eta_1\}\{\eta_2\}\{\eta_3\}\{\gamma_1\eta_2^*\eta_3 r_1\}\{\eta_1\gamma_2\eta_3^* r_2\}\\ &\quad\times\{\eta_1^*\eta_2\gamma_3 r_3\}\{\gamma_1\gamma_2\gamma_3 r_4\}.\end{aligned} \tag{3.5}$$

The phase in (3.5) is the same for all interchanges but in addition we chose this phase $+1$ for all multiplicity-free $6j$ symbols. This choice fixes the value of certain $3j$ phases and these are given in Table II. We note that the choice is also invariant under the transpose conjugate symmetry. Some $3j$ phases, in particular those of the form $\{\gamma\gamma\gamma' r\}$, are fixed by the character theory. These values are also given in Table II.

The $6j$ symbols of U_n are tabulated in Table III. The bold typeface headings denote the top line of the $6j$ symbols and each subsequent entry denotes a lower line (three irreps and four multiplicity labels), the interchange sign, and the value.

The irrep decompositions for $\mathrm{U}_{mn}\supset\mathrm{U}_m\times\mathrm{U}_n$ are given in Table IV. No branching multiplicity occurs for the cases

TABLE I. Irreps of U_n.

Composite label	$0;0$	$0;1$	$1;0$	$1;1$	$0;1^2$	$1^2;0$	$0;2$	$2;0$
Complex conjugate	$0;0$	$1;0$	$0;1$	$1;1$	$1^2;0$	$0;1^2$	$2;0$	$0;2$
Dimension formula	1	n	n	$(n+1)(n-1)$	$n(n-1)/2$	$n(n-1)/2$	$(n+1)n/2$	$(n+1)n/2$

Composite label	$1;1^2$	$1^2;1$	$1;2$	$2;1$
Complex conjugate	$1^2;1$	$1;1^2$	$2;1$	$1;2$
Dimension formula	$(n+1)n(n-2)/2$	$(n+1)n(n-2)/2$	$(n+2)n(n-1)/2$	$(n+2)n(n-1)/2$

Composite label	$0;1^3$	$1^3;0$	$0;21$	$21;0$	$0;3$	$3;0$
Complex conjugate	$1^3;0$	$0;1^3$	$21;0$	$0;21$	$3;0$	$0;3$
Dimension formula	$n(n-1)(n-2)/6$	$n(n-1)(n-2)/6$	$(n+1)n(n-1)/3$	$(n+1)n(n-1)/3$	$(n+2)(n+1)n/6$	$(n+2)(n+1)n/6$

TABLE II. U_n $3j$ phases and values.

$\{0;0\ 0;0\ 0;0\}$	$=\{0;0\}=+1$ [a]
$\{1;0\ 0;1\ 0;0\}$	$=\{0;1\}$ [b]
$\{1;1\ 1;0\ 0;1\}$	
$\{1;1\ 1;1\ 0;0\}$	$=\{1;1\}=+1$ [a]
$\{1;1\ 1;1\ 1;1\ r\}$	$=H_r\{1;1\ 1;0\ 0;1\}\{0;1\}$ [a,e]
$\{0;1^2\ 1;0\ 1;0\}$	$=-1$ [a]
$\{1^2;0\ 0;1^2\ 0;0\}$	$=\{0;1^2\}=+1$ [c]
$\{1^2;0\ 0;1^2\ 1;1\}$	$=\{1;1\ 1;0\ 0;1\}\{0;1\}$ [d]
$\{0;2\ 1;0\ 1;0\}$	$=+1$ [a]
$\{0;2\ 1^2;0\ 1;1\}$	$=\{0;2\ 0;1\ 0;1\}\{0;1^2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\{0;1\}$ [d]
$\{2;0\ 0;2\ 0;0\}$	$=\{0;2\}=+1$ [c]
$\{2;0\ 0;2\ 1;1\}$	$=\{1;1\ 1;0\ 0;1\}\{0;1\}$ [d]
$\{1;1^2\ 1;1\ 1;0\}$	
$\{1;1^2\ 1^2;0\ 0;1\}$	$=\{1;1^2\ 1;1\ 1;0\}\{0;1^2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\{0;1\}$ [d]
$\{1;1^2\ 1;1^2\ 1^2;0\}$	$=\{0;1^2\ 1;0\ 1;0\}\{1;1\}=-1$ [a,d]
$\{1;1^2\ 1;1^2\ 2;0\}$	$=\{0;2\ 1;0\ 1;0\}\{1;1\}=+1$ [a,d]
$\{1^2;1\ 1;1^2\ 0;0\}$	$=\{1;1^2\}=\{0;1\}$ [c]
$\{1^2;1\ 1;1^2\ 1;1\ r\}$	$=H'_r\{1;1\ 1;0\ 0;1\}\{0;1^2\}$ [d,e]
$\{1;2\ 1;1\ 1;0\}$	
$\{1;2\ 2;0\ 0;1\}$	$=\{1;2\ 1;1\ 1;0\}\{0;2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\{1;1\}$ [d]
$\{1;2\ 1;1^2\ 1^2;0\}$	$=\{1;2\ 1;1\ 1;0\}\{1;1^2\ 1;1\ 1;0\}\{0;1^2\ 1;0\ 1;0\}\{1;1\}$ [d]
$\{1;2\ 1;1^2\ 2;0\}$	$=\{1;2\ 1;1\ 1;0\}\{1;1^2\ 1;1\ 1;0\}\{0;2\ 1;0\ 1;0\}\{1;1\}$ [d]
$\{1;2\ 1^2;1\ 1;1\}$	$=\{1;2\ 1;1\ 1;0\}\{1;1^2\ 1;1\ 1;0\}\{1;1\ 1;0\ 0;1\}\{1;1\}$ [d]
$\{1;2\ 1;2\ 1^2;0\}$	$=\{0;1^2\ 1;0\ 1;0\}\{1;1\}=-1$ [a,d]
$\{1;2\ 1;2\ 2;0\}$	$=\{0;2\ 1;0\ 1;0\}\{1;1\}=+1$ [a,d]
$\{2;1\ 1;2\ 0;0\}$	$=\{1;2\}=\{0;1\}$ [c]
$\{2;1\ 1;2\ 1;1\ r\}$	$=H'_r\{1;1\ 1;0\ 0;1\}\{0;2\}$ [d,e]
$\{0;1^3\ 1^2;0\ 1;0\}$	
$\{0;1^3\ 1^2;1\ 1^2;0\}$	$=\{0;1^3\ 1^2;0\ 1;0\}\{1;1^2\ 1^2;0\ 0;1\}\{0;1^2\ 1;0\ 1;0\}\{0;1^2\}$ [d]
$\{0;1^3\ 1^2;1\ 2;0\}$	$=\{0;1^3\ 1^2;0\ 1;0\}\{1;1^2\ 1^2;0\ 0;1\}\{0;2\ 1;0\ 1;0\}\{0;1^2\}$ [d]
$\{1^3;0\ 0;1^3\ 0;0\}$	$=\{0;1^3\}=\{0;1\}$ [c]
$\{1^3;0\ 0;1^3\ 1;1\}$	$=\{1;1\ 1;0\ 0;1\}\{0;1^2\}$ [d]
$\{0;21\ 1^2;0\ 1;0\}$	
$\{0;21\ 2;0\ 1;0\}$	$=\{0;21\ 1^2;0\ 1;0\}\{0;2\ 1;0\ 1;0\}\{0;1^2\ 1;0\ 1;0\}\{0;1^2\}$ [d]
$\{0;21\ 1^2;1\ 1^2;0\}$	$=\{0;21\ 1^2;0\ 1;0\}\{1;1^2\ 1^2;0\ 0;1\}\{0;1^2\ 1;0\ 1;0\}\{0;1^2\}$ [d]
$\{0;21\ 1^2;1\ 2;0\}$	$=\{0;21\ 1^2;0\ 1;0\}\{1;1^2\ 1^2;0\ 0;1\}\{0;2\ 1;0\ 1;0\}\{0;1^2\}$ [d]
$\{0;21\ 2;1\ 1^2;0\}$	$=\{0;21\ 2;0\ 1;0\}\{1;2\ 2;0\ 0;1\}\{0;1^2\ 1;0\ 1;0\}\{0;2\}$ [d]
$\{0;21\ 2;1\ 2;0\}$	$=\{0;21\ 2;0\ 1;0\}\{1;2\ 2;0\ 0;1\}\{0;2\ 1;0\ 1;0\}\{0;2\}$ [d]
$\{0;21\ 1^3;0\ 1;1\}$	$=\{0;21\ 1^2;0\ 1;0\}\{0;1^3\ 1^2;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\{0;1^2\}$ [d]
$\{21;0\ 0;21\ 0;0\}$	$=\{0;21\}=\{0;1\}$ [c]
$\{21;0\ 0;21\ 1;1\ r\}$	$=H'_r\{1;1\ 1;0\ 0;1\}\{0;1^2\}$ [d,e]
$\{0;3\ 2;0\ 1;0\}$	
$\{0;3\ 2;1\ 1^2;0\}$	$=\{0;3\ 2;0\ 1;0\}\{1;2\ 2;0\ 0;1\}\{0;1^2\ 1;0\ 1;0\}\{0;2\}$ [d]
$\{0;3\ 2;1\ 2;0\}$	$=\{0;3\ 2;0\ 1;0\}\{1;2\ 2;0\ 0;1\}\{0;2\ 1;0\ 1;0\}\{0;2\}$ [d]
$\{0;3\ 21;0\ 1;1\}$	$=\{0;3\ 2;0\ 1;0\}\{0;21\ 2;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\{0;2\}$ [d]
$\{3;0\ 0;3\ 0;0\}$	$=\{0;3\}=\{0;1\}$ [c]
$\{3;0\ 0;3\ 1;1\}$	$=\{1;1\ 1;0\ 0;1\}\{0;2\}$ [d]

[a] Fixed by character theory.

[b] $\{0;1\}=\begin{cases}+1, & \text{for } SU_{2k+1},\\ +1, & \text{for } SU_{2k},\ k \text{ even},\\ -1, & \text{for } SU_{2k},\ k \text{ odd}.\end{cases}$

[c] Fixed by the condition $\{\gamma_1\}\{\gamma_2\}\{\gamma_3\}=+1$ for the triad $(\gamma_1\,\gamma_2\,\gamma_3 r)$ imposed to give a unit choice for all A matrices. This choice leads to imaginary values for some $6j$ symbols.

[d] Fixed by the requirement that as many $6j$ symbols are invariant under transposition of columns. This may give a reality criterion for some $6j$ symbols and simplifies the transpose conjugate symmetry of others.

[e] $H=\mathrm{diag}(+1,-1)$, $H'=\mathrm{diag}(+1,+1)$.

TABLE III. U_n 6j formulas. The boldfaced headings denote the top line of the 6j symbol. Each subsequent entry denotes a lower line (three irrep labels and four multiplicity labels), followed by the algebraic formula. We have also used the following notation:

$$\chi_r = \begin{cases} +1, & \text{if } \{1;1\ 1;1\ 1;1\ r\}\{1;1\ 1;0\ 0;1\}\{0;1\} = 1, \\ +i, & \text{if } \{1;1\ 1;1\ 1;1\ r\}\{1;1\ 1;0\ 0;1\}\{0;1\} = +1, \end{cases} \qquad x = \delta_{r_1}\delta_{r_2}\delta_{r_3},$$

$$\epsilon_{r_1r_2r_3r_4} = \tfrac{1}{2}[\chi_{r_1}\chi_{r_2}\chi_{r_3}\chi_{r_4} + \chi^*_{r_1}\chi^*_{r_2}\chi^*_{r_3}\chi^*_{r_4}], \qquad y = \delta_{r_1}\delta_{r_2} + \delta_{r_2}\delta_{r_3} + \delta_{r_3}\delta_{r_1},$$

$$\delta_r = \tfrac{1}{2}[1 + \{1;1\ 1;1\ 1;1\ r\}] = \begin{cases} 1, & \text{if } \{1;1\ 1;1\ 1;1\ r\} = +1, \\ 0, & \text{if } \{1;1\ 1;1\ 1;1\ r\} = -1, \end{cases} \qquad z = \delta_{r_1} + \delta_{r_2} + \delta_{r_3}.$$

1;1	**1;0**	**0;1**		
1;1	1;0	0;1	0000	$\{0;1\}\dfrac{-1}{(n+1)n(n-1)}$
1;1	**1;1**	**1;1**		
0;1	0;1	0;1	000r	$\chi_r \dfrac{+1}{(n+1)(n-1)}\sqrt{\dfrac{(n+2\delta_r)(n-2\delta_r)}{2n}}$
1;1	1;1	1;1	$r_1r_2r_3r_4$	$\epsilon_{r_1r_2r_3r_4}(n^2-4(3x-y+z))\dfrac{(n+2x)(n-2x)}{2(n+1)(n-1)}\prod_{i=1}^{4}\sqrt{\dfrac{1}{(n+2\delta_{r_i})(n-2\delta_{r_i})}}$
$0;1^2$	**1;0**	**1;0**		
1;1	0;1	1;0	0000	$\dfrac{+1}{n(n-1)}$
$1^2;0$	**$0;1^2$**	**1;1**		
1;0	1;0	0;1	0000	$\dfrac{+\sqrt{n-2}}{n(n-1)}$
1;1	1;1	0;1	00r0	$\chi_r\{0;1^2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\dfrac{(n-4\delta_r)}{(n+1)(n-1)}\sqrt{\dfrac{(n+2\delta_r)}{2n(n-2)(n-2\delta_r)}}$
$1^2;0$	$1^2;0$	1;1	0000	$\dfrac{(n-1+\sqrt{5})(n-1-\sqrt{5})}{(n+1)n(n-1)(n-2)}$
0;2	**1;0**	**1;0**		
1;1	0;1	1;0	0000	$\dfrac{+1}{(n+1)n}$
0;2	**$1^2;0$**	**1;1**		
0;1	0;1	1;0	0000	$\dfrac{+1}{(n+1)n(n-1)}$
1;1	1;1	0;1	00r0	$\chi_r\{0;2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\dfrac{+1}{(n+1)(n-1)}\sqrt{\dfrac{(n-2)(n+2\delta_r)}{2n(n-2\delta_r)}}$
$0;1^2$	$0;1^2$	1;1	0000	$\dfrac{-1}{(n+1)(n-1)}$
0;2	$0;1^2$	1;1	0000	$\dfrac{+1}{(n+1)(n-1)}$
2;0	**0;2**	**1;1**		
1;0	1;0	0;1	0000	$\dfrac{+\sqrt{n+2}}{(n+1)n}$
1;1	1;1	$0;1^2$	00r0	$\chi_r\{0;1^2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\dfrac{+1}{(n+1)(n-1)}\sqrt{\dfrac{(n+2)(n-2\delta_r)}{2(n+2\delta_r)n}}$
1;1	1;1	0;2	00r0	$\chi_r\{0;2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\dfrac{n+4}{(n+1)(n-1)}\sqrt{\dfrac{n-2\delta_r}{2(n+2)(n+2\delta_r)n}}$
$1^2;0$	$1^2;0$	1;1	0000	$\dfrac{+\sqrt{(n+2)(n-2)}}{(n+1)n(n-1)}$
2;0	$1^2;0$	1;1	0000	$\dfrac{-1}{(n+1)(n-1)}$
2;0	2;0	1;1	0000	$\dfrac{(n+1+\sqrt{5})(n+1-\sqrt{5})}{(n+1)n(n-1)(n-2)}$
$1;1^2$	**1;1**	**1;0**		
1;0	1;1	1;0	0000	$\{0;1^2\ 1;0\ 1;0\}\{0;1\}\dfrac{+1}{(n+1)(n-1)}$
1;1	0;1	1;1	0r00	$\chi_r\{1;1^2\ 1;1\ 1;0\}\{0;1^2\ 1;0\ 1;0\}\{0;1\}\dfrac{+1}{(n+1)(n-1)}\sqrt{\dfrac{n+2}{2n(n-2\delta_r)}}$
$1^2;1$	1;1	1;0	0000	$\{0;1\}\dfrac{+1}{(n+1)(n-1)(n-2)}$

$1;1^2$	**$1^2;0$**	**$0;1$**		
$0;1$	$1;1$	$1;0$	0000	$\frac{+1}{n-1}\sqrt{\frac{2}{(n+1)n}}$
$1;1$	$1;0$	$1^2;0$	0000	$\{1;1^2\ 1^2;0\ 0;1\}\{0;1^2\ 1;0\ 1;0\}\frac{-2}{(n+1)n(n-1)(n-2)}$
$0;1^2$	$0;1$	$1;1$	0000	$\{1;1^2\ 1^2;0\ 0;1\}\ \frac{+1}{n-1}\sqrt{\frac{2}{(n+1)n(n-2)}}$
$0;2$	$0;1$	$1;1$	0000	$\{1;1^2\ 1^2;0\ 0;1\}\ \frac{-1}{n+1}\sqrt{\frac{2}{n(n-1)}}$
$1;1^2$	$1;1$	$1;0$	0000	$\frac{-2}{(n+1)n(n-1)(n-2)}$
$1;1^2$	**$1;1^2$**	**$1^2;0$**		
$1;0$	$0;1$	$1;1$	0000	$\frac{+1}{n-1}\sqrt{\frac{n-3}{(n+1)n(n-2)}}$
$1^2;0$	$1;1$	$1;0$	0000	$\{0;1^2\ 1;0\ 1;0\}\ \frac{\sqrt{2(n-3)}}{(n+1)(n-1)(n-2)}$
$1;1^2$	**$1;1^2$**	**$2;0$**		
$1;0$	$0;1$	$1;1$	0000	$\frac{+1}{n+1}\sqrt{\frac{1}{n(n-2)}}$
$1^2;0$	$1;1$	$1;0$	0000	$\{0;2\ 1;0\ 1;0\}\ \frac{+1}{n+1}\sqrt{\frac{2}{n(n-1)(n-2)}}$
$1^2;1$	**$1;1^2$**	**$1;1$**		
$0;1$	$0;1$	$0;1^2$	0000	$\{1;1\ 1;0\ 0;1\}\{0;1\}\ \frac{2(2n-1)}{(n+1)n(n-1)}\sqrt{\frac{1}{(n-2)(n^3-2n^2-2n+2)}}$
$0;1$	$0;1$	$0;1^2$	0001	$\frac{i}{(n+1)}\sqrt{\frac{2(n+2)(n-3)}{(n-2)(n^3-2n^2-2n+2)}}$
$1;0$	$1;0$	$1;1$	0000	$\frac{+1}{(n+1)n(n-1)}\sqrt{\frac{n^3-2n^2-2n+2}{n-2}}$
$1;0$	$1;0$	$1;1$	0001	0
$1;1$	$1;1$	$0;1$	$00r0$	$\chi_r\{1;1\ 1;1\ 1;1\ r\}\ \frac{(n^2-2(1+\delta_r))(n+2)}{(n+1)(n-1)}\sqrt{\frac{(n+2\delta_r)}{2n(n-2\delta_r)(n-2)(n^3-2n^2-2n+2)}}$
$1;1$	$1;1$	$0;1$	$00r1$	$\chi_r\{1;1\ 1;1\ 1;1\ r\}\ \frac{(n+2-2\delta_r)}{(n+1)}\sqrt{\frac{(n+2\delta_r)(n-3)}{(n+2)_r(n-2\delta_r)(n-2)(n^3-2n^2-2n+2)}}$
$1^2;0$	$1^2;0$	$1;0$	0000	$\{1;1\ 1;0\ 0;1\}\{0;1\}\ \frac{2\sqrt{n^3-2n^2-2n+2}}{(n+1)n(n-1)(n-2)}$
$1^2;0$	$1^2;0$	$1;0$	0001	0
$1;1^2$	$0;1$	$0;1^2$	0000	$\{0;1\}\ \frac{n}{(n+1)(n-1)(n-2)}\sqrt{\frac{2(n-3)}{n^3-2n^2-2n+2}}$
$1;1^2$	$0;1$	$0;1^2$	0001	$\frac{-2i}{(n+1)n(n-2)}\sqrt{\frac{n+2}{n^3-2n^2-2n+2}}$
$1^2;1$	$1;0$	$1;1$	0000	$\{1^2;1\ 1;1^2\ 1;1\ 0\}\frac{(n^2-6n+4)}{(n+1)(n-1)(n-2)(n^3-2n^2-2n+2)}$
$1^2;1$	$1;0$	$1;1$	0001	$\{1^2;1\ 1;1^2\ 1;1\ 0\}\ \frac{(n-1)\sqrt{2(n+2)(n-3)}}{(n+1)(n-2)(n^3-2n^2-2n+2)}$
$1^2;1$	$1;0$	$1;1$	0101	$\{1^2;1\ 1;1^2\ 1;1\ 0\}\ \frac{2(2n-1)}{(n+1)n(n-2)(n^3-2n^2-2n+2)}$
$1;2$	**$1;1$**	**$1;0$**		
$1;0$	$1;1$	$1;0$	0000	$\{0;2\ 1;0\ 1;0\}\{0;1\}\frac{+1}{(n+1)(n-1)}$
$1;1$	$0;1$	$1;1$	$0r00$	$\chi_r\{1;2\ 1;1\ 1;0\}\{0;2\ 1;0\ 1;0\}\{0;1\}\ \frac{+1}{(n+1)(n-1)}\sqrt{\frac{n-2\delta_r}{2(n+2\delta_r)n}}$
$1^2;1$	$1;1$	$1;0$	0000	$\{0;1\}\frac{+1}{(n+2)n(n-1)}$
$2;1$	$1;1$	$1;0$	0000	$\{0;1\}\frac{+1}{(n+2)(n+1)(n-1)}$

TABLE III. (*Continued.*)

1;2	**2;0**	**0;1**		
0;1	1;1	1;0	0000	$\frac{+1}{n+1}\sqrt{\frac{2}{n(n-1)}}$
1;1	1;0	2;0	0000	$\{1;2\ 2;0\ 0;1\}\{0;2\ 1;0\ 1;0\}\ \frac{-2}{(n+1)n(n-1)}\sqrt{\frac{1}{n+2}}$
$0;1^2$	0;1	1;1	0000	$\{1;2\ 2;0\ 0;1\}\ \frac{-1}{n(n-1)}\sqrt{\frac{2}{n+1}}$
0;2	0;1	1;1	0000	$\{1;2\ 2;0\ 0;1\}\ \frac{+1}{n+1}\sqrt{\frac{2}{(n+2)n(n-1)}}$
1;2	1;1	1;0	0000	$\frac{-2}{(n+2)(n+1)n(n-1)}$
1;2	**$1;1^2$**	**$1^2;0$**		
1;0	0;1	1;1	0000	$+\frac{1}{n}\sqrt{\frac{1}{(n+1)(n-1)}}$
1;1	0;2	0;1	0000	$\{1;2\ 1;1\ 1;0\}\{1;1^2\ 1;1\ 1;0\}\ \frac{+1}{(n+1)n}\sqrt{\frac{2}{n-1}}$
$1^2;0$	1;1	1;0	0000	$\{1;2\ 1;1\ 1;0\}\{1;1^2\ 1^2;0\ 0;1\}\{1;1\ 1;0\ 0;1\}\ \frac{-1}{(n+1)}\sqrt{\frac{2}{n(n-1)(n-2)}}$
$1^2;1$	0;1	1;1	0000	$\{0;1\}\ \frac{-1}{(n+1)}\sqrt{\frac{2(n-2)}{(n-1)n(n^3-2n^2-2n+2)}}$
$1^2;1$	0;1	1;1	0100	$\{0;2\ 1;0\ 1;0\}\ \frac{-2i}{(n+1)}\sqrt{\frac{(n-3)}{(n+2)n(n-1)(n-2)(n^3-2n^2-2n+2)}}$
1;2	**$1;1^2$**	**2;0**		
1;0	0;1	1;1	0000	$+\frac{1}{n}\sqrt{\frac{1}{(n+1)(n-1)}}$
1;1	0;2	0;1	0000	$\{1;2\ 1;1\ 1;0\}\{1;1^2\ 1;1\ 1;0\}\ \frac{-1}{n-1}\sqrt{\frac{2}{(n+2)(n+1)n}}$
$1^2;0$	1;1	1;0	0000	$\{1;2\ 1;1\ 1;0\}\{1;1^2\ 1^2;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\ \frac{+1}{n(n-1)}\sqrt{\frac{2}{n+1}}$
1;2	**$1^2;1$**	**1;1**		
0;1	0;1	1;1	0000	$\frac{+1}{(n+1)(n-1)}$
0;1	$1;1^2$	1;1	0000	$\{0;2\ 1;0\ 1;0\}\{0;1\}\ \frac{-1}{(n+1)(n-2)}\sqrt{\frac{2}{n(n-1)}}$
1;0	$1^2;1$	$1^2;0$	0000	$\{0;1\}\ \frac{-1}{(n+1)(n-2)}\sqrt{\frac{2}{n(n-1)}}$
1;1	1;1	1;0	$00r0$	$\chi_r\{1;2\ 2;0\ 0;1\}\{1;1^2\ 1^2;0\ 0;1\}\{0;1\}\ \frac{(n-2(1-\delta_r))}{(n+1)(n-1)(n-2)}\sqrt{\frac{(n-2\delta_r)}{2n)(n+2\delta_r)}}$
$0;1^2$	0;2	0;1	0000	$\{1;2\ 1;1\ 1;0\}\{1;1^2\ 1;1\ 1;0\}\ \frac{-2}{(n+1)(n-1)}\sqrt{\frac{1}{n}}$
$1;1^2$	0;1	1;1	0000	$\{0;2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\ \frac{1}{(n+1)(n-1)}\sqrt{\frac{1}{(n-2)(n^3-2n^2-2n+2)}}$
$1;1^2$	0;1	1;1	0100	$\frac{-i}{n}\sqrt{\frac{2(n-3)}{(n+2)(n-2)(n^3+2n^2-2n-2)}}$
$1^2;1$	1;0	2;0	0000	$\{0;1\}\ \frac{+1}{(n+1)n}\sqrt{\frac{2}{(n-1)(n-2)}}$
1;2	0;1	1;1	0000	$\{0;1^2\ 1;0\ 1;0\}\{0;1\}\ \frac{-1}{(n+2)(n-1)}\sqrt{\frac{2}{(n+1)n}}$
2;1	1;0	2;0	0000	$\{0;1\}\ \frac{-1}{(n+2)(n-1)}\sqrt{\frac{2}{(n+1)n}}$
1;2	**1;2**	**$1^2;0$**		
1;0	0;1	1;1	0000	$\frac{+1}{(n-1)}\sqrt{\frac{1}{(n+2)n}}$
2;0	1;1	1;0	0000	$\{0;1^2\ 1;0\ 1;0\}\ \frac{+1}{n-1}\sqrt{\frac{2}{(n+2)(n+1)n}}$

TABLE III. (*Continued.*)

$1^2;0$	$0;1$	$1;1$	0000	$\{0;1\}\,\frac{+1}{n(n-1)}\sqrt{\frac{2}{(n+2)(n+1)}}$
1;2	**1;2**	**2;0**		
$1;0$	$0;1$	$1;1$	0000	$\frac{+1}{n+1}\sqrt{\frac{n+3}{(n+2)n(n-1)}}$
$2;0$	$1;1$	$1;0$	0000	$\{0;2\ 1;0\ 1;0\}\,\frac{\sqrt{2(n+3)}}{(n+2)(n+1)(n-1)}$
2;1	**1;2**	**1;1**		
$0;1$	$0;1$	$0;2$	0000	$\{1;1\ 1;0\ 0;1\}\{0;1\}\,\frac{2(2n+1)}{(n+1)n(n-1)}\sqrt{\frac{1}{(n+2)(n^3+2n^2-2n-2)}}$
$0;1$	$0;1$	$0;2$	0001	$\frac{i}{(n-1)}\sqrt{\frac{2(n+3)(n-2)}{(n+2)(n^3+2n^2-2n-2)}}$
$1;0$	$1;0$	$1;1$	0000	$\frac{1}{(n+1)n(n-1)}\sqrt{\frac{n^3+2n^2-2n-2}{n+2}}$
$1;0$	$1;0$	$1;1$	0001	0
$1;1$	$1;1$	$0;1$	$00r0$	$\chi_r\{1;1\ 1;1\ 1;1\ r\}\,\frac{(n^2+2(1+\delta_r))(n-2)}{(n+1)(n-1)}\sqrt{\frac{(n-2\delta_r)}{2(n+2)(n+2\delta_r)n(n^3+2n^2-2n-2)}}$
$1;1$	$1;1$	$0;1$	$00r1$	$\chi_r\{1;1\ 1;1\ 1;1\ r\}\,\frac{(n-2+2\delta_r)}{(n-1)}\sqrt{\frac{(n+3)(n-2\delta_r)}{(n+2)(n+2\delta_r)n(n-2)(n^3+2n^2-2n-2)}}$
$2;0$	$2;0$	$1;0$	0000	$\{1;1\ 1;0\ 0;1\}\{0;1\}\,\frac{2\sqrt{n^3+2n^2-2n-2}}{(n+2)(n+1)n(n-1)}$
$2;0$	$2;0$	$1;0$	0001	0
$1;1^2$	$0;1$	$0;2$	0000	$\{0;1\}\,\frac{-1}{(n-1)}\sqrt{\frac{2(n+2)}{(n+1)n(n^3+2n^2-2n-2)}}$
$1;1^2$	$0;1$	$0;2$	0001	$\{0;1^2\ 1;0\ 1;0\}\,\frac{2i}{(n-1)}\sqrt{\frac{1}{(n+2)(n+1)n(n-2)(n^3+2n^2-2n-2)}}$
$1^2;1$	$1;0$	$1;1$	0000	$\{0;2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\,\frac{1}{(n+1)(n-1)}\sqrt{\frac{1}{(n+2)(n^3+2n^2-2n-2)}}$
$1^2;1$	$1;0$	$1;1$	0001	$\frac{-i}{n}\sqrt{\frac{2(n+3)}{(n+2)(n-2)(n^3+2n^2-2n-2)}}$
$1;2$	$0;1$	$0;2$	0000	$\{0;1\}\,\frac{n}{(n+2)(n+1)(n-1)}\sqrt{\frac{2(n+3)}{n^3+2n^2-2n-2}}$
$1;2$	$0;1$	$0;2$	0001	$\frac{-2i}{(n+2)n(n-1)}\sqrt{\frac{n-2}{n^3+2n^2-2n-2}}$
$2;1$	$1;0$	$1;1$	0000	$\{2;1\ 1;2\ 1;1\ 0\}\,\frac{(n^2+6n+4)}{(n+2)(n+1)(n-1)(n^3+2n^2-2n-2)}$
$2;1$	$1;0$	$1;1$	0001	$\{2;1\ 1;2\ 1;1\ 0\}\,\frac{i(n+1)\sqrt{2(n+3)(n-2)}}{(n+2)(n-1)(n^3+2n^2-2n-2)}$
$2;1$	$1;0$	$1;1$	0101	$\{2;1\ 1;2\ 1;10\}\{0;1\}\,\frac{-2(2n+1)}{(n+2)n(n-1)(n^3+2n^2-2n-2)}$
$\mathbf{0;1^3}$	$\mathbf{1^2;0}$	**1;0**		
$0;1$	$0;1^2$	$1;0$	0000	$\{0;1^2\ 1;0\ 1;0\}\frac{+2}{n(n-1)}$
$1;1$	$0;1$	$1^2;0$	0000	$\{0;1^3\ 1^2;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\{0;1\}\,\frac{+1}{n(n-1)}\sqrt{\frac{1}{n-2}}$
$1;1^2$	$0;1^2$	$1;0$	0000	$\{0;1\}\frac{+2}{n(n-1)(n-2)}$
$\mathbf{0;1^3}$	$\mathbf{1^2;1}$	$\mathbf{1^2;0}$		
$1;0$	$0;1$	$1^2;0$	0000	$\frac{+1}{n(n-1)}\sqrt{\frac{2(n-3)}{n-2}}$
$1;0$	$1;1^2$	$1^2;0$	0000	$\{0;1^2\ 1;0\ 1;0\}\{0;1\}\frac{-4}{(n+1)n(n-1)(n-2)}$
$1;1$	$0;1^2$	$1;0$	0000	$\{0;1^3\ 1^2;0\ 1;0\}\{1;1^2\ 1^2;0\ 0;1\}\{1;1\ 1;0\ 0;1\}\,\frac{+2}{(n-1)(n-2)}\sqrt{\frac{n-3}{(n+1)n}}$
$1^2;1$	$0;1$	$1^2;0$	0000	$\{0;1^2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\,\frac{+2}{(n-1)(n-2)}\sqrt{\frac{2}{(n+1)n}}$

$0;1^3$	**$1^2;1$**	**$2;0$**		
$1;0$	$0;1$	$1^2;0$	0000	$+\frac{1}{n}\sqrt{\frac{2}{(n+1)(n-2)}}$
$1;1$	$0;1^2$	$1;0$	0000	$\{0;1^3\ 1^2;0\ 1;0\}\{1;1^2\ 1^2;0\ 0;1\}\{1;1\ 1;0\ 0;1\}\ \frac{-2}{n}\sqrt{\frac{1}{(n+1)(n-1)(n-2)}}$
$2;1$	$0;1$	$1^2;0$	0000	$\{0;1^2\ 1;0\ 1;0\}\ \frac{+2}{n(n-1)}\sqrt{\frac{2}{(n+1)(n-2)}}$
$1^3;0$	**$0;1^3$**	**$1;1$**		
$1;0$	$1;0$	$0;1^2$	0000	$\frac{+1}{n(n-1)}\sqrt{\frac{2(n-3)}{n-2}}$
$1^2;0$	$1^2;0$	$0;1$	0000	$\{1^2;0\ 0;1^2\ 1;1\}\{0;1^2\ 1;0\ 1;0\}\ \frac{+2}{n(n-1)}\sqrt{\frac{2(n-3)}{(n-2)}}$
$1^2;1$	$1;0$	$0;1^2$	0000	$\{1;1\ 1;0\ 0;1\}\{0;1\}\ \frac{+2}{(n-1)(n-2)}\sqrt{\frac{2}{(n+1)n}}$
$0;21$	**$1^2;0$**	**$1;0$**		
$0;1$	$0;1^2$	$1;0$	0000	$\{0;1^2\ 1;0\ 1;0\}\{0;1\}\frac{-1}{n(n-1)}$
$1;1$	$0;1$	$1^2;0$	0000	$\{0;21\ 1^2;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\{0;1\}\ \frac{-\sqrt{n-2}}{(n+1)n(n-1)}$
$1;1^2$	$0;1^2$	$1;0$	0000	$\{0;1\}\frac{+2}{(n+1)n(n-1)}$
$0;21$	**$2;0$**	**$1;0$**		
$0;1$	$0;1^2$	$1;0$	0000	$\{0;1\}\ \frac{1}{n}\sqrt{\frac{3}{(n+1)(n-1)}}$
$0;1$	$0;2$	$1;0$	0000	$\{0;2\ 1;0\ 1;0\}\{0;1\}\frac{-1}{(n+1)n}$
$1;1$	$0;1$	$1^2;0$	0000	$\{0;21\ 1^2;0\ 1;0\}\{0;1^2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\{0;1\}\ \frac{1}{(n+1)(n-1)}\sqrt{\frac{1}{n}}$
$1;1$	$0;1$	$2;0$	0000	$\{0;21\ 2;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\{0;1\}\ \frac{-\sqrt{n+2}}{(n+1)n(n-1)}$
$1;2$	$0;2$	$1;0$	0000	$\{0;1\}\frac{+2}{(n+1)n(n-1)}$
$0;21$	**$1^2;1$**	**$1^2;0$**		
$1;0$	$0;1$	$1^2;0$	0000	$\frac{+1}{n-1}\sqrt{\frac{2}{(n+1)n}}$
$1;0$	$1;1^2$	$1^2;0$	0000	$\{0;1^2\ 1;0\ 1;0\}\{0;1\}\frac{+2}{(n+1)n(n-1)(n-2)}$
$1;1$	$0;1^2$	$1;0$	0000	$\{0;21\ 1^2;0\ 1;0\}\{1;1^2\ 1^2;0\ 0;1\}\{1;1\ 1;0\ 0;1\}\ \frac{-1}{(n+1)(n-1)}\sqrt{\frac{1}{n-2}}$
$1;1$	$0;2$	$1;0$	0000	$\{0;21\ 2;0\ 1;0\}\{1;1^2\ 1^2;0\ 0;1\}\{0;1^2\ 1;0\ 1;0\}\ \frac{1}{(n+1)(n-1)}\sqrt{\frac{1}{n}}$
$1^2;1$	$0;1$	$1^2;0$	0000	$\{0;1^2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\ \frac{-1}{(n+1)n(n-1)}\sqrt{\frac{2(n-3)}{n-2}}$
$1^2;1$	$0;1$	$2;0$	0000	$\{0;1^2\ 1;0\ 1;0\}\ \frac{1}{(n+1)n}\sqrt{\frac{6}{(n-1)(n-2)}}$
$0;21$	**$1^2;1$**	**$2;0$**		
$1;0$	$0;1$	$1^2;0$	0000	$\frac{+1}{(n+1)n}\sqrt{\frac{2(n+2)}{n-1}}$
$1;1$	$0;1^2$	$1;0$	0000	$\{0;21\ 1^2;0\ 1;0\}\{1;1^2\ 1^2;0\ 0;1\}\{1;1\ 1;0\ 0;1\}\ \frac{\sqrt{n+2}}{(n+1)n(n-1)}$
$1;1$	$0;2$	$1;0$	0000	$\{0;21\ 2;0\ 1;0\}\{1;1^2\ 1;1\ 1;0\}\{1;1\ 1;0\ 0;1\}\{0;1\}\ \frac{1}{(n+1)(n-1)}\sqrt{\frac{3}{n}}$
$2;1$	$0;1$	$1^2;0$	0000	$\{0;2\ 1;0\ 1;0\}\ \frac{+1}{(n+1)n}\sqrt{\frac{2}{(n+2)(n-1)}}$
$2;1$	$0;1$	$2;0$	0000	$\frac{1}{n(n-1)}\sqrt{\frac{6}{(n+2)(n+1)}}$

0;21	**2;1**	**$1^2;0$**		
1;0	0;1	2;0	0000	$\frac{+1}{n(n-1)}\sqrt{\frac{2(n-2)}{n+1}}$
1;0	$1;1^2$	2;0	0000	$\{0;1\}\ \frac{-2}{(n+1)n(n-1)}\sqrt{\frac{3}{(n+2)(n-2)}}$
1;1	$0;1^2$	1;0	0000	$\{0;21\ 1^2;0\ 1;0\}\{1;2\ 1;1\ 1;0\}\{1;1\ 1;0\ 0;1\}\ \frac{+1}{(n+1)(n-1)}\sqrt{\frac{3}{n}}$
1;1	0;2	1;0	0000	$\{0;21\ 2;0\ 1;0\}\{1;2\ 2;0\ 0;1\}\{1;1\ 1;0\ 0;1\}\ \frac{+\sqrt{n-2}}{(n+1)n(n-1)}$
$1^2;1$	0;1	$1^2;0$	0000	$\frac{+1}{(n+1)n}\sqrt{\frac{6}{(n-1)(n-2)}}$
$1^2;1$	0;1	2;0	0000	$\{0;1^2\ 1;0\ 1;0\}\ \frac{+1}{n(n-1)}\sqrt{\frac{2}{(n+2)(n-2)}}$
0;21	**2;1**	**2;0**		
1;0	0;1	2;0	0000	$\frac{+1}{n+1}\sqrt{\frac{2}{n(n-1)}}$
1;0	1;2	2;0	0000	$\{0;2\ 1;0\ 1;0\}\{0;1\}\frac{+2}{(n+2)(n+1)n(n-1)}$
1;1	$0;1^2$	1;0	0000	$\{0;21\ 1^2;0\ 1;0\}\{1;2\ 2;0\ 0;1\}\{0;2\ 1;0\ 1;0\}\ \frac{+1}{(n+1)(n-1)}\sqrt{\frac{3}{n}}$
1;1	0;2	1;0	0000	$\{0;21\ 2;0\ 1;0\}\{1;2\ 2;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\ \frac{-1}{(n+1)(n-1)}\sqrt{\frac{1}{n+2}}$
2;1	0;1	$1^2;0$	0000	$\{0;2\ 1;0\ 1;0\}\ \frac{+1}{n(n-1)}\sqrt{\frac{6}{(n+2)(n+1)}}$
2;1	0;1	2;0	0000	$\{0;2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\ \frac{-1}{(n+1)n(n-1)}\sqrt{\frac{2(n+3)}{n+2}}$
0;21	**$1^3;0$**	**1;1**		
0;1	0;1	$1^2;0$	0000	$\frac{+1}{n-1}\sqrt{\frac{2}{(n+1)n}}$
0;1	$1;1^2$	$1^2;0$	0000	$\{0;1\}\ \frac{-1}{(n-1)(n-2)}\sqrt{\frac{2}{(n+1)n}}$
0;1	1;2	$1^2;0$	0000	$\{0;2\ 1;0\ 1;0\}\{0;1\}\ \frac{-1}{n(n-1)}\sqrt{\frac{6}{(n+1)(n-2)}}$
$0;1^2$	$0;1^2$	1;0	0000	$\{0;21\ 1^2;0\ 1;0\}\{0;1^3\ 1^2;0\ 1;0\}\{0;2\ 1;0\ 1;0\}\{0;1\}\ \frac{+1}{n-1}\sqrt{\frac{2}{(n+1)n(n-2)}}$
$0;1^2$	0;2	1;0	0000	$\{0;21\ 1^2;0\ 1;0\}\{0;1^3\ 1^2;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\{0;1\}\ \frac{-1}{n(n-1)}\sqrt{\frac{6}{n+1}}$
$1;1^2$	0;1	$1^2;0$	0000	$\{1;1\ 1;0\ 0;1\}\{0;1\}\ \frac{+1}{(n+1)n(n-1)}\sqrt{\frac{2(n-3)}{(n-2)}}$
$1;1^2$	0;1	0;2	0000	$\{0;1^2\ 1;0\ 1;0\}\{0;1\}\ \frac{+1}{(n+1)n}\sqrt{\frac{6}{(n-1)(n-2)}}$
21;0	**0;21**	**1;1**		
1;0	1;0	$0;1^2$	0000	$\{0;21\ 1^2;0\ 1;0\}\{0;2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\ \frac{-1}{n(n-1)}$
1;0	1;0	$0;1^2$	0001	$\{0;21\ 1^2;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\ \frac{i\sqrt{(n+2)(n-2)}}{(n+1)n(n-1)}$
1;0	1;0	0;2	0000	$\{0;21\ 2;0\ 1;0\}\{0;1^2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\ \frac{-1}{(n+1)n}$
1;0	1;0	0;2	0001	$\{0;21\ 2;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\ \frac{i\sqrt{(n+2)(n-2)}}{(n+1)n(n-1)}$
$1^2;0$	$1^2;0$	0;1	0000	$\{0;21\ 1^2;0\ 1;0\}\{0;1\}\ \frac{-\sqrt{n-2}}{(n+1)n(n-1)}$
$1^2;0$	$1^2;0$	0;1	0001	$\{0;21\ 1^2;0\ 1;0\}\{0;1^2\ 1;0\ 1;0\}\{0;1\}\ \frac{2i\sqrt{n+2}}{(n+1)n(n-1)}$
2;0	$1^2;0$	0;1	0000	$\{0;21\ 2;0\ 1;0\}\{0;2\ 1;0\ 1;0\}\{0;1\}\ \frac{+1}{(n+1)(n-1)}\sqrt{\frac{3}{n}}$

TABLE III. (*Continued.*)

2;0	1^2;0	0;1	0001	0
2;0	2;0	0;1	0000	$\{0;21\ 2;0\ 1;0\}\{0;1\}\ \frac{-\sqrt{n+2}}{(n+1)n(n-1)}$
2;0	2;0	0;1	0001	$\{0;21\ 2;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\ \frac{2i\sqrt{n-2}}{(n+1)n(n-1)}$
1^2;1	1;0	$0;1^2$	0000	$\{0;21\ 1^2;0\ 1;0\}\{0;2\ 1;0\ 1;0\}\{0;1\}\ \frac{2}{(n+1)n(n-2)}$
1^2;1	1;0	$0;1^2$	0001	$\{0;21\ 1^2;0\ 1;0\}\{0;1\}\ \frac{-i}{(n+1)n(n-1)}\sqrt{\frac{n+2}{n-2}}$
1^2;1	1;0	0;2	0000	$\{0;21\ 2;0\ 1;0\}\{0;1\}\ \frac{+1}{(n+1)(n-1)}\sqrt{\frac{3}{(n+2)n}}$
1^2;1	1;0	0;2	0001	$\{0;21\ 2;0\ 1;0\}\{0;1\}\ \frac{+i}{(n+1)(n-1)}\sqrt{\frac{3}{n(n-2)}}$
2;1	1;0	$0;1^2$	0000	$\{0;21\ 1^2;0\ 1;0\}\{0;1\}\ \frac{+1}{(n+1)(n-1)}\sqrt{\frac{3}{n(n-2)}}$
2;1	1;0	$0;1^2$	0001	$\{0;21\ 1^2;0\ 1;0\}\{0;1\}\ \frac{+i}{(n+1)(n-1)}\sqrt{\frac{3}{(n+2)n}}$
2;1	1;0	0;2	0000	$\{0;21\ 2;0\ 1;0\}\{0;1^2\ 1;0\ 1;0\}\{0;1\}\ \frac{-2}{(n+1)n(n-2)}$
2;1	1;0	0;2	0001	$\{0;21\ 2;0\ 1;0\}\{0;1\}\ \frac{-i}{(n+1)n(n-1)}\sqrt{\frac{n-2}{n+2}}$
0;3	**2;0**	**1;0**		
0;1	0;2	1;0	0000	$\{0;2\ 1;0\ 1;0\}\{0;1\}\frac{+2}{(n+1)n}$
1;1	0;1	2;0	0000	$\{0;3\ 2;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\{0;1\}\ \frac{+2}{(n+1)n}\sqrt{\frac{1}{n+2}}$
1;2	0;2	1;0	0000	$\{0;1\}\frac{+2}{(n+2)(n+1)n}$
0;3	**2;1**	**1^2;0**		
1;0	0;1	0;2	0000	$+\frac{1}{n}\sqrt{\frac{2}{(n+2)(n-1)}}$
1;1	0;2	1;0	0000	$\{0;3\ 2;0\ 1;0\}\{1;2\ 2;0\ 0;1\}\{1;1\ 1;0\ 0;1\}\ \frac{+2}{n}\sqrt{\frac{1}{(n+2)(n+1)(n-1)}}$
1^2;1	0;1	2;0	0000	$\{0;2\ 1;0\ 1;0\}\ \frac{+2}{(n+1)n}\sqrt{\frac{2}{(n+2)(n-1)}}$
0;3	**2;1**	**2;0**		
1;0	0;1	0;2	0000	$\frac{+1}{(n+1)n}\sqrt{\frac{2(n+3)}{n+2}}$
1;0	1;2	2;0	0000	$\{0;2\ 1;0\ 1;0\}\{0;1\}\frac{-4}{(n+2)(n+1)n(n-1)}$
1;1	0;2	1;0	0000	$\{0;3\ 2;0\ 1;0\}\{1;2\ 2;0\ 0;1\}\{1;1\ 1;0\ 0;1\}\ \frac{+2}{(n+2)(n+1)}\sqrt{\frac{n+3}{n(n-1)}}$
2;1	0;1	2;0	0000	$\{0;2\ 1;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\ \frac{+2}{(n+2)(n+1)}\sqrt{\frac{2}{n(n-1)}}$
0;3	**21;0**	**1;1**		
0;1	0;1	2;0	0000	$\frac{+1}{n+1}\sqrt{\frac{2}{n(n-1)}}$
0;1	1;2	1^2;0	0000	$\{0;2\ 1;0\ 1;0\}\{0;1\}\ \frac{+1}{n(n-1)}\sqrt{\frac{6}{(n+2)(n+1)}}$
0;1	1;2	2;0	0000	$\{0;1\}\ \frac{+1}{(n+1)n(n-1)}\sqrt{\frac{2(n-3)}{n+2}}$
$0;1^2$	0;2	1;0	0000	$\{0;3\ 2;0\ 1;0\}\{0;21\ 2;0\ 1;0\}\{1;1\ 1;0\ 0;1\}\{0;1\}\ \frac{-1}{(n+1)n}\sqrt{\frac{6}{n-1}}$
0;2	0;2	1;0	0000	$\{0;3\ 2;0\ 1;0\}\{0;21\ 2;0\ 1;0\}\{0;1^2\ 1;0\ 1;0\}\{0;1\}\ \frac{-1}{(n+1)}\sqrt{\frac{2}{(n+2)n(n-1)}}$

TABLE III. (*Continued.*)

1;1^2	0;1	2;0	0000	$\{0;1^2\ 1;0\ 1;0\}\{0;1\}\dfrac{-1}{(n+1)n}\sqrt{\dfrac{6}{(n+2)(n-1)}}$
1;2	0;1	2;0	0000	$\{0;1\}\dfrac{-1}{(n+2)(n+1)}\sqrt{\dfrac{2}{n(n-1)}}$
3;0	**0;3**	**1;1**		
1;0	1;0	0;2	0000	$\dfrac{+1}{(n+1)n}\sqrt{\dfrac{2(n+3)}{n+2}}$
2;0	2;0	0;1	0000	$\{2;0\ 0;2\ 1;1\}\{0;2\ 1;0\ 1;0\}\dfrac{+2\sqrt{2(n+3)}}{(n+2)(n+1)n}$
2;1	1;0	0;2	0000	$\{1;1\ 1;0\ 0;1\}\{0;1\}\dfrac{+2}{(n+2)(n+1)}\sqrt{\dfrac{2}{n(n-1)}}$

considered. Thus the *2jm* symbol associated with each ket branching becomes only a phase factor for which a sign can always be chosen, that is,

$$\begin{pmatrix} \gamma^* & \gamma \\ 0 & 0 \\ \eta^*\kappa^* & \eta\kappa \end{pmatrix} = \begin{pmatrix} \gamma \\ \eta\kappa \end{pmatrix} = \pm 1. \tag{3.6}$$

The symmetry relation

$$\begin{pmatrix} \gamma^* \\ \eta^*\kappa^* \end{pmatrix} = \{\gamma\}\{\eta\}^*\{\kappa\}^*\begin{pmatrix} \gamma \\ \eta\kappa \end{pmatrix} \tag{3.7}$$

determines the relationship between *2jm* phases of complex conjugate pairs of ket branchings.

The *3jm* symbols are zero unless the top and bottom rows form triads of the group and subgroup, respectively, and unless the columns form ket branchings. The symmetries of the *3jm* (Refs. 9–11) are as follows: invariance under cyclic permutations of the columns; a possible sign change for a column interchange, for example, the (12) interchange is

$$\begin{aligned}&\begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 \\ \eta_1\kappa_1 & \eta_2\kappa_2 & \eta_3\kappa_3 \end{pmatrix}^r_{st} \\ &\quad= \begin{pmatrix} \gamma_2 & \gamma_1 & \gamma_3 \\ \eta_2\kappa_2 & \eta_1\kappa_1 & \eta_3\kappa_3 \end{pmatrix}^r_{st} \\ &\qquad\times\{\gamma_1\gamma_2\gamma_3 r\}\{\eta_1\eta_2\eta_3 s\}\{\kappa_1\kappa_2\kappa_3 t\}\end{aligned} \tag{3.8}$$

(we are only using simple phase irreps); and a possible sign change under the complex conjugation symmetry

$$\begin{aligned}&= \begin{pmatrix} \gamma_1^* & \gamma_2^* & \gamma_3^* \\ \eta_1^*\kappa_1^* & \eta_2^*\kappa_2^* & \eta_3^*\kappa_3^* \end{pmatrix}^{r^*}_{st} \\ &\quad\times\begin{pmatrix} \gamma_1 \\ \eta_1\kappa_1 \end{pmatrix}\begin{pmatrix} \gamma_2 \\ \eta_2\kappa_2 \end{pmatrix}\begin{pmatrix} \gamma_3 \\ \eta_3\kappa_3 \end{pmatrix}.\end{aligned} \tag{3.9}$$

The list of some $U_{mn}\supset U_m\times U_n$ nontrivial *3jm* symbols is given in Table V. The *3jm* table uses the group triad $(\gamma_1\gamma_2\gamma_3 r)$ as a header. Each subsequent entry gives the allowed subgroup irrep labels η,κ, the subgroup product multiplicity labels s and t, and the algebraic formula. The *3jm* symmetries are also used to reduce the size of the table. The group triad is ordered such that the irreps appear as in Table I with highest power first down to the lowest power.

TABLE IV. $U_{mn}\supset U_m\times U_n$ branching rules. The values of the *2jm* symbols of the ket branchings below are all chosen $+1$. Those associated with the complex conjugate ket branchings are obtained using Eq. (3.7). The values of the m–n transposition phases for each ket branching are $+1$ except $\{1;1^2\ 0\ 0;1\times 0;1\} = -1$ and $\{0;1^3\ 0\ 0;21\times 0;21\} = -1$ [see Eq. (6.7)].

0;0 ↓	0;0×0;0
0;1 ↓	0;1×0;1
1;1 ↓	0;0×1;1 + 1;1×0;0 + 1;1×1;1
0;1^2 ↓	0;1^2×0;2 + 0;2×0;1^2
0;2 ↓	0;1^2×0;1^2 + 0;2×0;2
1;1^2 ↓	0;1×0;1 + 0;1×1;1^2 + 0;1×1;2 + 1;1^2×0;1 + 1;1^2×1;2 + 1;2 ×0;1 + 1;2 ×1;1^2
1;2 ↓	0;1×0;1 + 0;1×1;1^2 + 0;1×1;2 + 1;1^2×0;1 + 1;1^2×1;1^2 + 1;2 ×0;1 + 1;2 ×1;2
0;1^3 ↓	0;1^3×0;3 + 0;21×0;21 + 0;3 ×0;1^3
0;21 ↓	0;1^3×0;21 + 0;21×0;1^3 + 0;21×0;21 + 0;21×0;3 + 0;3×0;21
0;3 ↓	0;1^3×0;1^3 + 0;21×0;21 + 0;3×0;3

IV. THE METHOD OF CALCULATION

The *6j* and *3jm* symbols are calculated recursively by building up from the trivial *6j* and *3jm* symbols. The method takes advantage of the "phase freedoms" within the Racah–Wigner algebra that are allowed by Schur's lemmas. These phase freedoms describe transformations in the product or branching multiplicity space and exist for each triad and ket branching. In the multiplicity-free case, the phase freedom reduces to just a phase; hence the origin of the term phase freedom. As stated earlier, simplifications occur within the Racah–Wigner algebra because the unitary groups are quasiambivalent and no nonsimple phase irreps occur in the present calculation.

Detailed accounts of the method[9,11] with numerical examples have been given elsewhere. But in outline the basic procedure is the solution of linear equations generated by (i) the unitary condition, (ii) the Racah backcoupling relation, and (iii) the Biedenharn–Elliott sum rule, for *6j* symbols; and (i) the unitary conditions and (ii) the Wigner relation, for *3jm* symbols. Since the *6j* symbols are independent of any group-subgroup scheme, these are calculated first. A small set of the *6j* symbols for both group and subgroup are required in the Wigner relation to calculate the *3jm* symbols.

V. THE TRANSPOSE CONJUGATE SYMMETRY

In Sec. II we discussed the similarities in the Kronecker

TABLE V. $U_{mn} \supset U_m \times U_n$ $3jm$ formulas. The boldfaced headings denote a U_{mn} triad of the $3jm$ symbol. Subsequent entries denote the $U_m \times U_n$ subgroup triad followed by the algebraic formula. The $1j$ and $3j$ phases appear with subscripts mn, m, or n corresponding to the unitary group in which its value must be taken. The following notation has also been used:

$$\chi_r = \begin{cases} +1, & \text{if } \{1;1\ 1;1\ 1;1\ r\}_{mn} = +1, \\ +i, & \text{if } \{1;1\ 1;1\ 1;1\ r\}_{mn} = -1, \end{cases} \qquad \delta^r_{00} = \tfrac{1}{2}[1 + \{1;1\ 1;1\ 1;1\ r\}_{mn}],$$

$$\chi_s = \begin{cases} +1, & \text{if } \{1;1\ 1;1\ 1;1\ s\}_{m} = +1, \\ +i, & \text{if } \{1;1\ 1;1\ 1;1\ s\}_{m} = -1, \end{cases} \qquad \delta^r_{s0} = \tfrac{1}{2}[1 + \{1;1\ 1;1\ 1;1\ r\}_{mn}\{1;1\ 1;1\ 1;1\ s\}_m\{1;1\}_n],$$

$$\chi_t = \begin{cases} +1, & \text{if } \{1;1\ 1;1\ 1;1\ t\}_{n} = +1, \\ +i, & \text{if } \{1;1\ 1;1\ 1;1\ t\}_{n} = -1, \end{cases} \qquad \delta^r_{0t} = \tfrac{1}{2}[1 + \{1;1\ 1;1\ 1;1\ r\}_{mn}\{1;1\}_m\{1;1\ 1;1\ 1;1\ t\}_n],$$

$$\delta^r_{st} = \tfrac{1}{2}[1 + \{1;1\ 1;1\ 1;1\ r\}_{mn}\{1;1\ 1;1\ 1;1\ s\}_m\{1;1\ 1;1\ 1;1\ t\}_n].$$

1;1	**1;0**	**0;1**	**0**	
$0;0\times1;1$	$1;0\times1;0$	$0;1\times0;1$	00	$+\sqrt{\frac{(n+1)(n-1)}{(mn+1)(mn-1)}}$
$1;1\times0;0$	$1;0\times1;0$	$0;1\times0;1$	00	$+\sqrt{\frac{(m+1)(m-1)}{(mn+1)(mn-1)}}$
$1;1\times1;1$	$1;0\times1;0$	$0;1\times0;1$	00	$+\sqrt{\frac{(m+1)(m-1)(n+1)(n-1)}{(mn+1)(mn-1)}}$
1;1	**1;1**	**1;1**	**r**	
$0;0\times1;1$	$0;0\times1;1$	$0;0\times1;1$	$0t$	$\{0;1\}_{mn}\{0;1\}_m\chi_r^*\chi_t\delta^r_{0t}\sqrt{\frac{(n+2\delta_t)(n+1)(n-1)(n-2\delta_t)}{(mn+2\delta_r)(mn+1)(mn-1)(mn-2\delta_r)}}$
$1;1\times0;0$	$1;1\times0;0$	$1;1\times0;0$	$s0$	$\{0;1\}_{mn}\{0;1\}_m\chi_r^*\chi_s\delta^r_{s0}\sqrt{\frac{(m+2\delta_s)(m+1)(m-1)(m-2\delta_s)}{(mn+2\delta_r)(mn+1)(mn-1)(mn-2\delta_r)}}$
$1;1\times1;1$	$1;1\times0;0$	$0;0\times1;1$	00	$\chi_r^*\delta^r_{00}\sqrt{\frac{(m+1)(m-1)(n+1)(n-1)}{(mn+2\delta_r)(mn+1)(mn-1)(mn-2\delta_r)}}$
$1;1\times1;1$	$1;1\times1;1$	$0;0\times1;1$	$0t$	$\chi_r^*\chi_t\delta^r_{0t}\sqrt{\frac{(m+1)(m-1)(n+2\delta_t)(n+1)(n-1)(n-2\delta_t)}{(mn+2\delta_r)(mn+1)(mn-1)(mn-2\delta_r)}}$
$1;1\times1;1$	$1;1\times1;1$	$1;1\times0;0$	$s0$	$\chi_r^*\chi_s\delta^r_{s0}\sqrt{\frac{(m+2\delta_s)(m+1)(m-1)(m-2\delta_s)(n+1)(n-1)}{(mn+2\delta_r)(mn+1)(mn-1)(mn-2\delta_r)}}$
$1;1\times1;1$	$1;1\times1;1$	$1;1\times1;1$	st	$\chi_r^*\chi_s\chi_t\delta^r_{st}\sqrt{\frac{(m+2\delta_s)(m+1)(m-1)(m-2\delta_s)(n+2\delta_t)(n+1)(n-1)(n-2\delta_t)}{(mn+2\delta_r)(mn+1)(mn-1)(mn-2\delta_r)}}$
$0;1^2$	**1;0**	**1;0**	**0**	
$0;1^2\times0;2$	$1;0\times1;0$	$1;0\times1;0$	00	$\{0;1^2\ 1;0\ 1;0\}_{mn}\sqrt{\frac{(m-1)(n+1)}{(mn-1)2}}$
$0;2\times0;1^2$	$1;0\times1;0$	$1;0\times1;0$	00	$\{0;1^2\ 1;0\ 1;0\}_{mn}\sqrt{\frac{(m+1)(n-1)}{(mn-1)2}}$
$1^2;0$	**$0;1^2$**	**1;1**	**0**	
$1^2;0\times2;0$	$0;1^2\times0;2$	$0;0\times1;1$	00	$\{0;1^2\ 1;0\ 1;0\}_m\{0;1\}_m\sqrt{\frac{(m-1)(n+2)(n+1)(n-1)}{(mn+1)(mn-1)(mn-2)2}}$
$1^2;0\times2;0$	$0;1^2\times0;2$	$1;1\times0;0$	00	$\{0;2\ 1;0\ 1;0\}_n\{0;1\}_n\sqrt{\frac{(m+1)(m-1)(m-2)(n+1)}{(mn+1)(mn-1)(mn-2)2}}$
$1^2;0\times2;0$	$0;1^2\times0;2$	$1;1\times1;1$	00	$\{0;1\}_m\sqrt{\frac{(m+1)(m-1)(m-2)(n+2)(n+1)(n-1)}{(mn+1)(mn-1)(mn-2)4}}$
$2;0\times1^2;0$	$0;1^2\times0;2$	$1;1\times1;1$	00	$+\sqrt{(m+1)m(m-1)(n+1)n(n-1)/(mn+1)(mn-1)(mn-2)4}$
$2;0\times1^2;0$	$0;2\times0;1^2$	$0;0\times1;1$	00	$\{0;2\ 1;0\ 1;0\}_m\{0;1\}_m\sqrt{\frac{(m+1)(n+1)(n-1)(n-2)}{(mn+1)(mn-1)(mn-2)2}}$
$2;0\times1^2;0$	$0;2\times0;1^2$	$1;1\times0;0$	00	$\{0;1^2\ 1;0\ 1;0\}_n\{0;1\}_n\sqrt{\frac{(m+2)(m+1)(m-1)(n-1)}{(mn+1)(mn-1)(mn-2)2}}$
$2;0\times1^2;0$	$0;2\times0;1^2$	$1;1\times1;1$	00	$\{0;1\}_m\sqrt{\frac{(m+2)(m+1)(m-1)(n+1)(n-1)(n-2)}{(mn+1)(mn-1)(mn-2)4}}$
0;2	**1;0**	**1;0**	**0**	
$0;1^2\times0;1^2$	$1;0\times1;0$	$1;0\times1;0$	00	$\{0;2\ 1;0\ 1;0\}_{mn}\sqrt{\frac{(m+1)(n-1)}{(mn+1)2}}$
$0;2\times0;2$	$1;0\times1;0$	$1;0\times1;0$	00	$\{0;2\ 1;0\ 1;0\}_{mn}\sqrt{\frac{(m+1)(n+1)}{(mn+1)2}}$
0;2	**$1^2;0$**	**1;1**	**0**	
$0;1^2\times0;1^2$	$1^2;0\times2;0$	$0;0\times1;1$	00	$-\{0;1^2\ 1;0\ 1;0\}_m\{0;1\}_{mn}\{0;1\}_m\{0;1\}_n\sqrt{\frac{(m-1)(n+1)n(n-1)}{(mn+1)mn(mn-1)2}}$

TABLE V. (*Continued.*)

$0;1^2\times0;1^2$	$2;0\times1^2;0$	$1;1\times0;0$	00	$-\{0;1^2\,1;0\,1;0\}_n\{0;1\}_{mn}\{0;1\}_m\{0;1\}_n\sqrt{\frac{(m+1)m(m-1)(n-1)}{(mn+1)mn(mn-1)2}}$
$0;1^2\times0;1^2$	$1^2;0\times2;0$	$1;1\times1;1$	00	$-\{0;1\}_{mn}\{0;1\}_m\{0;1\}_n\sqrt{\frac{(m+1)(m-1)(m-2)(n+1)n(n-1)}{(mn+1)mn(mn-1)4}}$
$0;1^2\times0;1^2$	$2;0\times1^2;0$	$1;1\times1;1$	00	$-\{0;1\}_{mn}\{0;1\}_m\{0;1\}_n\sqrt{\frac{(m+1)m(m-1)(n+1)(n-1)(n-2)}{(mn+1)mn(mn-1)4}}$
$0;2\times0;2$	$1^20;\times2;0$	$1;1\times0;0$	00	$-\{0;2\,1;0\,1;0\}_n\{0;1\}_{mn}\{0;1\}_m\{0;1\}_n\sqrt{\frac{(m+1)m(m-1)(n+1)}{(mn+1)mn(mn-1)2}}$
$0;2\times0;2$	$1^2;0\times2;0$	$1;1\times1;1$	00	$-\{0;1\}_{mn}\{0;1\}_m\{0;1\}_n\sqrt{\frac{(m+1)m(m-1)(n+2)(n+1)(n-1)}{(mn+1)mn(mn-1)4}}$
$0;2\times0;2$	$2;0\times1^2;0$	$0;0\times1;1$	00	$-\{0;2\,1;0\,1;0\}_m\{0;1\}_{mn}\{0;1\}_m\{0;1\}_n\sqrt{\frac{(m+1)(n+1)n(n-1)}{(mn+1)mn(mn-1)2}}$
$0;2\times0;2$	$2;0\times1^2;0$	$1;1\times1;1$	00	$-\{0;1\}_{mn}\{0;1\}_m\{0;1\}_n\sqrt{\frac{(m+2)(m+1)(m-1)(n+1)(n-1)}{(mn+1)mn(mn-1)4}}$
2;0	**0;2**	**1;1**	**0**	
$1^2;0\times1^2;0$	$0;1^2\times0;1^2$	$0;0\times1;1$	00	$\{0;1^2\,1;0\,1;0\}_m\{0;1\}_m\sqrt{\frac{(m-1)(n+1)(n-1)(n-2)}{(mn+2)(mn+1)(mn-1)2}}$
$1^2;0\times1^2;0$	$0;1^2\times0;1^2$	$1;1\times0;0$	00	$\{0;1^2\,1;0\,1;0\}_n\{0;1\}_n\sqrt{\frac{(m+1)(m-1)(m-2)(n-1)}{(mn+2)(mn+1)(mn-1)2}}$
$1^2;0\times1^2;0$	$0;1^2\times0;1^2$	$1;1\times1;1$	00	$\{0;1\}_m\sqrt{\frac{(m+1)(m-1)(m-2)(n+1)(n-1)(n-2)}{(mn+2)(mn+1)(mn-1)4}}$
$2;0\times2;0$	$0;1^2\times0;1^2$	$1;1\times1;1$	00	$+\sqrt{\frac{(m+1)m(m-1)(n+1)n(n-1)}{(mn+2)(mn+1)(mn-1)4}}$
$2;0\times2;0$	$0;2\times0;2$	$0;0\times1;1$	00	$\{0;2\,1;0\,1;0\}_m\{0;1\}_m\sqrt{\frac{(m+1)(n+2)(n+1)(n-1)}{(mn+2)(mn+1)(mn-1)2}}$
$2;0\times2;0$	$0;2\times0;2$	$1;1\times0;0$	00	$\{0;2\,1;0\,1;0\}_n\{0;1\}_n\sqrt{\frac{(m+2)(m+1)(m-1)(n+1)}{(mn+2)(mn+1)(mn-1)2}}$
$2;0\times2;0$	$0;2\times0;2$	$1;1\times1;1$	00	$\{0;1\}_m\sqrt{\frac{(m+2)(m+1)(m-1)(n+1)(n-1)(n-2)}{(mn+2)(mn+1)(mn-2)4}}$
1;1²	**1;1**	**1;0**	**0**	
$0;1\times0;1$	$0;0\times1;1$	$1;0\times1;0$	00	$+\sqrt{\frac{(m+1)(m-1)nn}{(mn+1)mn(mn-1)(mn-2)}}$
$0;1\times0;1$	$1;1\times0;0$	$1;0\times1;0$	00	$\{0;1^2\,1;0\,1;0\}_{mn}\sqrt{\frac{mm(n+1)(n-1)}{(mn+1)mn(mn-1)(mn-2)}}$
$0;1\times0;1$	$1;1\times1;1$	$1;0\times1;0$	00	$-\{0;1^2\,1;0\,1;0\}_{mn}\frac{(m-n)}{\sqrt{(mn+1)mn(mn-1)(mn-2)}}$
$0;1\times1;1^2$	$0;0\times1;1$	$1;0\times1;0$	00	$+\sqrt{\frac{(m+1)(n+1)n(n-2)}{(mn+1)mn(mn-2)2}}$
$0;1\times1;1^2$	$1;1\times1;1$	$1;0\times1;0$	00	$\{0;1^2\,1;0\,1;0\}_{mn}\{0;1^2\,1;0\,1;0\}_n\sqrt{\frac{(m-1)(n+1)n(n-2)}{(mn+1)mn(mn-2)2}}$
$0;1\times1;2$	$0;0\times1;1$	$1;0\times1;0$	00	$+\sqrt{\frac{(m-1)(n+2)n(n-1)}{(mn+1)mn(mn-2)2}}$
$0;1\times1;2$	$1;1\times1;1$	$1;0\times1;0$	00	$\{0;1^2\,1;0\,1;0\}_{mn}\{0;2\,1;0\,1;0\}_n\sqrt{\frac{(m+1)(n+2)n(n-1)}{(mn+1)mn(mn-2)2}}$
$1;1^2\times0;1$	$1;1\times0;0$	$1;0\times1;0$	00	$+\sqrt{\frac{(m+1)m(m-2)(n+1)}{(mn+1)mn(mn-2)2}}$
$1;1^2\times0;1$	$1;1\times1;1$	$1;0\times1;0$	00	$\{0;1^2\,1;0\,1;0\}_{mn}\{0;1^2\,1;0\,1;0\}_m\sqrt{\frac{(m+1)m(m-2)(n-1)}{(mn+1)mn(mn-2)2}}$
$1;1^2\times1;2$	$1;1\times1;1$	$1;0\times1;0$	00	$+\sqrt{\frac{(m+1)m(m+2)(n+2)n(n-1)}{(mn+1)mn(mn-2)2}}$
$1;2\times0;1$	$1;1\times0;0$	$1;0\times1;0$	00	$+\sqrt{\frac{(m+2)m(m-1)(n-1)}{(mn+1)mn(mn-2)2}}$
$1;2\times0;1$	$1;1\times1;1$	$1;0\times1;0$	00	$\{0;1^2\,1;0\,1;0\}_{mn}\{0;2\,1;0\,1;0\}_m\sqrt{\frac{(m+2)m(m-1)(n+1)}{(mn+2)mn(mn-1)2}}$
$1;2\times1;1^2$	$1;1\times1;1$	$1;0\times1;0$	00	$+\sqrt{\frac{(m+2)m(m-1)(n+1)n(n-2)}{(mn+1)mn(mn-2)2}}$

TABLE V. (*Continued.*)

$1;1^2$	**$1^2;0$**	**$0;1$**	**0**	
$0;1\times 0;1$	$1^2;0\times 2;0$	$0;1\times 0;1$	00	$-\{0;1^2\ 1;0\ 1;0\}_{mn}\{0;1^2\ 1;0\ 1;0\}_m\sqrt{\frac{(m+1)(n-1)}{(mn+1)(mn-1)(mn-2)}}$
$0;1\times 0;1$	$2;0\times 1^2;0$	$0;1\times 0;1$	00	$+\{0;1^2\ 1;0\ 1;0\}_{mn}\{0;2\ 1;0\ 1;0\}_m\sqrt{\frac{(m-1)(n+1)}{(mn+1)(mn-1)(mn-2)}}$
$0;1\times 1;1^2$	$2;0\times 1^2;0$	$0;1\times 0;1$	00	$+\{0;1^2\ 1;0\ 1;0\}_{mn}\{0;2\ 1;0\ 1;0\}_m\sqrt{\frac{(n+1)(n-2)}{(mn+1)(mn-2)}}$
$0;1\times 1;2$	$1^2;0\times 2;0$	$0;1\times 0;1$	00	$+\{0;1^2\ 1;0\ 1;0\}_{mn}\{0;1^2\ 1;0\ 1;0\}_m\sqrt{\frac{(n+2)(n-1)}{(mn+2)(mn-1)}}$
$1;1^2\times 0;1$	$1^2;0\times 2;0$	$0;1\times 0;1$	00	$+\{0;1^2\ 1;0\ 1;0\}_{mn}\{0;2\ 1;0\ 1;0\}_n\sqrt{\frac{(m+1)(m-2)}{(mn+1)(mn-2)}}$
$1;1^2\times 1;2$	$1^2;0\times 2;0$	$0;1\times 0;1$	00	$+\{0;1^2\ 1;0\ 1;0\}_{mn}\sqrt{\frac{(m+1)(m-2)(n+2)(n-1)}{(mn+1)(mn-2)2}}$
$1;2\times 0;1$	$2;0\times 1^2;0$	$0;1\times 0;1$	00	$+\{0;1^2\ 1;0\ 1;0\}_{mn}\{0;1^2\ 1;0\ 1;0\}_n\sqrt{\frac{(m+2)(m-1)}{(mn+1)(mn-2)}}$
$1;2\times 1;1^2$	$2;0\times 1^2;0$	$0;1\times 0;1$	00	$+\{0;1^2\ 1;0\ 1;0\}_{mn}\sqrt{\frac{(m+2)(m-1)(n+1)(n-2)}{(mn+1)(mn-2)2}}$
$1;2$	**$1;1$**	**$1;0$**	**0**	
$0;1\times 0;1$	$0;0\times 1;1$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+1)(m-1)nn}{(mn+2)(mn+1)mn(mn-1)}}$
$0;1\times 0;1$	$1;1\times 0;0$	$1;0\times 1;0$	00	$\{0;2\ 1;0\ 1;0\}_{mn}\sqrt{\frac{mm(n+1)(n-1)}{(mn+2)(mn+1)mn(mn-1)}}$
$0;1\times 0;1$	$1;1\times 1;1$	$1;0\times 1;0$	00	$-\{0;2\ 1;0\ 1;0\}_{mn}\frac{(m+n)}{\sqrt{(mn+2)(mn+1)mn(mn-1)}}$
$0;1\times 1;1^2$	$0;0\times 1;1$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m-1)(n+1)n(n-2)}{(mn+2)mn(mn-1)2}}$
$0;1\times 1;1^2$	$1;1\times 1;1$	$1;0\times 1;0$	00	$\{0;2\ 1;0\ 1;0\}_{mn}\{0;1^2\ 1;0\ 1;0\}_n\sqrt{\frac{(m+1)(n+1)n(n-2)}{(mn+2)mn(mn-1)2}}$
$0;1\times 1;2$	$0;0\times 1;1$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+1)(n+2)n(n-1)}{(mn+2)mn(mn-1)2}}$
$0;1\times 1;2$	$1;1\times 1;1$	$1;0\times 1;0$	00	$\{0;2\ 1;0\ 1;0\}_{mn}\{0;2\ 1;0\ 1;0\}_n\sqrt{\frac{(m-1)(n+2)n(n-1)}{(mn+2)mn(mn-1)2}}$
$1;1^2\times 0;1$	$1;1\times 0;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+1)m(m-2)(n-1)}{(mn+2)mn(mn-1)2}}$
$1;1^2\times 0;1$	$1;1\times 1;1$	$1;0\times 1;0$	00	$\{0;2\ 1;0\ 1;0\}_{mn}\{0;1^2\ 1;0\ 1;0\}_m\sqrt{\frac{(m+1)m(m-2)(n+1)}{(mn+2)mn(mn-1)2}}$
$1;1^2\times 1;1^2$	$1;1\times 1;1$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+1)m(m-2)(n+1)n(n-2)}{(mn+2)mn(mn-1)2}}$
$1;2\times 0;1$	$1;1\times 0;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+2)m(m-1)(n+1)}{(mn+2)mn(mn-1)}}$
$1;2\times 0;1$	$1;1\times 1;1$	$1;0\times 1;0$	00	$\{0;2\ 1;0\ 1;0\}_{mn}\{0;2\ 1;0\ 1;0\}_m\sqrt{\frac{(m+2)m(m-1)(n-1)}{(mn+2)mn(mn-1)2}}$
$1;2\times 1;2$	$1;1\times 1;1$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+2)m(m-1)(n+2)n(n-1)}{(mn+2)mn(mn-1)2}}$
$1;2$	**$2;0$**	**$0;1$**	**0**	
$0;1\times 0;1$	$1^2;0\times 1^2;0$	$0;1\times 0;1$	00	$-\{0;2\ 1;0\ 1;0\}_{mn}\{0;1^2\ 1;0\ 1;0\}_m\sqrt{\frac{(m+1)(n+1)}{(mn+2)(mn+1)(mn-1)}}$
$0;1\times 0;1$	$2;0\times 2;0$	$0;1\times 0;1$	00	$+\{0;2\ 1;0\ 1;0\}_{mn}\{0;2\ 1;0\ 1;0\}_m\sqrt{\frac{(m-1)(n-1)}{(mn+2)(mn+1)(mn-1)}}$
$0;1\times 1;1^2$	$1^2;0\times 1^2;0$	$0;1\times 0;1$	00	$+\{0;2\ 1;0\ 1;0\}_{mn}\{0;1^2\ 1;0\ 1;0\}_m\sqrt{\frac{(n+1)(n-2)}{(mn+2)(mn-1)}}$
$0;1\times 1;2$	$2;0\times 2;0$	$0;1\times 0;1$	00	$+\{0;2\ 1;0\ 1;0\}_{mn}\{0;2\ 1;0\ 1;0\}_m\sqrt{\frac{(n+2)(n-1)}{(mn+2)(mn-1)}}$
$1;1^2\times 0;1$	$1^2;0\times 1^2;0$	$0;1\times 0;1$	00	$+\{0;2\ 1;0\ 1;0\}_{mn}\{0;1^2\ 1;0\ 1;0\}_n\sqrt{\frac{(m+1)(m-2)}{(mn+2)(mn-1)}}$
$1;1^2\times 1;1^2$	$1^2;0\times 1^2;0$	$0;1\times 0;1$	00	$+\{0;2\ 1;0\ 1;0\}_{mn}\sqrt{\frac{(m+1)(m-2)(n+1)(n-2)}{(mn+2)(mn-1)2}}$

TABLE V. (*Continued.*)

$1;2\times 0;1$	$2;0\times 2;0$	$0;1\times 0;1$	00	$+\{0;2\ 1;0\ 1;0\}_{mn}\{0;2\ 1;0\ 1;0\}_{n}\sqrt{\frac{(m+2)(m+1)}{(mn+2)(mn-1)}}$
$1;2\times 1;2$	$2;0\times 2;0$	$0;1\times 0;1$	00	$+\{0;2\ 1;0\ 1;0\}_{mn}\sqrt{\frac{(m+2)(m-1)(n+2)(n-1)}{(mn+2)(mn-1)2}}$
$0;1^3$	**$1^2;0$**	**$1;0$**	**0**	
$0;1^3\times 0;3$	$1^2;0\times 2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m-1)(m-2)(n+2)(n+1)}{(mn-1)(mn-2)6}}$
$0;21\times 0;21$	$1^2;0\times 2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+1)(m-1)(n+1)(n-1)}{(mn-1)(mn-2)3}}$
$0;21\times 0;21$	$2;0\times 1^2;0$	$1;0\times 1;0$	00	$\{0;1^2\ 1;0\ 1;0\}_{mn}\sqrt{\frac{(m+1)(m-1)(n+1)(n-1)}{(mn-1)(mn-2)3}}$
$0;3\times 0;1^3$	$2;0\times 1^2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+2)(m+1)(n-1)(n-2)}{(mn-1)(mn-2)6}}$
$0;21$	**$1^2;0$**	**$1;0$**	**0**	
$0;1^3\times 0;21$	$1^2;0\times 2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m-1)(m-2)(n+1)(n-1)}{(mn+1)(mn-1)6}}$
$0;21\times 0;1^3$	$2;0\times 1^2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+1)(m-1)(n-1)(n-2)}{(mn+1)(mn-1)6}}$
$0;21\times 0;21$	$1^2;0\times 2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+1)(m-1)(n+1)(n-1)}{(mn+1)(mn-1)6}}$
$0;21\times 0;21$	$2;0\times 1^2;0$	$1;0\times 1;0$	00	$-\{0;1^2\ 1;0\ 1;0\}_{mn}\sqrt{\frac{(m+1)(m-1)(n+1)(n-1)}{(mn+1)(mn-1)6}}$
$0;21\times 0;3$	$1^2;0\times 2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+1)(m-1)(n+2)(n+1)}{(mn+1)(mn-1)6}}$
$0;3\times 0;21$	$2;0\times 1^2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+2)(m+1)(n+1)(n-1)}{(mn+1)(mn-1)6}}$
$0;21$	**$2;0$**	**$1;0$**	**0**	
$0;1^3\times 0;21$	$1^2;0\times 1^2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m-1)(m-2)(n+1)(n-1)}{(mn+1)(mn-1)6}}$
$0;21\times 0;1^3$	$1^2;0\times 1^2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+1)(m-1)(n-1)(n-2)}{(mn+1)(mn-1)6}}$
$0;21\times 0;21$	$1^2;0\times 1^2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+1)(m-1)(n+1)(n-1)}{(mn+1)(mn-1)6}}$
$0;21\times 0;21$	$2;0\times 2;0$	$1;0\times 1;0$	00	$-\{0;2\ 1;0\ 1;0\}_{mn}\sqrt{\frac{(m+1)(m-1)(n+1)(n-1)}{(mn+1)(mn-1)6}}$
$0;21\times 0;3$	$2;0\times 2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+1)(m-1)(n+2)(n+1)}{(mn+1)(mn-1)6}}$
$0;3\times 0;21$	$2;0\times 2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+2)(m+1)(n+1)(n-1)}{(mn+1)(mn-1)6}}$
$0;3$	**$2;0$**	**$1;0$**	**0**	
$0;1^3\times 0;1^3$	$1^2;0\times 1^2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m-1)(m-2)(n-1)(n-2)}{(mn+2)(mn+1)6}}$
$0;21\times 0;21$	$1^2;0\times 1^2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+1)(m-1)(n+1)(n-1)}{(mn+2)(mn+1)3}}$
$0;21\times 0;21$	$2;0\times 2;0$	$1;0\times 1;0$	00	$\{0;2\ 1;0\ 1;0\}_{mn}\sqrt{\frac{(m+1)(m-1)(n+1)(n-1)}{(mn+2)(mn+1)3}}$
$0;3\times 0;3$	$2;0\times 2;0$	$1;0\times 1;0$	00	$+\sqrt{\frac{(m+2)(m+1)(n+2)(n+1)}{(mn+2)(mn+1)6}}$

product rule and the branching rules arising from the transpose conjugate symmetry of partitions (λ) and $(\tilde{\lambda})$. We also noted how the dimension formula for one irrep can be obtained from that of its transpose conjugate partner. This transpose conjugate symmetry carries over to the U_n $6j$ symbols and $\mathrm{U}_{mn}\supset \mathrm{U}_m\times \mathrm{U}_n$ $3jm$ symbols, and relates the pair of $6j$ symbols

$$\begin{Bmatrix}\gamma_1 & \gamma_2 & \gamma_3\\ \eta_1 & \eta_2 & \eta_3\end{Bmatrix}_{r_1r_2r_3r_4}$$

and (5.1)

$$\begin{Bmatrix}\tilde{\gamma}_1 & \tilde{\gamma}_2 & \tilde{\gamma}_3\\ \tilde{\eta}_1 & \tilde{\eta}_2 & \tilde{\eta}_3\end{Bmatrix}_{r_1r_2r_3r_4}$$

and the four $3jm$ symbols

$$\begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 \\ \eta_1\kappa_1 & \eta_2\kappa_2 & \eta_3\kappa_3 \end{pmatrix}^r_{st},$$
$$\begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 \\ \tilde\eta_1\tilde\kappa_1 & \tilde\eta_2\tilde\kappa_2 & \tilde\eta_3\tilde\kappa_3 \end{pmatrix}^r_{st},$$
$$\begin{pmatrix} \tilde\gamma_1 & \tilde\gamma_2 & \tilde\gamma_3 \\ \eta_1\tilde\kappa_1 & \eta_2\tilde\kappa_2 & \eta_3\tilde\kappa_3 \end{pmatrix}^r_{st}.$$
and (5.2)
$$\begin{pmatrix} \tilde\gamma_1 & \tilde\gamma_2 & \tilde\gamma_3 \\ \tilde\eta_1\kappa_1 & \tilde\eta_2\kappa_2 & \tilde\eta_3\kappa_3 \end{pmatrix}^r_{st}.$$
Here γ,η,κ denote composite labels. If $\{\gamma\} = \{\mu;\nu\}$ then $\{\tilde\gamma\} = \{\tilde\mu;\tilde\nu\}$. Note we have also used η as a label for U_n in (5.1) and U_m in (5.2). However, no confusion should arise.

As a consequence of the combinatoric similarities arising from the transpose conjugate symmetry, the equations used to solve for pairs of $6j$ symbols related by (5.1) or pairs of $3jm$ symbols related by (5.2) are closely related. Indeed the solution, that is, the algebraic formula, of one may be obtained from the other by a simple procedure.

(1) For $6j$ symbols, (i) replace all factors $(n+a)$ by $(n-a)$, (ii) replace any $3j$ phase $\{\gamma_1\gamma_2\gamma_3 r\}$ by the transpose conjugate $3j$ phase $\{\tilde\gamma_1\tilde\gamma_2\tilde\gamma_3 r\}$, and (iii) replacing any $1j$ phase $\{\gamma\}$ by $\{\tilde\gamma\}$. We note that if a $6j$ symbol is a self-transpose conjugate, to within the row and column symmetries discussed in Sec. III, the algebraic formula will reflect this by having pairs of factors $(n+a)(n-a)$, and having self-transpose conjugate $3j$ (or $1j$) phases, or pairs of $3j$ (or $1j$) phases which are transpose conjugates.

(2) For $3jm$ symbols, if the irreps of any two of the groups U_{mn}, U_m, or U_n have been transpose conjugated then (i) the corresponding factor $(mn+a)$, $(m+a)$, or $(n-a)$ must be replaced by $(mn-a)$, $(m-a)$, or $(n-a)$, respectively, (ii) the $3j$ and $1j$ phases of these two groups must be replaced by their tranpose conjugate $3j$ and $1j$ phases, and (iii) in $2jm$ phases the irrep label of the two groups must be replaced by the transpose conjugated irrep. Similar to $6j$ symbols, self-transpose conjugate $3jm$ symbols have pairs of factors $(mn+a)(mn-a)$, $(m+a)(m-a)$, or $(n+a)(n-a)$ and self-transpose conjugate pairs of $3j$, $1j$, or $2j$ phases.

During the calculation, the phase freedom choices[5,9,11] are chosen in a manner so that the above transpose conjugate symmetry is satisfied. These phase choices sometimes impose conditions on some $3j$ phases and transposition phases (see Sec. VI) which would otherwise be chosen freely. For example, we have from the $6j$ calculation

$$\begin{Bmatrix} 0;2 & 1^2;0 & 1;1 \\ 0;1 & 0;1 & 1;0 \end{Bmatrix}_{0000} = \sqrt{1/(n+1)n(n-1)}, \tag{5.3}$$

$$\begin{aligned}\begin{Bmatrix} \tilde0;\tilde2 & \tilde1^2;\tilde0 & \tilde1;\tilde1 \\ \tilde0;\tilde1 & \tilde0;\tilde1 & \tilde1;\tilde0 \end{Bmatrix}_{0000} &= \begin{Bmatrix} 0;1^2 & 2;0 & 1;1 \\ 0;1 & 0;1 & 1;0 \end{Bmatrix}_{0000} \\ &= \begin{Bmatrix} 0;2 & 1^2;0 & 1;1 \\ 0;1 & 0;1 & 1;0 \end{Bmatrix}^*_{0000} \\ &\quad \times\{0;2\ 1^2;0\ 1;1\ 0\}\{0;2\ 1;0\ 1;0\ 0\} \\ &\quad \times\{0;1^2\ 1;0\ 1;0\ 0\}\{1;1\ 1;0\ 0;1\ 0\}\{0;1\ 0\},\end{aligned} \tag{5.4}$$

after applying a (12) column interchange and a complex conjugation. If the transpose conjugate symmetry is not to introduce additional phase factors then the product of $3j$ phases in (5.3) must be chosen as unity. Such a choice also satisfies the requirement that a multiplicity-free $6j$ symbol be invariant under an odd column interchange. Furthermore, it implies that this $6j$ symbol is real.

VI. THE m–n TRANSPOSITION SYMMETRY

The $U_{mn}\supset U_m\times U_n$ has a further symmetry due to the occurrence of the direct product subgroup $U_m\times U_n$. The subgroup obtained by transposing U_m and U_n, that is, $U_n\times U_m$, is also contained in U_{mn}. The relationship between the basis vectors of the two group-subgroup schemes is written formally as

$$\begin{aligned}&\tau\cdot|\gamma(U_{mn})a\eta(U_m)j\kappa(U_n)k\rangle \\ &\quad = |\gamma(U_{mn})a'\kappa(U_n)k\eta(U_m)j\rangle\, T(\gamma,\eta\kappa)^{a'}{}_a,\end{aligned} \tag{6.1}$$

where τ is the transposition operator acting on the direct product subgroup and $T(\gamma,\eta\kappa)^{a'}{}_a$ are elements, which we call transposition factors, of an $m^\gamma_{\eta\kappa}\times m^\gamma_{\eta\kappa}$ unitary matrix. In giving this result we have required that the irrep matrices of $U_m\times U_n$ and $U_n\times U_m$ be the same. The factorization of the subgroup basis then follows from Schur's lemmas. The involutary nature of the transposition operator gives the relation (in matrix form)

$$T(\gamma,\kappa\eta) = T(\gamma,\eta\kappa)^\dagger, \tag{6.2}$$

and the complex conjugation symmetry appropriate to any transformation factor gives

$$T(\gamma^*,\eta^*\kappa^*) = A(\gamma,\kappa\eta)T(\gamma,\eta\kappa)^*A(\gamma,\eta\kappa)^{\dagger *}, \tag{6.3}$$

where $A(\gamma,\eta\kappa)$ is the complex conjugation matrix formed from the $2jm$ symbols

$$A(\gamma,\eta\kappa)^{a'}{}_a = \begin{pmatrix} \gamma^* & \gamma \\ a' & a \\ \eta^*\kappa^* & \eta\kappa \end{pmatrix}. \tag{6.4}$$

The transportation factors can be chosen diagonal

$$T(\gamma,\eta\kappa)^{a'}{}_a = \{\gamma a\eta\kappa\}\delta^{a'}{}_a,$$

consistent with the following choice of the complex conjugation:

$$A(\gamma,\eta\kappa) = \begin{cases} 1, & \text{if } \{\gamma\}\{\eta\}^*\{\kappa\}^* = +1, \\ J, & \text{if } \{\gamma\}\{\eta\}^*\{\kappa\}^* = -1. \end{cases} \tag{6.5}$$

(J is the symplectic matrix

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

and 1 the unit matrix.) The transposition phase (or τ phase), $\{\gamma a\eta\kappa\}$, satisfies from Eqs. (6.2) and (6.3)

$$\{\gamma a\kappa\eta\} = \{\gamma a\eta\kappa\}^* = \{\gamma^* a\eta^*\kappa^*\}. \tag{6.6}$$

For $\eta(U_m)\neq\kappa(U_n)$ the transposition phases are arbitrary subject only to the conditions (6.6). The interesting case occurs when $\eta(U_m)=\kappa(U_n)$, for which the transposition phase $\{\gamma a\eta\eta\}$ is only a sign. This sign is determined by the inner plethysm of Schur functions in the following manner. If γ and η are both irrep labels of the form $\{0;\lambda\}$ and are partitions of the same integer, then

$$\{\gamma a\eta\eta\} = \begin{cases} +1, & \text{if } \eta\odot\{2\} \text{ contains the } a\text{th occurrence of } \gamma, \\ -1, & \text{if } \eta\odot\{1^2\} \text{ contains the } a\text{th occurrence of } \gamma \end{cases} \tag{6.7}$$

($\odot$ denotes the inner plethysm operation).

If γ and η are not of the form $\{0;\lambda\}$ we use (2.27) and write $\{\gamma a\eta\eta\}$ as a product of two transposition phases, one labeled by one-dimensional irreps and the other by irrep labels of the form $\{0;\lambda\}$. Both these phases are determined by (6.7). As examples we give for $U_{n^2}\supset U_n\times U_n$

$$\begin{aligned} &\{0;1\,0\,0;1\,0;1\} = +1,\\ &\{0;2\,0\,0;2\,0;2\} = \{0;2\,0\,0;1^2\,0;1^2\} = +1, \\ &\{0;1^2\,0\,0;2\,0;1^2\} = \{0;1^2\,0\,0;1^2\,0;2\} = +1, \end{aligned} \tag{6.8}$$

while we have

$$\begin{aligned} &\{1;1\,0\,1;1\,1;1\}\\ &\quad = \{-1^{n^2}0-n^n\ -n^n\}\\ &\qquad \times\{0;21^{n-2}\,0\,0;n+1,n^{n-2},n-1\\ &\qquad\quad 0;n+1,n^{n-2},n-1\}, \\ &\{1;1^2\,0\,0;1\,0;1\}\\ &\quad = \{-1^{n^2}0\ -n^n\ -n^n\}\\ &\qquad \times\{0;2^21^{n-3}\,0\,0;n+1,n^{n-1}\,0;n+1,n^{n-1}\}. \end{aligned} \tag{6.9}$$

From (6.6) and (6.7)

$$\{-1^{n^2}0-n^n-n^n\} = \{1^{n^2}0\,n^n\,n^n\}^* = (-)^z,$$

where $z=n(n-1)/2$. [see Ref. 21, Eq. (6.12)]. The two other transposition phases of (6.9) can be determined from Eq. (6.2) of Ref. 21

$$\begin{aligned} &\{0;21^{n^2-2}\,0\,0;n+1,n^{n-2},n-1\,0;n+1,n^{n-2},n-1\}\\ &\quad = (-)^z,\\ &\{0;2^21^{n^2-3}\,0\,0;n+1,n^{n-1}\,0;n+1,n^{n-1}\}\\ &\quad = (-)^{z+1}. \end{aligned}$$

Hence we have

$$\begin{aligned} &\{1;1\,0\,1;1\,1;1\} = +1,\\ &\{1;1^2\,0\,0;1\,0;1\} = -1. \end{aligned}$$

The values of these phases have a striking consequence for one particular $3jm$ symbol, as we will see shortly.

Using (6.1), the m–n transposition symmetry for the $3jm$ symbols can be written

$$\begin{aligned} &\begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3\\ a_1 & a_2 & a_3\\ \eta_1\kappa_1 & \eta_2\kappa_2 & \eta_3\kappa_3 \end{pmatrix}^r_{st}\\ &\quad = \begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3\\ a_1' & a_2' & a_3'\\ \kappa_1\eta_1 & \kappa_2\eta_2 & \kappa_3\eta_3 \end{pmatrix}^{r'}_{st}\\ &\qquad \times T(\gamma_1,\eta_1\kappa_1)^{a_1'}{}_{a_1} T(\gamma_2,\eta_2\kappa_2)^{a_2'}{}_{a_2}\\ &\qquad \times T(\gamma_3,\eta_3\kappa_3)^{a_3'}{}_{a_3}. \end{aligned} \tag{6.10}$$

The prime is used here to distinguish the two sets of $3jm$ symbols, those belonging to $U_{mn}\supset U_m\times U_n$ and those of $U_{mn}\supset U_m\times U_n$. Since no branching multiplicity occurs and the transposition factors are diagonal, (6.9) simplifies to

$$\begin{aligned} &\begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3\\ \eta_1\kappa_1 & \eta_2\kappa_2 & \eta_3\kappa_3 \end{pmatrix}^r_{st}\\ &\quad = \begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3\\ \kappa_1\eta_1 & \kappa_2\eta_2 & \kappa_3\eta_3 \end{pmatrix}^{r'}_{st}\\ &\qquad \times\{\gamma_1 0\eta_1\kappa_1\}\{\gamma_2 0\eta_2\kappa_2\}\{\gamma_3 0\eta_3\kappa_3\}. \end{aligned} \tag{6.11}$$

An immediate result is that the magnitudes of prime and unprimed $3jm$ symbols in (6.9) are equal while the phase difference is given by the product of transposition phases.

Arising in the $3jm$ table is a striking example for which the m–n transposition phases determine a nontrivial zero. The product of transposition phases for the $U_{n^2}\supset U_n\times U_n$ $3jm$ symbol

$$\begin{pmatrix} 1;1^2 & 1;1 & 1;0\\ 0;1\times 0;1 & 1;1\times 1;1 & 1;0\times 1;0 \end{pmatrix} \tag{6.12}$$

is fixed and from (6.8)–(6.12) has the value

$$\begin{aligned} &\{1;1^2\,0\,0;1\,0;1\}\{1;1\,0\,1;1\,1;1\}\\ &\quad\times\{1;0\,0\,1;0\,1;0\} = -1. \end{aligned}$$

Hence by (6.11) the $3jm$ of (6.12) must be zero for $m=n$. The algebraic formula must have an $(m-n)$ dependency. Indeed the formula obtained using the recursive method has such a dependency. This is the first occurrence of the $(m-n)$ factor—all other $3jm$ symbols have factors of the form $(mn+a)$, $(m+b)$, and $(n+c)$.

VII. THE COMPOSITE LABELING MODIFICATION RULES

In the last section we showed that the m–n transportation symmetry can determine a vanishing of the $3jm$ symbol for certain values of m and n. A partial explanation, which is linked to the modification rules for the U_n irreps, can be given for vanishings of the algebraic formulas for the other nontrivial $6j$ and $3jm$ symbols. Consider the irrep label $\{1;1^2\}$ for various n. From the modification rules this label vanishes for $n=0,2$ while its transpose conjugate label $\{1;2\}$ vanishes for $n=1$. Thus the dimension formula for $\{1;1^2\}$ must contain factors which vanish for $n=0,2,-1$. The latter value is given by the transpose symmetry. These three factors must be the only ones and so

$$\{1;1^2\}\propto(n+1)(n)(n-2). \tag{7.1}$$

In a similar manner many of the factors $(n-a)$ appearing in the tables of $6j$ and $3jm$ symbols may be explained by this argument. However, in some cases the tabulated result has arisen after cancellation of pairs of such factors and no explicit $(n-a)$ factor appears despite the inadmissibility of an irrep label for $n=a$. It must be remembered though that when an inadmissible label occurs the $6j$ or $3jm$ symbol no longer satisfies the triad and ket branching criterion and hence of necessity must be zero.

A more subtle aspect of the modification rules is illustrated by the label $\{1^2;1^2\}$ for $n=2$. This label occurs in (2.8) and since

$$\{1^2;1^2\} = -\{1;1\}, \quad \text{for } U_2, \tag{7.2}$$

it canceled another term in (2.10). This cancellation changed the multiplicity of $\{1;1\}$ in $\{1;1\}\times\{1;1\}$ or equivalently the U_n triad

$$(\{1;1\}\{1;1\}\{1;1\}r) \tag{7.3}$$

from 2 to 1 for U_2. Thus the appearance of the factor $(n-2\delta_r)$, where δ_r takes on the value 1 or 0, in the algebraic formulas for all *6j* and *3jm* symbols containing triad (7.3) is an obvious consequence of this irrep label cancellation. The tables presented here, however, are of insufficient size to understand the full consequences of modification rules as applied to *6j* and *3jm* symbols.

VIII. CONCLUDING REMARKS

We have obtained a set of algebraic formulas for some nontrivial U_n *6j* symbols and $U_{mn}\supset U_m\times U_n$ *3jm* symbols valid for all *m* and *n*, by a straightforward application of the building-up method previously used to obtain numerical values of SU_6 and SU_3. The composite labeling has allowed the triads of U_n and ket branchings of $U_{mn}\supset U_m\times U_n$ to be written in a rank-independent manner. Linear equations generated from the unitarity condition, Racah backcoupling relation and Biedenharn–Elliott sum rule for *6j* symbols, and the Wigner relation for *3jm* symbols have been solved to give the rank-dependent algebraic formulas for these unitary group *6j* and *3jm* symbols. The combinatoric properties of the composite labels lead to symmetries and vanishings. These symmetries are quite different from the row and column interchange symmetries satisfied by all *6j* and *3jm* symbols of all groups[22–24] and the Regge symmetries[25,26] satisfied by all SU_n. Those symmetries equate to within phases and dimension factors two *6j* symbols (or *3jm* symbols) for different irreps of the one group (respectively, the one-group-subgroup chain). Modification rules for composite labels relate in a more complicated way two *6j* or *3jm* symbols involving different irreps, that is, different triads and ket branchings.

Other rank-dependent formulas for the unitary groups have been obtained by Biedenharn and his co-workers.[27–30] They have exploited the canonical embedding $U_n\supset U_{n-1}$ to derive recursively the matrix elements of coupled and uncoupled tensor operators (equivalently, our *6j* and *3jm* symbols). However, they consider only the regular partition labels for U_n and although the canonical chain $U_n\supset U_{n-1}\cdots\supset U_1$ has the advantage of being multiplicity-free, noncanonical bases such as $U_{mn}U_m\times U_n$ and $U_{m+n}\supset U_m\times U_n$ do occur in physical applications. The rank dependence and the vanishings and cancellations are similar where they focus on the properties of the "shift operator" of the Kronecker products to give a unique product multiplicity separation. Our method of applying modification rules to irrep and multiplicity labels is different.

In our present considerations we have not exploited two important symmetries: the Schur–Weyl duality which connects the unitary group with the symmetric group, and the relationship between U_n and SU_n which is generated by the equivalence of the one-dimensional irreps of U_n in SU_n. The Schur–Weyl duality provides many rank-independent relations between *6j* symbols and similarly *3jm* symbols to within phases and dimension factors. These symmetries, and the formulas that follow from them, have beeen studied by Jucys,[31,32] Vanagas,[33] Kramer and Seligman,[34,35] and more recently Sullivan[36] and Chen.[37] See also Ref. 38. Sullivan in particular has discussed the consistency of phase choices with respect to the Schur–Weyl duality and SU_n. The isomorphism between the direct product groups $C_n\times U_n$ and $U_1\times SU_n$ (where C_n is the *n*-fold cyclic group) provides a further symmetry between associated irreps of U_n, although a rank-dependent one. Bickerstaff[39] discusses this relationship especially in connection with the reality criterion of U_n and SU_n to *6j* and *3jm* symbols. Since C_n and U_1 are Abelian groups it seems possible to choose phases such that numerical values of the U_n *6j* and *3jm* symbols are simultaneously the same as the corresponding SU_n symbol. Both the "U_n–SU_n simultaneity" and the Schur–Weyl duality are attractive symmetries to include since they give great simplifications and ease in calculations of the unitary group *6j* and *3jm* symbols. However, a simple structure for the matrices that express the U_n–SU_n simultaneity and the Schur–Weyl duality may not be compatible with (i) the already-existing choices of permutation matrix and *A* matrix expressing the column interchange and complex conjugation symmetries of *6j* and *3jm* symbols, (ii) the new transpose conjugate and $m-n$ transposition symmetries discussed earlier, and (iii) the special choice of *6j* invariance under column interchange made in our calculation. Indeed the imposition of U_n–SU_n simultaneity can lead to awkward resolutions of product multiplicity if the separation choice is assumed rank-independent.

ACKNOWLEDGMENTS

One of us (RWH) would like to thank Dr. R. Dirl. RWH also acknowledges the support of the Austrian FWF (Fonds zur Förderung der wissenschaftlichen Forschung).

[1]H. Jahn, Proc. R. Soc. London. Ser. A **201**, 516 (1950).
[2]G. Racah, Phys. Rev. **76**, 1352 (1949).
[3]Y. Tanabe and S. Sugano, J. Phys. Soc. Jpn. **9**, 766 (1954).
[4]R. L. Jaffe, Phys. Rev. D **15**, 267 (1977).
[5]R. P. Bickerstaff and B. G. Wybourne, J. Phys. G **7**, 272 (1981).
[6]E. F. Kustov, Phys. Status Solidi B **81**, 421 (1977).
[7]H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).
[8]R. P. Bickerstaff, P. H. Butler, M. B. Butts, R. W. Haase, and M. F. Reid, J. Phys. A **15**, 1087 (1982).
[9]P. H. Butler *Point Group Symmetry Application: Methods and Tables* (Plenum, New York, 1981).
[10]P. H. Butler, Int. J. Quantum Chem. **10**, 599 (1976).
[11]P. H. Butler and B. G. Wybourne, Int. J. Quantum Chem. **10**, 581 (1976).
[12]F. D. Murnaghan *The Theory of Group Representations* (Johns Hopkins, Baltimore, MD 1938).
[13]D. E. Littlewood, *The Theory of Group Characters* (Oxford U.P., 1940).
[14]Y. J. Abramsky and R. C. King, Nuovo Cimento A **67**, 153 (1970).
[15]R. C. King, J. Math. Phys. **11**, 280 (1970).
[16]R. C. King, J. Phys. A **8**, 429 (1975).
[17]G. R. E. Black, R. C. King, and B. G. Wybourne, J. Phys. A **16**, 1555 (1983).
[18]G. Robinson, *Representations of the Symmetry Group* (Univ. of Toronto, Toronto, 1961).
[19]N. El Sarma and R. C. King, J. Phys. A **12**, 2317 (1979).

[20]B. G. Wybourne, *Symmetry Principles and Atomic Spectroscopy* (with an appendix of tables by P. H. Butler) (Wiley, New York, 1970).
[21]P. H. Butler and R. C. King, J. Math. Phys. **14**, 9, 1176 (1972).
[22]J. R. Derome and W. T. Sharp, J. Math. Phys. **6**, 1584 (1965).
[23]J. R. Derome, J. Math. Phys. **7**, 612 (1966).
[24]P. H. Butler, Philos. Trans. R. Soc. London, Ser. A **277** 454 (1975).
[25]T. Regge, Nuovo Cimento **11**, 116 (1959).
[26]P. Kramer and T. H. Seligman, Z. Phys. **219** 105 (1969).
[27]L. C. Biedenharn, J. D. Louck, E. Chacon, and M. Ciftan, J. Math. Phys. **14**, 1957 (1972).
[28]J. D. Louck and L. C. Biedenharn, J. Math. Phys. **14**, 1336 (1973).
[29]J. D. Louck, M. A. Lohe, and L. C. Biedenharn, J. Math. Phys. **16**, 2408 (1975).
[30]M. A. Lohe, L. C. Biedenharn, and J. D. Louck, J. Math. Phys. **18**, 1883 (1977).
[31]A. A. Jucys, Liet. Fiz. Rinkinys **10**, 5 (1970).
[32]S. J. Alisauskas, A. A. Jucys, and A. P. Jucys, J. Math. Phys. **13**, 9, 1329 (1972).
[33]V. V. Vanagas, *Algebraic Methods in Nuclear Physics* (Mintis, Vilnius, USSR, 1971), in Russian.
[34]P. Kramer, Z. Phys. **205**, 181 (1967); **216**, 68 (1968).
[35]P. Kramer and T. H. Seligman, Z. Phys. **219**, 105 (1969); Nucl. Phys. A **136**, 545 (1969).
[36]J. J. Sullivan, J. Math. Phys. **14**, 387 (1973); **16**, 756 (1975); **16**, 1707 (1975); **19**, 1674 (1978); **19**, 1681 (1978); **21**, 227 (1980); **24**, 424 (1983); **24**, 2542 (1983).
[37]C. J. Q., J. Math. Phys. **22**, 1 (1981).
[38]R. W. Haase and P. H. Butler, J. Phys. A **17**, 61 (1984).
[39]R. P. Bickerstaff, J. Math. Phys. **25**, 2808 (1984).

Group theoretical analysis of the Hartree–Fock–Bogoliubov equation. I. General theory

Masa-aki Ozaki
Department of Physics, Kyoto University, Kyoto 606, Japan

(Received 27 September 1984; accepted for publication 1 February 1985)

This paper describes a general group theoretical analysis of the temperature–Hartree–Fock–Bogoliubov (HFB) equation and its solution. The action of the symmetry group G_0 of the system on the HFB Hamiltonian, the HFB density matrix, and the HFB Green function are defined. It is shown that the HFB equation and its solution are classified by a subgroup G of G_0, which is the invariance group of the HFB Hamiltonian, the HFB density matrix, and the HFB Green function corresponding to the solution. General expression of the instability of a solution and its decomposition into R-rep (single-valued irreducible representation over the real number field) components of the invariance group of the solution are obtained. The self-consistent field (SCF) condition is decomposed into R-rep components of G_0.

I. INTRODUCTION

Recently a number of new states of solids with multiple-order parameters have been suggested. These states are the coexisting states of a charge density wave or a magnetism with the superconductivity.[1] Many of them have been discussed by the temperature–Hartree–Fock–Bogoliubov (HFB) equation[2] of several types. In spite of many works on particular types of the HFB equation, there has been no complete general theory on the classification and the interconnection of the HFB solution without reference to the structure of the system. Such a general theory is very much needed for finding systematically the possible phases and phase transitions.

In Fukutome[3] and our previous papers,[4] a theory for classification, characterization of the unrestricted Hartree–Fock (UHF) solution, and a group theoretical bifurcation theory of the UHF equation has been developed.

In this paper we show that the above group theoretical analysis for the UHF equation can be extended to the HFB equation. Here we give the general group theoretical analysis of the HFB equation. The special cases of the system with the symmetry of the lattice group (electrons in solid) or Euclidean group (Fermion gas) will be studied in subsequent papers.[5] In Sec. II we give a short review of the HFB equation. In Sec. III we describe the group action on the HFB Hamiltonian, the HFB density matrix, and the HFB Green function. In Sec. IV we give the group theoretical analysis of the self-consistent field (SCF) condition. In Sec. V we give the group theoretical classification of the HFB equation and its solution. In Sec. VI we give the general expression of the instability of a HFB solution and the decomposition of the instability into R-irreducible components.

II. A SHORT REVIEW OF THE HFB EQUATION

We consider a system of fermions with a Hamiltonian

$$\mathscr{H} = \sum_{i,j=1}^{N} (T_{ij} - \mu\delta_{ij}) a_i^+ a_j + \frac{1}{2} \sum_{i,j,m,n=1}^{N} \langle in|V|jm\rangle a_i^+ a_n^+ a_m a_j, \tag{2.1}$$

where the sums run through a complete set of single particle states (N) and μ is the chemical potential. We take the HFB Hamiltonian as follows:

$$\begin{aligned} H(Z) &= \sum_{i,j=1}^{N} \{(x'_{ij} - \mu\delta_{ij}) a_i^+ a_j + \tfrac{1}{2} y_{ij} a_i^+ a_j^+ + \tfrac{1}{2}(y^+)_{ij} a_i a_j\} \\ &= \frac{1}{2}(A^+ZA) + \frac{1}{2}\sum_{i=1}^{N} x_{ii}, \end{aligned} \tag{2.2}$$

where

$$\begin{aligned} &A^+ = (a_1^+,...,a_N^+, a_1,...,a_N), \\ &x_{ij} = x'_{ij} - \mu\delta_{ij}, \quad x^+ = x, \quad y^t = -y, \\ &Z = \begin{pmatrix} x, & y \\ -y^*, & -x^* \end{pmatrix}, \end{aligned} \tag{2.3}$$

and x^+ denotes the Hermite conjugate of x and y^t denotes the transposed matrix of y. The HFB free energy $F(Z)$ is given by

$$F(Z) = \langle \mathscr{H} - H(Z)\rangle_Z - (1/\beta)\log Q(Z), \tag{2.4}$$

where

$$Q(Z) = \mathrm{Tr}[e^{-\beta H(Z)}], \tag{2.5}$$

$$\langle\cdots\rangle_Z = \mathrm{Tr}[e^{-\beta H(Z)}\cdots/Q(Z)], \tag{2.6}$$

and $\beta = (1/kT)$. We note the second term $\frac{1}{2}\Sigma x_{ii}$ in (2.2) does not contribute to the HFB free energy $F(Z)$. Then in the following we can omit this term.

The HFB density matrix $R(Z)$ corresponding to $H(Z)$ is given by

$$R(Z) = \langle AA^+\rangle_Z = \begin{pmatrix} \rho & \lambda \\ -\lambda^* & 1-\rho^* \end{pmatrix}, \tag{2.7}$$

where

$$\rho(Z)_{ij} = \langle a_j^+ a_i\rangle_Z, \quad \lambda(Z)_{ij} = \langle a_j a_i\rangle_Z, \tag{2.8}$$

and they satisfy

$$\rho^+ = \rho, \quad \lambda^t = -\lambda. \tag{2.9}$$

$R(Z)$ is expressed explicitly by Z as

$$R(Z) = (1 + e^{\beta Z})^{-1}. \tag{2.10}$$

 0022-2488/85/071514-07$02.50

Thus we can consider R an independent variable instead of Z. In terms of R the HFB free energy $F(R)$ is given by

$$F(R) = \sum_{i,j=1}^{N} T_{ij}\rho_{ji} + \frac{1}{2}\sum_{\substack{i,j\\m,n}} [in|V|jm]\rho_{ji}\rho_{mn} + \frac{1}{2}\sum_{\substack{i,j\\m,n}} \langle in|V|jm\rangle(\lambda^{+})_{ni}\lambda_{jm} + \tfrac{1}{2}(1/\beta)\mathrm{tr}[R\ln R + (1-R)\ln(1-R)], \tag{2.11}$$

where

$$[in|V|jm] = \langle in|V|jm\rangle - \langle in|V|mj\rangle \tag{2.12}$$

and tr denotes the trace of the matrix. Here Z (or R) which minimizes (2.4) [or (2.11)] is the HFB solution.

The first-order variation of $F(R)$ by R is

$$\delta F = \tfrac{1}{2}\,\mathrm{tr}[\delta R\,[\tilde{\epsilon} + (1/\beta)\ln\{R(1-R)^{-1}\}], \tag{2.13}$$

where

$$\tilde{\epsilon} = \begin{pmatrix} \epsilon & \Delta \\ -\Delta^* & -\epsilon^* \end{pmatrix}, \tag{2.14}$$

$$\epsilon_{ij} = T_{ij} - \mu\delta_{ij} + \sum_{mn}[in|V|jm]\rho_{mn}, \qquad \Delta_{in} = \sum_{jm}\langle in|V|jm\rangle\lambda_{jm}. \tag{2.15}$$

Thus we have the HFB equation

$$R = (1 + e^{\beta\tilde{\epsilon}})^{-1}. \tag{2.16}$$

From (2.16) and (2.10) we have

$$x_{ij} = T_{ij} - \mu\delta_{ij} + \sum_{mn}[in|V|jm]\rho_{mn}, \qquad y_{in} = \sum_{jm}\langle jin|V|jm\rangle\lambda_{jm}, \tag{2.17}$$

which are just the SCF conditions. Here R can be obtained from (2.10) by diagonalizing Z by Bogoliubov transformation U as follows. Now Z of the type (2.3) can be diagonalized by an unitary matrix U,

$$U^{+}ZU = \begin{pmatrix} E & 0 \\ 0 & -E \end{pmatrix}, \tag{2.18}$$

or

$$\begin{pmatrix} x & y \\ -y^* & -x^* \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} A \\ B \end{pmatrix}E, \tag{2.19}$$

where

$$U = \begin{pmatrix} A & B^* \\ B & A^* \end{pmatrix}, \tag{2.20}$$

$$U^{+}U = UU^{+} = 1, \tag{2.21}$$

$$A^{+}A + B^{+}B = AA^{+} + B^{*}B^{t} = 1, \qquad A^{t}B + B^{t}A = A^{*}B^{+} + B^{*}A^{t} = 0, \tag{2.22}$$

and

$$E = \begin{pmatrix} \epsilon_1 & & \\ & \ddots & \\ & & \epsilon_N \end{pmatrix}. \tag{2.23}$$

By this transformation R becomes

$$U^{+}RU = \begin{pmatrix} f_1 & & & & & \\ & \ddots & & & & \\ & & f_N & & & \\ & & & 1-f_1 & & \\ & & & & \ddots & \\ & & & & & 1-f_N \end{pmatrix}, \tag{2.24}$$

where

$$f_i = (1 + e^{\beta\epsilon_i})^{-1}. \tag{2.25}$$

Then R is obtained by

$$\begin{aligned} R &= U\begin{pmatrix} f & 0 \\ 0 & 1-f \end{pmatrix}U^{+} \\ &= \begin{pmatrix} A & B^* \\ B & A^* \end{pmatrix}\begin{pmatrix} f & 0 \\ 0 & 1-f \end{pmatrix}\begin{pmatrix} A^{+} & B^{+} \\ B^{T} & A \end{pmatrix} \\ &= \begin{pmatrix} AfA^{+} + B^{*}(1-f)B^{t}, & AfB^{+} + B^{*}(1-f)A \\ BfA^{+} + A^{*}(1-f)B^{t}, & BfB^{+} + A^{*}(1-f)A^{t} \end{pmatrix}. \end{aligned} \tag{2.26}$$

Usually the HFB equations (2.10) and (2.17) are solved by an iterative procedure as follows. Begin with an educated guess for the HFB Hamiltonian $z = (x,y)$. Then diagonalize Z by U. Then we obtain $R = (\rho,\lambda)$ by (2.26). Inserting this (ρ,λ) into (2.17) we obtain a new set of (x,y)'s. This process is repeated until the same densities are obtained on successive iterations.

As in solid state physics, the Green function is used frequently, now we give the HFB equation in terms of the Green function. Define the finite temperature Heisenberg operator by[6]

$$A^{+}_{H(Z),i}(\tau) = e^{H(Z)\tau/\hbar}a_i^{+}e^{-H(Z)\tau/\hbar}, \qquad A_{H(Z),i}(\tau) = e^{H(Z)\tau/\hbar}a_i e^{-H(Z)\tau/\hbar}. \tag{2.27}$$

The single particle temperature Green functions are defined by

$$\begin{aligned} G^{H(Z)}_{ij}(\tau-\tau') &= -\langle TA_{H(Z),i}(\tau)A^{+}_{H(Z),j}(\tau')\rangle_Z, \\ F^{H(Z)}_{ij}(\tau-\tau') &= -\langle TA_{H(Z),i}(\tau)A_{H(Z),j}(\tau')\rangle_Z, \\ \tilde{F}^{H(Z)}_{ij}(\tau-\tau') &= -\langle TA^{+}_{H(Z),i}(\tau)A^{+}_{H(Z),j}(\tau')\rangle_Z, \\ \tilde{G}^{H(Z)}_{ij}(\tau-\tau') &= -\langle TA^{+}_{H(Z),i}(\tau)A_{H(Z),j}(\tau')\rangle_Z. \end{aligned} \tag{2.28}$$

In terms of the matrix of the Green function, the HFB Green function $\mathscr{G}$ is defined by

$$\mathscr{G}^{H(Z)}(\tau-\tau') = \begin{pmatrix} \mathbf{G}^{H(Z)}(\tau-\tau'), & \mathbf{F}^{H(Z)}(\tau-\tau') \\ \tilde{\mathbf{F}}^{H(Z)}(\tau-\tau'), & \tilde{\mathbf{G}}^{H(Z)}(\tau-\tau') \end{pmatrix}, \tag{2.29}$$

where $\mathbf{G}^{H(Z)}$, etc., are the matrices of the Green function whose ijth elements are given by (2.28). The matrix Green functions in (2.29) satisfy the following relations:

$$\{\mathbf{G}^{H(Z)}(\tau-\tau')\}^{+}=\mathbf{G}^{H(Z)}(\tau-\tau'),$$
$$\{\mathbf{F}^{H(Z)}(\tau-\tau')\}^{t}=-\mathbf{F}^{H(Z)}(\tau'-\tau),$$
$$\widetilde{\mathbf{G}}^{H(Z)}(\tau-\tau')=-\{\mathbf{G}^{H(Z)}(\tau'-\tau)\}^{*},\tag{2.30}$$
$$\widetilde{\mathbf{F}}^{H(Z)}(\tau-\tau')=-\{\mathbf{F}^{H(Z)}(\tau'-\tau)\}^{*}.$$

The Fourier transforms of $\mathscr{G}$ are defined by

$$\mathscr{G}^{H(Z)}(\omega_n)=\begin{pmatrix}\mathbf{G}^{H(Z)}(\omega_n), & \mathbf{F}^{H(Z)}(\omega_n)\\ \widetilde{\mathbf{F}}^{H(Z)}(\omega_n), & \widetilde{\mathbf{G}}^{H(Z)}(\omega_n)\end{pmatrix}$$
$$=\int_0^{\beta\hbar}d\tau\, e^{i\omega_n\tau}\mathscr{G}^{H(Z)}(\tau),\tag{2.31}$$

where

$$\omega_n=(2n+1)\pi/\beta\hbar,\tag{2.32}$$

and the $\mathbf{G}^{H(Z)}(\omega_n)$, etc., satisfy the following relations:

$$\{\mathbf{G}^{H(Z)}(\omega_n)\}^{+}=\mathbf{G}^{H(Z)}(-\omega_n),$$
$$\{\mathbf{F}^{H(Z)}(\omega_n)\}^{t}=-\mathbf{F}^{H(Z)}(-\omega_n),$$
$$\widetilde{\mathbf{G}}^{H(Z)}(\omega_n)=-\{\mathbf{G}^{H(Z)}(\omega_n)\}^{*},\tag{2.33}$$
$$\widetilde{\mathbf{F}}^{H(Z)}(\omega_n)=-\{\mathbf{F}^{H(Z)}(\omega_n)\}^{*}.$$

The HFB equation in terms of the Green function is written by

$$\frac{d}{d\tau}\mathscr{G}^{H(Z)}(\tau)=-\hbar\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\delta(\tau)-\tilde{\epsilon}\,\mathscr{G}^{H(Z)}(\tau).\tag{2.34}$$

In terms of the Fourier transform,

$$\mathscr{G}^{H(Z)}(\omega_n)=\left\{i\hbar\omega_n\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}-\tilde{\epsilon}\right\}^{-1}\begin{Bmatrix}\hbar 1 & 0\\ 0 & \hbar 1\end{Bmatrix},\tag{2.35}$$

ϵ_{ij} and Δ_{jm} in $\tilde{\epsilon}$ can be written in terms of Green functions

$$\epsilon_{ij}=T_{ij}-\mu\delta_{ij}+\frac{1}{\beta}\sum_n \mathrm{tr}(\mathbf{\Gamma}^{ij}\mathbf{G}^{H(Z)}(\omega_n))$$
$$\tag{2.36}$$
$$\Delta_{jm}=\frac{1}{\beta}\sum_n \mathrm{tr}(\mathbf{\Lambda}^{jm}\mathbf{F}^{H(Z)}(\omega_n)),$$

where $\mathbf{\Gamma}^{ij}$ and $\mathbf{\Lambda}^{jm}$ are matrices such that

$$(\mathbf{\Gamma}^{ij})_{nm}=[in|V|jm]$$
$$\tag{2.37}$$
$$(\mathbf{\Lambda}^{jm})_{ik}=\langle jm|V|ik\rangle.$$

III. THE SYMMETRY GROUP ACTION ON THE HFB HAMILTONIAN, THE HFB DENSITY MATRIX, AND THE HFB GREEN FUNCTION

Let $\hat{g}$ be an unitary or antiunitary canonical transformation. For a complex number f,

$$\hat{g}(fa_i^{+})\hat{g}^{-1}=f^{(*)}\sum_{i'}a_{i'}^{+}g_{i'i},$$
$$\tag{3.1}$$
$$\hat{g}(fa_i)\hat{g}^{-1}=f^{(*)}\sum_{i'}a_{i'}g_{i'i}^{*},$$

where (*) denotes the complex conjugate in the case of antiunitary $\hat{g}$ and $\{g_{i'i}\}=g$ is a unitary matrix corresponding to $\hat{g}$. We use the same notation g for $\hat{g}$ or g in the following.

The symmetry group G_0 of the system is defined by

$$G_0=\{g|\hat{g}\mathscr{H}\hat{g}^{-1}=\mathscr{H}\}.\tag{3.2}$$

Then $g\in G_0$ means that

$$\sum_{i'j'}g_{ii'}T_{i'j'}^{(*)}(g^{+})_{j'j}=T_{ij},$$

and (3.3)

$$\sum_{i'j'k'l'}g_{ii'}g_{jj'}\langle i'j'|V|k'l'\rangle^{(*)}(g^{+})_{k'k}(g^{+})_{l'l}$$
$$=\langle ij|V|kl\rangle.$$

Here we give two examples of the symmetry group.

Example 1: Electron gas or liquid ^{3}He in n-dimensional space. In this case

$$G_0=E^n\times S\times\Phi+t\{E^n\times S\times\Phi\}=(E^n\times S)\times M,\tag{3.4}$$

where E^n is the Euclidean group in n-dimensional space, S is the group of the spin rotation,[3,4] t is the time reversal, Φ is the group of phase transformations such as, for $\phi\in\Phi$,

$$\phi\cdot a_j^{+}=e^{i\phi}a_j^{+},\quad \phi\cdot a_j=e^{-i\phi}a_j,\tag{3.5}$$

and

$$M=\Phi+t\Phi.\tag{3.6}$$

Example 2: Electron in solid with a lattice group. In this case

$$G_0=P(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)\times S\times M,\tag{3.7}$$

where $P(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)$ is the triclinic space group with the basis vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$.

In the case of the molecular system with a spatial point symmetry we have considered the UHF equation in the previous paper.[4] In subsequent papers[5] we will study in detail the case of the above examples.

The G_0 action on the HFB Hamiltonian $H(Z)$ is defined by

$$g\cdot H(Z)\equiv\hat{g}H(Z)\hat{g}^{-1}=H(Z_g),\quad \text{for } g\in G_0,\tag{3.8}$$

where

$$Z_g=\begin{pmatrix}x_g & y_g\\ -y_g^{*} & -x_g^{*}\end{pmatrix}=\tilde{g}Z^{(*)}\tilde{g}^{\dagger},\tag{3.9}$$

$$\tilde{g}=\begin{pmatrix}g & 0\\ 0 & g^{*}\end{pmatrix},\tag{3.10}$$

$$x_g=g\,x^{(*)}g^{\dagger},\quad y_g=g\,y^{(*)}g^{t}.\tag{3.11}$$

The G_0 action on the HFB density matrix $R(Z)$ is defined by

$$g\cdot R(Z)\equiv\langle AA^{+}\rangle_{Z_g}=\begin{pmatrix}\rho_g & \lambda_g\\ -\lambda_g^{*} & 1-\rho_g^{*}\end{pmatrix}=\tilde{g}R(Z)^{(*)}\tilde{g}^{\dagger}.\tag{3.12}$$

Then

$$\rho_g=g\rho^{(*)}g^{+},\quad \lambda_g=g\lambda^{(*)}g^{t}.\tag{3.13}$$

The G_0 action on the Green function is defined by

$$g\cdot G_{ij}^{H(Z)}(\tau-\tau')=-\langle TA_{H(Z_g),i}(\tau)A_{H(Z_g),j}(\tau')\rangle_{Z_g}\tag{3.14}$$

and so on. Then we have

$$g\cdot\mathscr{G}^{H(Z)}(\tau)=\tilde{g}\,\mathscr{G}\,(\tau)\tilde{g}^{\dagger},\tag{3.15}$$

that is,

$$g\cdot\mathbf{G}^{H(Z)}(\tau)=g\mathbf{G}^{H(Z)}(\tau)g^{+},\quad g\cdot\mathbf{F}^{H(Z)}(\tau)=g\mathbf{F}^{H(Z)}(\tau)g^{t},$$
$$g\cdot\widetilde{\mathbf{F}}^{H(Z)}(\tau)=g^{*}\widetilde{\mathbf{F}}^{H(Z)}(\tau)g^{+},\quad g\cdot\widetilde{\mathbf{G}}^{H(Z)}(\tau)=g^{*}\widetilde{\mathbf{G}}^{H(Z)}(\tau)g^{t}. \tag{3.16}$$

The G_0 action on the Fourier transform of the Green function is given by

$$g\cdot\mathscr{G}^{H(Z)}(\omega_n)=\begin{cases}\tilde{g}\mathscr{G}^{H(Z)}(\omega_n)\tilde{g}^{\dagger}, & \text{for unitary } g,\\ \tilde{g}\mathscr{G}^{H(Z)*}(-\omega_n)\tilde{g}^{\dagger}, & \text{for antiunitary } g.\end{cases} \tag{3.17}$$

IV. DECOMPOSITION OF THE SCF CONDITION INTO IRREDUCIBLE COMPONENTS

The set of $N\times N$ Hermitian matrices x(or ρ) is a $(\tau_1=N^2)$-dimensional vector space W_x (or W_ρ) over the real number field R with the bases, $E_{\alpha\alpha}$, $E_{\alpha\beta}+E_{\beta\alpha}$, and $i(E_{\alpha\beta}-E_{\beta\alpha})$, where $1\leqslant\alpha,\beta\leqslant N$ and $E_{\alpha\beta}$ is an $N\times N$ matrix whose $\alpha\beta$th entry is 1 and the others are zero. The set of $N\times N$ antisymmetric matrices y (or λ) is a $(\tau_2=N^2-N)$-dimensional vector space W_y (or W_λ) over R with the bases $E_{\alpha\beta}-E_{\beta\alpha}$ and $i(E_{\alpha\beta}-E_{\beta\alpha})$. From (3.11) W_x and W_y become representation spaces of G_0 over R. We decompose W_x and W_y into R-irreducible (R-rep) spaces of G_0, then for $x\in W_x$ and $y\in W_y$ we have

$$x=\sum_{\gamma}\sum_{p=1}^{a_\gamma}\sum_{\iota=1}^{n_\gamma}x_{p\iota}^{\gamma}e_{p\iota}^{\gamma},$$
$$y=\sum_{\gamma}\sum_{p=1}^{b_\gamma}\sum_{\iota=1}^{n_\gamma}y_{p\iota}^{\gamma}f_{p\iota}^{\gamma}, \tag{4.1}$$

where $x_{p\iota}^{\gamma}$ and $y_{p\iota}^{\gamma}$ are real numbers, γ denotes an irreducible representation D^{γ} of G_0 over R, a_γ and b_γ are the multiplicity of D^{γ} in W_x and W_y, n_γ is the dimension of D^{γ}, $e_{p\iota}^{\gamma}$ $(\iota=1,\ldots,n_\gamma)$ form the pth basis of D^{γ} in W_x and $f_{p\iota}^{\gamma}$ $(\iota=1,\ldots,n_\gamma)$ form the pth basis of D^{γ} in W_y:

$$ge_{p\iota}^{\gamma(*)}g^{+}=\sum_{\kappa\iota}D_{\kappa\iota}^{\gamma}(g)e_{p\kappa}^{\gamma},$$
$$gf_{p\iota}^{\gamma(*)}g^{t}=\sum_{\kappa}D_{\kappa\iota}^{\gamma}(g)f_{p\kappa}^{\gamma}. \tag{4.2}$$

In the same way we can decompose W_ρ and W_λ corresponding to the HFB density matrix R into R-irreducible vector spaces; for $\rho\in W_\rho$ and $\lambda\in W_\lambda$ we have

$$\rho=\sum_{\gamma}\sum_{p=1}^{a_\gamma}\sum_{\iota=1}^{n_\gamma}\rho_{p\iota}^{\gamma}e_{p\iota}^{\gamma},$$
$$\lambda=\sum_{\gamma}\sum_{p=1}^{a_\gamma}\sum_{\iota=1}^{n_\gamma}\lambda_{p\iota}^{\gamma}f_{p\iota}^{\gamma}, \tag{4.3}$$

where $\rho_{p\iota}^{\gamma}$ and $\lambda_{p\iota}^{\gamma}$ are real numbers.

Now we consider group theoretical properties of the SCF condition (2.17).

For this purpose we have a proposition.

Proposition 4.1: Let ρ be a Hermitian matrix and λ be an antisymmetric matrix. Define matrices $x[\rho]$ and $y[\lambda]$ by

$$\{x[\rho]\}_{ij}\equiv\sum_{mn}[in|V|jm]\rho_{mn},$$
$$\{y[\rho]\}_{ij}=\sum_{mn}\langle ij|V|nm\rangle\lambda_{nm}. \tag{4.4}$$

Then for $g\in G_0$

$$x[g\rho^{(*)}g^{+}]=g\{x[\rho]\}^{(*)}g^{+},$$
$$y[g\lambda^{(*)}g^{t}]=g\{y[\lambda]\}^{(*)}g^{t}. \tag{4.5}$$

The proof of the Proposition is given in the Appendix.

Now we consider $e_{p\iota}^{\gamma}$ and $f_{p\iota}^{\gamma}$ as elements of W_ρ and W_λ. Define $\epsilon_{p\iota}^{\gamma}$ and $\Delta_{p\iota}^{\gamma}$ by

$$\epsilon_{p\iota}^{\gamma}\equiv x[e_{p\iota}^{\gamma}],\quad \Delta_{p\iota}^{\gamma}\equiv y[f_{p\iota}^{\gamma}]. \tag{4.6}$$

Then we have a theorem.

Theorem 4.2: $\epsilon_{p\iota}^{\gamma}$ $(\iota=1,\ldots,n)$ and $\Delta_{p\iota}^{\gamma}$ form the bases of D^{γ} in W_x and W_y:

$$g\epsilon_{p\iota}^{\gamma(*)}g^{\dagger}=\sum_{\kappa}D_{\kappa\iota}^{\gamma}(g)\epsilon_{p\kappa}^{\gamma},$$
$$g\Delta_{p\iota}^{\gamma(*)}g^{t}=\sum_{\kappa}D_{\kappa\iota}^{\gamma}(g)\Delta_{p\kappa}^{\gamma}. \tag{4.7}$$

Proof of the Theorem: From (4.2), (4.5), and (4.6),

$$g\epsilon_{p\iota}^{\gamma(*)}g^{\dagger}=x[ge_{p\iota}^{\gamma(*)}g^{\dagger}]=x\Big[\sum_{\kappa}D_{\kappa\iota}^{\gamma}(g)e_{p\kappa}^{\gamma}\Big]$$
$$=\sum_{\kappa}D_{\kappa\iota}^{\gamma}(g)x[e_{p\kappa}^{\gamma}]=\sum_{\kappa}D_{\kappa\iota}^{\gamma}(g)\epsilon_{p\kappa}^{\gamma}.$$

In the same way we have (4.7) for $\Delta_{p\iota}^{\gamma}$.

Thus in the SCF condition the decomposition of W_λ and W_ρ is propagated into the decomposition of W_x and W_y. Then $\epsilon_{p\iota}^{\gamma}$ and $\Delta_{p\iota}^{\gamma}$ can be written as follows:

$$\epsilon_{p\iota}^{\gamma}=\sum_{p'\iota'}X^{\gamma}(p'\iota'|p\,\iota)e_{p'\iota'}^{\gamma},$$
$$\Delta_{p\iota}^{\gamma}=\sum_{p'\iota'}Y^{\gamma}(p'\iota'|p\,\iota)f_{p'\iota'}, \tag{4.8}$$

where $X^{\gamma}(p'\,\iota'|p\,\iota)$ and $Y^{\gamma}(p'\,\iota'|p\,\iota)$ are real numbers.

V. GROUP THEORETICAL CLASSIFICATION OF THE HFB EQUATION AND SOLUTION

The HFB free energy $F(R)=F(\rho,\lambda)$ contains external parameters, temperature T, pressure p, and others. We write these parameters by Π for brevity's sake and then the HFB free energy is written as $F(\Pi,\rho,\lambda)$. From the G_0 invariance (3.2) of the Hamiltonian we can easily obtain the G_0 invariance of the HFB free energy $F(\Pi,\rho,\lambda)$, that is,

$$F(\Pi,\rho_g,\lambda_g)=F(\Pi,\rho,\lambda),\quad \text{for } g\in G_0. \tag{5.1}$$

Thus the problem is to determine the extreme point of the G_0 invariant function $F(\Pi,\rho,\lambda)$ of (ρ,λ) with parameters Π. Mathematically this is just the same problem as the UHF theory (generally it is called[7] the Landau problem). By similar consideration with the previous paper,[3,4] we can develop the group theoretical classification of the HFB equation and solution and the group theoretical bifurcation theory of the HFB equation. Here we give only results on the group theoretical analysis on the HFB equation. Their derivations are similar with those of the previous paper and we do not give them here.

Definition 5.1: Let R be a HFB solution. The invariance group $G(R)\subset G_0$ is defined by

$$G(R) = \{g \in G_0 | g \cdot R = R\}. \tag{5.2}$$

If R is a HFB solution, for any $g \in G_0$, $g \cdot R$ is also a solution with the same HFB free energy. The conjugate subgroup $gG(R)g^{-1}$ is the invariance group of $g \cdot R$, that is,

$$G(g \cdot R) = g \cdot G(R) \cdot g^{-1}. \tag{5.3}$$

As the relation of the conjugation of subgroups is an equivalence relation (we denote it by $\sim$), we can decompose the set $S(G_0)$ of all the subgroups of G_0 into equivalence classes according to this equivalence relation, and we obtain a quotient set $Q(G_0) = S(G_0)/\sim$. The quotient set $Q(G_0)$ represents physically distinct types of HFB solutions. Thus we can classify the HFB solutions by $Q(G_0)$.

From the SCF condition (2.17) and Proposition 4.1 it follows for $Z = (x,y)$ corresponding to a solution $R = (\rho, \lambda)$,

$$x_g = x, \quad y_g = y, \quad \text{for } g \in G(R), \tag{5.4}$$

or

$$g \cdot Z = Z. \tag{5.5}$$

From the definition of G_0 action on the Green function we have for $g \in G(R)$

$$g \cdot \mathscr{G}(\tau) = \mathscr{G}(\tau), \quad g \cdot \mathscr{G}(\omega_n) = \mathscr{G}(\omega_n). \tag{5.6}$$

So far we classified the HFB solution. Now we consider the classification of the HFB equation. Let $G \subset G_0$ be a subgroup of G_0 and $W_\rho(G)$ and $W_\lambda(G)$ be vector subspaces of W_ρ and W_λ such that

$$\begin{aligned} W_\rho(G) &= \{\rho \in W_\rho | \rho_g = \rho, \quad \text{for } g \in G\}, \\ W_\lambda(G) &= \{\lambda \in W_\lambda | \lambda_g = \lambda, \ \text{for } g \in G\}. \end{aligned} \tag{5.7}$$

In the same way we define $W_x(G)$ and $W_y(G)$. Thus we define the G–HFB equation with an invariance group $G \subset G_0$ as the variational problem of the HFB free energy $F(\Pi, \rho, \lambda)$ in the restricted subspace $W_\rho(G) \oplus W_\lambda(G)$. Solving the G–HFB equation in the iterative procedure is begun with a guessed $R_0 = (\rho_0, \lambda_0)$ in $W_\rho(G) \oplus W_\lambda(G)$. From (4.5) we obtain $Z_0 = (x_0, y_0)$ in $W_x(G) \oplus W_y(G)$. From (2.10) we obtain the next $R_1 = (\rho_1, \lambda_1)$, which is in $W_\rho(G) \oplus W_\lambda(G)$.

Thus in the process of iteration, the space of the variation is conserved. Then we can classify the HFB equation by $Q(G_0)$.

VI. GROUP THEORETICAL CLASSIFICATION OF THE INSTABILITY OF THE HFB SOLUTION

Since the HFB free energy $F(\Pi, \rho, \lambda)$ is G_0 invariant, W_ρ and W_λ are the representation spaces of G_0, we can apply Sattinger's[8] group theoretical bifurcation theory to the HFB equation as in the previous paper.[4(c),4(d)] Now we give some notation and definition. We shall represent the phrase "irreducible single valued representation over R" by the abbreviation "R-rep" and we denote an R-rep of a group $G \subset G_0$ by $\check{G}$. We denote the representation space of $\check{G}$ in $W_\rho \oplus W_\lambda$ by $V(\check{G})$. Let G' be a subgroup of G such that $\operatorname{Ker}(\check{G}) \subseteq G' \subset G$ and for some $R'(\neq 0) \in V(\check{G})$

$$g \cdot R' = R', \quad \text{for all } g \in G'. \tag{6.1}$$

We call such G' the invariance group (IG) of $\check{G}$ and R' the invariance vector (IV) of G'. Then we get the following Proposition by the similar argument with the paper.[4(c)]

Proposition 6.1: The instability of a HFB solution R is characterized by an R-rep $\check{G}(R)$ of the invariance group $G(R)$ of R. The invariance group of the solution bifurcating from the instability is the one of the invariance groups of $\check{G}(R)$.

There exists a criterion [Criterion 6.1 of Ref. 4(c)] to determine whether a subgroup $G' \subseteq G$ can be an invariance group of $\check{G}$ or not, but the IG of $\check{G}$ cannot be obtained in general form. This problem will be discussed individually in special cases in the subsequent paper.[5]

Here we shall give the general expression of the instability of a HFB solution and its decomposition into R-irreducible components. The instability of a HFB solution $R = (\rho, \lambda)$ is given by the second-order variation $\delta^2 F(\Pi, \rho, \lambda)$. Using the same method with Mermin's one,[9] the second-order variation becomes

$$\begin{aligned} 4\delta^2 F &= \operatorname{tr}\{\delta R \delta \tilde{\epsilon}\} + (1/\beta)\operatorname{tr}[\delta R \delta \ln\{R/(1-R)\}] \\ &= \sum_{k,l,k'l'}^{N} \{\delta\rho_{kl}[ll'|V|kk']\delta\rho_{k'l'} + \text{c.c.} \\ &\quad - \delta\lambda_{kl}\langle lk|V|k'l'\rangle^* \delta\lambda^*_{k'l'} - \text{c.c.}\} \\ &\quad + \sum_{i,j=1}^{2N} (\delta R')_{ji} \frac{E_j - E_i}{F_i - F_j} (\delta R')_{ij}, \end{aligned} \tag{6.2}$$

where $\{F_i\}$ and $\{E_i\}$ are elements of the diagonal matrices

$$\begin{aligned} U^+ R U &= \begin{pmatrix} F_1 & & & 0 \\ & F_N & & \\ & & F_{N+1} & \\ 0 & & & F_{2N} \end{pmatrix} \\ &= \begin{pmatrix} f_1 & & & & & 0 \\ & \ddots & & & & \\ & & f_N & & & \\ & & & 1-f_1 & & \\ & & & & \ddots & \\ 0 & & & & & 1-f_N \end{pmatrix}, \end{aligned} \tag{6.3}$$

$$\begin{aligned} U^+ \tilde{\epsilon} U &= \begin{pmatrix} E_1 & & & & & 0 \\ & \ddots & & & & \\ & & E_N & & & \\ & & & E_{N+1} & & \\ & & & & \ddots & \\ 0 & & & & & E_{2N} \end{pmatrix} \\ &= \begin{pmatrix} \epsilon_1 & & & & & 0 \\ & \ddots & & & & \\ & & \epsilon_N & & & \\ & & & -\epsilon_1 & & \\ & & & & \ddots & \\ 0 & & & & & -\epsilon_N \end{pmatrix}, \end{aligned}$$

$$\delta R' = U^+ \delta R U, \tag{6.4}$$

$$\delta R = \begin{pmatrix} \delta\rho & \delta\lambda \\ -\delta\lambda^* & -\delta\rho^* \end{pmatrix}, \tag{6.5}$$

$$U = \begin{pmatrix} A & B^* \\ B & A^* \end{pmatrix}, \tag{6.6}$$

and U diagonalizes R. Expressing $\delta R'$ in (6.2) by $(\delta\rho,\delta\lambda)$ we get

$$4\delta^2 F = \widetilde{\Lambda}^t \Omega\, \Lambda, \tag{6.7}$$

where Ω and Λ are the matrices represented by

$$\Omega = \begin{pmatrix} Q_{\rho\rho} & Q_{\rho\rho^*} & Q_{\rho\lambda} & Q_{\rho\lambda^*} \\ Q^*_{\rho\rho^*} & Q^*_{\rho\rho} & Q^*_{\rho\lambda^*} & Q^*_{\rho\lambda} \\ Q^t_{\rho\lambda} & Q^+_{\rho\lambda^*} & Q_{\lambda\lambda} & Q_{\lambda\lambda^*} \\ Q^t_{\rho\lambda^*} & Q^+_{\rho\lambda} & Q^*_{\lambda\lambda^*} & Q^*_{\lambda\lambda} \end{pmatrix}, \tag{6.8}$$

$$\widetilde{\Lambda} = \begin{pmatrix} \delta\rho \\ \delta\rho^* \\ \delta\lambda \\ \delta\lambda^* \end{pmatrix}, \tag{6.9}$$

$$\delta\rho = (\delta\rho_{kl}), \quad \delta\lambda = (\delta\lambda_{kl}) \tag{6.10}$$

and $Q_{\rho\rho}$, etc., are the matrices whose $(kl-k'l')$th elements are given as follows:

$$\begin{aligned}
Q_{\rho\rho}(kl,k'l') &= [ll'|V|kk'] + \sum_{i,j=1}^{N} \frac{\epsilon_j-\epsilon_i}{f_i-f_j}\{A^+_{jk}A_{li}A^+_{ik'}A_{l'j} + B^t_{jk}B^*_{li}B^t_{ik'}B^*_{l'j}\} \\
&\quad + \sum_{i,j=1}^{N} \frac{\epsilon_i+\epsilon_j}{1-f_i-f_j}\{A^+_{jk}B^*_{li}B^t_{jk'}A_{l'j} + B^t_{jk}A_{li}A^+_{ik'}B^*_{l'j}\}, \\
Q_{\rho\rho^*}(kl,k'l') &= -\sum_{i,j=1}^{N} \frac{\epsilon_j-\epsilon_i}{f_i-f_j}\{A^+_{jk}A_{li}B^+_{ik'}B_{l'j} + B^t_{jk}B^*_{li}A^t_{ik'}A^*_{l'j}\} \\
&\quad - \sum_{i,j=1} \frac{\epsilon_i+\epsilon_j}{1-f_i-f_j}\{A^+_{jk}B^*_{li}A^t_{ik'}B_{l'j} + B^t_{jk}A_{li}B^*_{ik'}A^*_{l'j}\}, \\
Q_{\rho\lambda}(kl,k'l') &= \sum_{i,j=1}^{N} \frac{\epsilon_j-\epsilon_i}{f_i-f_j}\{A^+_{jk}A_{li}A^+_{ik'}B_{l'j} + B^t_{jk}B^*_{li}B^t_{ik'}A^*_{l'j}\} + \sum_{i,j=1}^{N} \frac{\epsilon_i+\epsilon_j}{1-f_i-f_j}\{A^+_{jk}B^*_{li}B^t_{ik'}B_{l'j} + B^t_{jk}A_{li}A^+_{ik'}A^*_{l'j}\}, \\
Q_{\rho\lambda^*}(kl,k'l') &= -Q_{\rho\lambda}(lk,l'k')^*, \\
Q_{\lambda\lambda}(kl,k'l') &= \sum_{i,j=1}^{N} \frac{\epsilon_j-\epsilon_i}{f_i-f_j}\{A^+_{jk}B_{li}A^+_{ik'}B_{l'j} + B^t_{jk}A^*_{li}B^t_{ik'}A^*_{l'j}\} \\
&\quad + \sum_{i,j=1}^{N} \frac{\epsilon_i+\epsilon_j}{1-f_i-f_j}\{A^+_{jk}A^*_{li}B^t_{ik'}B_{l'j} + B^t_{jk}B_{li}A^+_{ik'}A^*_{l'j}\}, \\
Q_{\lambda\lambda^*}(kl,k'l') &= -\langle lk|V|k'l'\rangle^* - \sum_{i,j=1}^{N} \frac{\epsilon_j-\epsilon_i}{f_i-f_j}\{A^+_{jk}B_{li}B^+_{ik'}A_{l'j} + B^t_{jk}A^*_{li}A^t_{ik'}B^*_{l'j}\} \\
&\quad - \sum_{i,j=1}^{N} \frac{\epsilon_i+\epsilon_j}{1-f_i-f_j}\{A^+_{jk}A^*_{li}A^t_{ik'}A_{l'j} + B^t_{jk}B_{li}B^+_{ik'}B^*_{l'j}\}.
\end{aligned} \tag{6.11}$$

They satisfy the following relations:

$$\begin{aligned}
&Q^t_{\rho\rho} = Q_{\rho\rho}, \quad Q_{\rho\rho}(lk,l'k') = Q_{\rho\rho}(kl,k'l')^*, \\
&Q^t_{\lambda\lambda} = Q_{\lambda\lambda}, \quad Q_{\lambda\lambda}(lk,l'k') = Q_{\lambda\lambda}(kl,k'l'), \\
&Q^+_{\lambda\lambda^*} = Q_{\lambda\lambda^*}, \quad Q^+_{\rho\rho^*} = Q_{\rho\rho^*}.
\end{aligned} \tag{6.12}$$

Now let us decompose $4\delta^2 F$ into R-irreducible components of the group $G(R)$. Then, $\delta\rho$ and $\delta\lambda$ can be decomposed into R-irreducible components of $G(R)$ as follows:

$$\begin{aligned}
\delta\rho &= \sum_\mu\sum_m\sum_\alpha \delta\rho^\mu_{m\alpha} e^\mu_{m\alpha}, \\
\delta\lambda &= \sum_\mu\sum_n\sum_\alpha \delta\lambda^\mu_{n\beta} f^\mu_{n\beta},
\end{aligned} \tag{6.13}$$

where $\delta\rho^\mu_{m\alpha}$ and $\delta\rho^\mu_{n\beta}$ are real numbers, μ denotes an R-rep of $G(R)$, and $e^\mu_{m\alpha}$ and $f^\mu_{n\beta}$ form the mth and nth bases of the R-rep D^μ of $G(R)$ in W_ρ and W_λ. Inserting (6.13) into (6.7) and using the orthogonality conditions of the R-rep $G(R)$, we get

$$4\delta^2 F = \sum_\mu 4\delta^2 F^\mu, \tag{6.14}$$

where

$$4\delta^2 F^\mu = \widetilde{\Lambda}^t_\mu \Omega_\mu \widetilde{\Lambda}_\mu, \tag{6.15}$$

$$\Omega_\mu = \begin{pmatrix} Q^\mu_{\rho\rho} & Q^\mu_{\rho\lambda} \\ (Q^\mu_{\rho\lambda})^t & Q^\mu_{\lambda\lambda} \end{pmatrix}, \tag{6.16}$$

$$\widetilde{\Lambda}_\mu = \begin{pmatrix} \delta\rho^\mu_{m\alpha} \\ \delta\lambda^\mu_{n\beta} \end{pmatrix}, \tag{6.17}$$

and the matrices $Q^\mu_{\rho\rho}$, etc., are given by

$$\begin{aligned}
Q^\mu_{\rho\rho}(m\alpha,m'\alpha') &= (e^\mu_{m\alpha}|Q_{\rho\rho}|e^\mu_{m'\alpha'}) + \text{c.c.} \\
&\quad + (e^\mu_{m\alpha}|Q_{\rho\rho^*}|e^{\mu*}_{m'\alpha'}) + \text{c.c.}, \\
Q^\mu_{\rho\lambda}(m\alpha,n\beta) &= (e^\mu_{m\alpha}|Q_{\rho\lambda}|f^\mu_{n\beta}) + \text{c.c.} \\
&\quad + (e^\mu_{m\alpha}|Q_{\rho\lambda^*}|f^{\mu*}_{n\beta}) + \text{c.c.}, \\
Q^\mu_{\lambda\lambda}(n\beta,n'B') &= (f^\mu_{n\beta}|Q_{\lambda\lambda}|f^\mu_{n'\beta'}) + \text{c.c.} \\
&\quad + (f_{n\beta}|Q_{\lambda\lambda^*}|f^*_{n'\beta'}) + \text{c.c.},
\end{aligned} \tag{6.18}$$

and satisfy the relations

$$(Q^\mu_{\rho\rho})^t = Q^\mu_{\rho\rho}, \quad (Q^\mu_{\lambda\lambda})^t = Q^\mu_{\lambda\lambda}. \tag{6.19}$$

As in the case of the paper,[4(c)] if D^μ is absolutely irreducible, the $Q^\mu_{\rho\rho}$, etc., become diagonal with respect to α, β, etc., that is,

$$Q^{\mu}_{\rho\rho}(m\alpha,m'\alpha')=\delta_{\alpha\alpha'}Q^{\mu}_{\rho\rho}(m\alpha,m'\alpha)$$
(independent of α),
$$Q^{\mu}_{\rho\lambda}(m\alpha,n\beta)=\delta_{\alpha\beta}Q^{\mu}_{\rho\lambda}(m\alpha,n\alpha) \tag{6.20}$$
(independent of α),
$$Q^{\mu}_{\lambda\lambda}(n\beta,n'\beta')=\delta_{\beta\beta'}Q^{\mu}_{\lambda\lambda}(n\beta,n\beta)$$
(independent of β).

Thus we have obtained the R-irreducible instability matrix Ω_μ characterized by an R-rep D^μ of $G(R)$.

ACKNOWLEDGMENTS

The author would like to express his sincere thanks to Professor H. Fukutome and Dr. Machida for helpful advice and discussion.

APPENDIX: PROOF OF PROPOSITION 4.1

In (3.3), replacing (j,j',k',k,l',l) by (n',n'',j',j,m'',m') we get

$$\sum_{\substack{i'j'\\m''n''}} g_{ii'}\, g_{n'n''}\, \langle i'n''|V|j'm''\rangle^{(*)} g^{+}_{j'j}\, g^{+}_{m''m'} = \langle in'|V|jm'\rangle. \tag{A1}$$

Multiplying $g^{+}_{nn'}g_{m'm}$ on both sides and summing by $n'm'$, we get

$$\sum_{i'j'} g_{ii'}\langle i'n|V|j'm\rangle^{(*)}g^{+}_{j'j} = \sum_{n'm'} g^{+}_{nn'}\langle in'|V|jm'\rangle g_{m'm}. \tag{A2}$$

$$\begin{aligned}
x[g\rho^{(*)}g^{+}]_{ij}
&= \sum_{m'n'}[in'|V|jm'](g\rho^{(*)}g^{+})_{m'n'}\\
&= \sum_{mn}\left\{\sum_{m'n'} g^{+}_{nn'}[in'|V|jm']g_{m'm}\right\}\rho^{(*)}_{mn}\\
&= \sum_{mn}\left\{\sum_{i'j'} g_{ii'}[i'n|V|j'm]^{(*)}g^{+}_{j'j}\right\}\rho^{(*)}_{mn} \qquad \text{[from (A2)]}\\
&= \sum_{i'j'}\left\{g_{ii'}\sum_{mn}[i'n|V|j'm]^{(*)}\rho^{(*)}_{mn}g^{+}_{j'j}\right\}\\
&= (gx[\rho]^{(*)}g^{+})_{ij}.
\end{aligned}$$

This completes the proof of the first equation of (4.5). The second equation of (4.5) is proved in a similar manner.

[1] M. Ishikawa and P. Fisher, Solid State Commun. **24**, 747 (1977); H. C. Hamaker, L. D. Woolf, H. B. Mackay, Z. Fisk, and M. B. Maple, *ibid.* **32**, 289 (1979); C. A. Balseiro and L. M. Falicov, in *Superconductivity in d and f-band Metals*, edited by H. Suhl and M. B. Maple (Academic, New York, 1980), pp. 105–112; E. Gratz and E. P. Wohlfarth, J. Mag. Magn. Mat. **15**, 903 (1980).

[2] K. Machida and T. Matsubara, J. Phys. Soc. Jpn. **50**, 2195 (1981); H. Nakanishi, M. Machida, and T. Matsubara, Solid State Commun. **43**, 899 (1982); E. W. Fenton and G. C. Psaltakis, *ibid.* **45**, 5 (1982).

[3] H. Fukutome, Prog. Theor. Phys. **52**, 115 (1974); **52**, 1766 (1974).

[4] (a)M. Ozaki and H. Fukutome, Prog. Theor. Phys. **60**, 1322 (1978); (b) M. Ozaki, *ibid.* **62**, 1183 (1979); (c) **67**, 83 (1982); (d) **67**, 415 (1982).

[5] M. Ozaki, J. Math. Phys. 26, 1521 (1985).

[6] A. L. Fetter and J. D. Waleska, *Quantum Theory of Many-Particle System* (McGraw–Hill, New York, 1971).

[7] L. Michel, Rev. Mod. Phys. **52**, 617 (1980).

[8] D. H. Sattinger, *Group Theoretical Method in Bifurcation Theory, Lecture Notes in Mathematics* (Springer, Berlin, 1979).

[9] N. D. Mernin, Ann. Phys. **21**, 99 (1963).

Group theoretical analysis of the Hartree–Fock–Bogoliubov equation. II. The case of the electronic system with triclinic lattice symmetry

Masa-aki Ozaki
Department of Physics, Kyoto University, Kyoto 606, Japan

(Received 27 September 1984; accepted for publication 1 February 1985)

This paper describes a group theoretical classification of the temperature–Hartree–Fock–Bogoliubov (HFB) equation in a crystalline solid system and the electronic state of the system. It is shown that the state with a single-order parameter (charge density wave, spin density wave, etc.) is classified into 47 classes and the BCS state coexistent with other nonsuperconducting orders, such as magnetic superconductors, is classified into 26 classes. The standard HFB Hamiltonian for each class is obtained. It is found that in each of above coexistence states (except BCS + ferromagnetism) an abnormal Cooper pair occurs.

I. INTRODUCTION

The electron system in a crystalline solid can be in many different states such as normal paramagnetic (NP), ferromagnetic (FM), spin density wave (SDW), charge density wave (CDW), BCS superconducting states. Each of these states can be described as the zeroth approximation by a symmetry-broken solution of a certain type of the temperature–Hartree–Fock–Bogoliubov (HFB) equation.

Recently new states with multiple-order parameters such as BCS states coexistent with magnetism have been suggested.[1] Many of these states have been studied individually by the HFB equation.[2–6] In the future other new states will probably be found.

Therefore, it is interesting and important to solve the following general problems.

(a) What kinds of phases (broken-symmetry HFB solutions) with single- or multiple-order parameters are group theoretically possible?

(b) What kinds of instabilities (the second-order phase transition) are group theoretically possible?

In Fukutome's[7] and our previous[8] papers, these problems have been solved for the molecular system (without[7] or with[8] a spatial point symmetry) in the frame of the zero-temperature Hartree–Fock (HF) equation.

In order to solve these problems in the case at finite temperature including superconductivity, in a previous paper[9] (referred to as I) we have developed the general group theoretical analysis of the HFB equation. In the paper we have shown that the group theoretical analysis of the HF equation can be generalized to the HFB equation.

In this paper we will apply the general theory to the electron system in crystal with triclinic lattice symmetry. By obtaining all R-reps (irreducible single-valued representation over the real number field) of the symmetry group G_0 of the system and the invariance group G_s of the ordinary BCS superconducting state, we have 24 types of instability for NP and 17 types for the ordinary BCS state.

For each R-rep of G_0 and G_s we have obtained all invariance groups and their corresponding HFB Hamiltonians. Thus we have obtained 47 classes of states with a single-order parameter and 26 classes of the BCS states coexistent with other nonsuperconducting orders.

In Sec. II we describe the symmetry group G_0 of the system and its elements in detail. In Sec. III the single-valued irreducible representations of G_0 over the real number field R (R-reps) are obtained. The bases of R-reps of G_0 in the HFB Hamiltonian space are obtained. In Sec. IV we give the classification of the states with single-order parameters and the standard HFB Hamiltonian and the self-consistent condition for each class. In Sec. V the BCS states coexistent with other nonsuperconducting orders are discussed and classified. The standard HFB Hamiltonian for each class is obtained.

The notations used in this paper are the same as those used in I and our previous papers.[7,8]

II. THE SYMMETRY GROUP OF THE SYSTEM

We consider the electron system of the crystal with a triclinic lattice symmetry. For simplicity we take a single band model which has the following Hamiltonian:

$$\mathscr{H} = \sum_{ks} \epsilon(k) a^+_{ks} a_{ks} + \frac{1}{2} \sum_{\substack{q,k,k' \\ s,s'}} V(k,k',q) a^+_{k+qs} a^+_{k'-qs'} a_{k's'} a_{ks}, \tag{2.1}$$

where (k,k') and (s,s') denote the momentum and the spin states $(\uparrow,\downarrow)$, respectively, and

$$V(k,k',q) = \langle (k+q)s,(k'-q)s' | V | ks, k's' \rangle \tag{2.2}$$

(independent of s and s')

in the notation of I. The Hamiltonian $\mathscr{H}$ has the following symmetry group G_0:

$$G_0 = \{ L(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3) + IL(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3) \} \times S \times \{ \Phi + t\Phi \} = P(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3) \times S \times M, \tag{2.3}$$

where $L(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)$ is the three-dimensional translational group with the basis vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$, I is the inversion, S is the group of the spin rotation, Φ is the group of phase transformation, t is the time reversal, $P(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3) = L(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3) + IL(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)$, and $M = \Phi + t\Phi$.

$L(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)$, I, S, and Φ act on (a^+_{ks}, a_{ks}) as follows:

$$\mathbf{n} \cdot a^+_{ks} = e^{-ikn} a^+_{ks}, \quad \mathbf{n} \cdot a_{ks} = e^{ikn} a_{ks}, \tag{2.4}$$

$$I \cdot a^+_{ks} = a^+_{-ks}, \quad I \cdot a_{ks} = a_{-ks}, \tag{2.5}$$

 0022-2488/85/071521-08$02.50

$$u(\mathbf{e},\theta)\cdot a^{+}_{ks} = \sum_{s'} \{u(\mathbf{e},\theta)\}_{s's} a^{+}_{ks'},$$
$$u(\mathbf{e},\theta)\cdot a_{ks} = \sum_{s'} \{u(\mathbf{e},\theta)\}^{*}_{s's} a_{ks'}, \tag{2.6}$$

$$\tilde{\phi}\cdot a^{+}_{ks} = e^{i\phi} a^{+}_{ks}, \quad \tilde{\phi}\cdot a_{ks} = e^{-i\phi} a_{ks}, \tag{2.7}$$

where $\mathbf{n}$ is an element of $L(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)$; $\mathbf{n} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$, $u(\mathbf{e},\theta) \in S$ is a spin rotation by θ radian around the $\mathbf{e}$ axis, and $\tilde{\phi} \in \Phi$ is a phase charge by ϕ $(0 \leqslant \phi \leqslant 2\pi)$. The time reversal t acts on (a^{+}_{ks}, a_{ks}) as follows:

$$t\cdot(f a^{+}_{k\uparrow}) = -f^{*} a^{+}_{-k\downarrow}, \quad t\cdot(f a_{k\uparrow}) = -f^{*} a_{-k\downarrow},$$
$$t\cdot(f a^{+}_{k\downarrow}) = f^{*} a_{-k\uparrow}, \quad t\cdot(f a_{k\downarrow}) = f^{*} a_{-k\uparrow}, \tag{2.8}$$

where f is a complex number. Note that $t\tilde{\phi} \neq \tilde{\phi} t$.

From the commutation relations of (a^{+}_{ks}, a_{ks}), Hermiticity of $\mathscr{H}$, and invariance of $\mathscr{H}$ for t and I, we have conditions on $V(k,k',q)$:

$$V(k,k',q) = V(-k,-k',-q) = V(k+q,k'-q,-q)$$
$$= V(k',k,-q) = V^{*}(k,k',q). \tag{2.9}$$

Note that $V(k,k',q)$ is real.

The following group theoretical analysis works almost in the same way even in the case of multiple bands or the other space group, although with some cumbersomeness.

III. IRREDUCIBLE SINGLE-VALUED REPRESENTATION OF THE SYMMETRY GROUP OVER THE REAL NUMBER FIELD

As the R-rep of G_0 characterizes[8,9] the instability of the normal paramagnetic state (NP) and is needed in the following section, in this section we give all R-reps of G_0 in the representation space $W_H = \{a^{+}_{ks} a_{k's'}, a^{+}_{ks} a^{+}_{k's'}, a_{ks} a_{k's'}\}^{h}_{c}$ where $\{A,B,...\}^{h}_{c}$ denotes a vector space spanned over the complex number by A, B, etc., such that its elements are Hermite.

Here we give some notations and definitions; G: a group, $\check{G}$: an R-rep of G, $[\check{G}]$: an invariance group of $\check{G}$, $V[\check{G}]$: the invariance vector space of $[\check{G}]$ in W_H. The invariance group of $\check{G}$ is the maximal subgroup $G' \subseteq G$ such that for some element $x(\neq 0)$ in the representation space,

$$g\cdot x = x, \quad \text{for } g \in G'. \tag{3.1}$$

An R-rep $\check{G}_0$ of G_0 is represented by the Kronecker product of the R-reps of $P(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)$, S, and M; $\check{G}_0 = \check{P}\otimes\check{S}\otimes\check{M}$. For $P(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)$ there are the following five types of R-reps: $\check{P}^{0\pm}$, $\check{P}^{\mathbf{Q}\pm}$, and $\check{P}^{\mathbf{q}}$, where

$$\check{P}^{0+}(\mathbf{n}) = \check{P}^{0+}(I\mathbf{n}) = 1, \tag{3.2}$$
$$\check{P}^{0-}(\mathbf{n}) = 1, \quad \check{P}^{0-}(I\mathbf{n}) = -1, \tag{3.3}$$
$$\check{P}^{\mathbf{Q}\pm}(\mathbf{n}) = \exp(i\mathbf{Q}\mathbf{n}), \quad \check{P}^{\mathbf{Q}\pm}(I\mathbf{n}) = \pm\exp(i\mathbf{Q}\mathbf{n}), \tag{3.4}$$
$$\check{P}^{\mathbf{q}}(\mathbf{n}) = \begin{pmatrix} \cos(\mathbf{qn}), & \sin(\mathbf{qn}) \\ -\sin(\mathbf{qn}), & \cos(\mathbf{qn}) \end{pmatrix},$$
$$\check{P}^{\mathbf{q}}(I\mathbf{n}) = \begin{pmatrix} \cos(\mathbf{qn}), & \sin(\mathbf{qn}) \\ \sin(\mathbf{qn}), & -\cos(\mathbf{qn}) \end{pmatrix}, \tag{3.5}$$

and

$$\mathbf{Q} = Q_1\mathbf{b}_1 + Q_2\mathbf{b}_2 + Q_3\mathbf{b}_3 = \tfrac{1}{2}\mathbf{K} \neq 0 \quad (Q_i = 0 \text{ or } \pm\tfrac{1}{2}),$$

$\mathbf{b}_i (i = 1,2,3)$ is the reciprocal basis vector,

$\mathbf{K}$ is a reciprocal lattice point, and $\mathbf{q}(\neq 0,\mathbf{Q})$ is a vector in the first Brillouin zone.

The R-reps of S in W_H are $\check{S}^0$ and $\check{S}^1$:

$$\check{S}^{0}(u(\mathbf{e},\theta)) = 1 \tag{3.6}$$

and

$$\check{S}^{1}(u(\mathbf{e},\theta)) = R(u(\mathbf{e},\theta)), \tag{3.7}$$

where $R(u(\mathbf{e},\theta))$ is a 3×3 orthogonal matrix satisfying

$$u\sigma_i u^{+} = \sum_j R_{ij}\sigma_j, \quad i,j = 1,2,3, \tag{3.8}$$

where $\sigma_i (i = x,y,z)$ are the Pauli matrices.

TABLE I. Bases of R-reps of G_0.

$\check{G}_0$	Basis of $\check{G}_0$
$\check{G}_0^{p\pm,x,\mu}$ [a]	$e^{p\pm,x,\mu}_{1,j,1}(k) = i^{(x+(1\mp\mu)/2)} \sum_{ss'} (a^{+}_{k+ps} a_{ks'} \pm a^{+}_{-k-ps} a_{-ks'})(\sigma_j)_{ss'}$ [b] $+ \text{H.C.}$
$p = 0, \mathbf{Q}$	
$\mu = \pm 1$	
$\check{G}_0^{q,x,\mu}$	$e^{q,x,\mu}_{1,j,1}(k) = i^{(x+(1\mp\mu)/2)} \sum_{ss'} (a^{+}_{k+qs} a_{ks'} + a^{+}_{-k-qs} a_{-ks'})(\sigma_j)_{ss'}$ $+ \text{H.C.}$
$\mu = \pm 1$	$e^{q,x,\mu}_{2,j,1}(k) = i^{(x+(3\mp\mu)/2)} \sum_{ss'} (a^{+}_{k+qs} a_{ks'} - a^{+}_{-k-qs} a_{-ks'})(\sigma_j)_{ss'}$ $+ \text{H.C.}$
$q \neq 0, \mathbf{Q}$	
$\check{G}_0^{p\pm,x,2}$	$e^{p\pm,x,2}_{1,j,m}(k) = i^{(m-x+(1\pm 1)/2)} \sum_{ss'} (a^{+}_{k+ps} a^{+}_{-ks'} \pm a^{+}_{-k-ps} a^{+}_{ks'})(i\sigma_j\sigma_2)_{ss'}$ $+ \text{H.C.}$
	$m = 1,2$
$P = 0,\mathbf{Q}$	
$\check{G}_0^{q,x,2}$	$e^{q,x,2}_{1,j,m}(k) = i^{(x+m-1)} \sum_{ss'} (a^{+}_{k+qs} a^{+}_{-ks'} + a^{+}_{-k-qs} a^{+}_{ks'})(i\sigma_j\sigma_2)_{ss'}$ $+ \text{H.C.}$
$q \neq 0, \mathbf{Q}$	$e^{q,x,2}_{2,j,m}(k) = i^{(x+m)} \sum_{ss'} (a^{+}_{k+qs} a^{+}_{-ks'} - a^{+}_{-k-qs} a^{+}_{ks'})(i\sigma_j\sigma_2)_{ss'}$ $+ \text{H.C.}$

[a] $x = 0,1$.
[b] $j = 0$, for $x = 0$, $j = 1,2,3$, for $x = 1$, $\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

The R-reps of M in W_H are $\check{M}^{\pm 1}$ and $\check{M}^2$ as follows:

$$\check{M}^{\pm 1}(\tilde{\phi}) = 1, \quad \check{M}^{\pm 1}(t\tilde{\phi}) = \pm 1, \tag{3.9}$$

$$\check{M}^2(\tilde{\phi}) = \begin{pmatrix} \cos 2\phi, & -\sin 2\phi \\ \sin 2\phi, & \cos 2\phi \end{pmatrix},$$
$$\check{M}^2(t\tilde{\phi}) = \begin{pmatrix} \cos 2\phi, & -\sin 2\phi \\ -\sin 2\phi, & -\cos 2\phi \end{pmatrix}. \tag{3.10}$$

Thus we have $5\times 2\times 3 = 30$ types of R-reps of G_0,

$$\check{G}_0^{\lambda,x,\mu} = \check{P}^\lambda \otimes \check{S}^x \otimes \check{M}^\mu, \tag{3.11}$$

where λ is $0\pm$, $\mathbf{Q}\pm$, and $\mathbf{q}$, $x = 0$ or 1, $\mu = \pm 1$ and 2. Denoting the basis of $\check{G}_0^{\lambda,x,\mu}$ by $e_{\kappa,j,\nu}^{\lambda,x,\mu}$ we have for $g = pum$ ($p\in P$, $u\in S$, $m\in M$),

$$g\cdot e_{\kappa,j,\nu}^{\lambda,x,\mu} = \sum_{\kappa',j',\nu'} \check{P}^\lambda(p)_{\kappa'\kappa}\check{S}(u)_{j'j}\check{M}(m)_{\nu'\nu}e_{\kappa',j',\nu'}^{\lambda,x,\mu}. \tag{3.12}$$

In these 30-reps, the following 6-reps of $\check{G}_0^{0-,0,1}$, $\check{G}_0^{0+,0,-1}$, $\check{G}_0^{0+,1,+1}$, $\check{G}_0^{0-,1,-1}$, $\check{G}_0^{0-,0,2}$, and $G_0^{0+,1,2}$ are forbidden in W_H for Hermite or antisymmetric conditions. The basis of each R-rep is summarized in Table I.

IV. CLASSIFICATION OF THE STATE WITH A SINGLE-ORDER PARAMETER AND THE STANDARD FORM OF THE HFB HAMILTONIAN FOR EACH CLASS

In the present section we give the invariance group $[\check{G}_0]$ of $\check{G}_0$, the HFB Hamiltonian corresponding to $[\check{G}_0]$. Also the self-consistent (SC) condition on the order parameter is obtained. The state corresponding to $[\check{G}_0]$ has a single-order parameter. We have obtained all the invariance groups $[\check{G}_0]$ in a manner similar to the previous paper[8] and we will not describe the procedure to obtain them.

For simplicity we consider only the cases of $\check{G}_0^{\lambda,x,\mu}$ with $\lambda = 0_\pm$, $\mathbf{Q}_\pm$, $\mathbf{q}$, where $\mathbf{Q} = \frac{1}{2}\mathbf{b}_1$ and $\mathbf{q} = (m_1/n_1)\mathbf{b}_1$ (m_1,n_1 are mutually prime integers, $m_1 \leqslant n_1/2$).

For other λ, $[\check{G}_0^{\lambda,x,\mu}]$ has a similar characteristic to the above cases though with some cumbersome notation. Then we do not discuss the latter cases.

We use the following definitions and notations: $x_0(k)$, $y_0(k)$, $x_1(k)$, $x_1^j(k)$, $y_1(k)$, and $y_1^j(k)$ are real numbers; $z(k)$, $z^j(k)$, $w(k)$, and $w^j(k)$ are complex numbers; $L(\mathbf{a}_1) = L(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)$, $L(2\mathbf{a}_1) = L(2\mathbf{a}_1,\mathbf{a},\mathbf{a}_3)$, $P(\mathbf{a}_1) = P(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3)$, and $P(2\mathbf{a}_1) = (1+I)L(2\mathbf{a}_1)$; $(\overline{n\mathbf{a}_1})$ is a translation by $n\mathbf{a}_1$ when n is an integer, zero translation when n is not an integer; $A(\mathbf{e}_i) = \{u(\mathbf{e}_i,\theta)|0\leqslant\theta\leqslant 4\pi\}$; u_{2i} ($i = x,y,z$) is a spin rotation by π radian around the i axis; $E_\pi = (\widetilde{0,\pi})\subset\Phi$; $M_\pi = \{E_\pi + tE_\pi\}$; CCW is the charge current wave; CDW is the charge density wave; BOW is the bond order wave; ASCW is the axial spin current wave[7]; FM is ferromagnetism; ASDW is the axial spin density wave[7]; SBOW is the spin bond order wave; HSCW is the helical spin current wave; HSDW is the helical spin density wave; SSC is the singlet superconducting state; and TSC is the triplet superconducting state.

In Table II we give $[\check{G}_0]$ and the corresponding HFB Hamiltonian. In the table the total HFB Hamiltonian for G_i ($i\neq 0$) should be understood as

$$H = H_0 + H_i + H_i^+ \ (\text{if } H_i^+ \neq H_i), \tag{4.1}$$

$x_1(k)$, $x^j(k)$, etc., are the order parameters.

Now we consider the self-consisting (SC) condition on the order parameter. The SC condition is easily obtained by (2.15) and (4.8) of I. We introduce the following notation:

$$
\begin{aligned}
{}^0W^p(k,k') &= 2V(k',k'+p,p) - V(k+p,k',k'-k),\\
{}^1W^p(k,k') &= -V(k+p,k',k'-k),\\
{}^2W^p(k,k') &= V(k+p,-k,k'-k).
\end{aligned} \tag{4.2}
$$

Then we have the SC conditions. For G_0,

$$x_0(k') = \epsilon(k') + \sum_k \langle a_{k\uparrow}^+ a_{k\uparrow}\rangle {}^0W^0(k,k'). \tag{4.3}$$

For G_1,

TABLE II. Invariance group and its HFB Hamiltonian of states with a single-order parameter.

$\check{G}_0$	$[\check{G}_0]$	HFB Hamiltonian
$\check{G}_0^{0+,0,+1}$	$G_0 = P\times S\times M$ normal paramagnetic state	$H_0 = \sum_k x_0(k)(a_{k\uparrow}^+a_{k\uparrow} + a_{k\downarrow}^+a_{k\downarrow})$ $x_0(-k) = x_0(k)$
$\check{G}_0^{0-,0,-1}$	$G_1 = \{1+tI\}L(\mathbf{a}_1)S\Phi$ CCW	$H_1 = \sum_k x_1(k)(a_{k\uparrow}^+a_{k\uparrow} + a_{k\downarrow}^+a_{k\downarrow})$ $x_1(-k) = -x_1(k)$
$\check{G}_0^{\mathbf{Q}+,0,+1}$	$G_2 = P(2\mathbf{a}_1)\times S\times M$ CDW	$H_2 = \sum_k x_1(k)(a_{k+Q\uparrow}^+a_{k\uparrow} + a_{k+Q\downarrow}^+a_{k\downarrow})$ $x_1(-k) = x_1(k)$
$\check{G}_0^{\mathbf{Q}-,0,+1}$	$G_3 = (1+\overline{\mathbf{a}_1}I)L(2\mathbf{a}_1)SM$ BOW	$H_3 = \sum_k ix_1(k)(a_{k+Q\uparrow}^+a_{k\uparrow} + a_{k+Q\downarrow}^+a_{k\downarrow})$ $x_1(-k) = -x_1(k)$
$\check{G}_0^{\mathbf{Q}+,0,-1}$	$G_4 = (1+t\,\overline{\mathbf{a}_1})P(2\mathbf{a}_1)S\Phi$ CCW	$H_4 = \sum_k ix_1(k)(a_{k+Q\uparrow}^+a_{k\uparrow} + a_{k+Q\downarrow}^+a_{k\downarrow})$ $x_1(-k) = x_1(k)$
$\check{G}_0^{\mathbf{Q}-,0,-1}$	$G_5 = (1+I\bar{\mathbf{a}}_1)(1+tI)L(2\mathbf{a}_1)S\Phi$ CCW	$H_5 = \sum_k x_1(k)(a_{k+Q\uparrow}^+a_{k\uparrow} + a_{k+Q\downarrow}a_{k\downarrow})$ $x_1(-k) = -x_1(k)$
$\check{G}_0^{\mathbf{q},0,1}$	$G_6 = P(n_1\mathbf{a})\times S\times M$ CDW	$H_6 = \sum_j\sum_k x_1^j(k)(a_{k_j\uparrow}^+a_{k\uparrow} + a_{k_j\downarrow}^+a_{k\downarrow} + a_{-k_j\uparrow}^+a_{-k\uparrow} + a_{-k_j\downarrow}^+a_{-k\downarrow})$ [a]

TABLE II. (*Continued*).

	$G_7=(1+(\overline{n_1\mathbf{a}_1/2})I)L(n_1\mathbf{a}_1)SM \quad n_1=4n+2$	CDW	$H_7=\sum_j{}'\sum_k ix_1^j(k)(a^+_{k_j\uparrow}a_{k\uparrow}+a^+_{k_j\downarrow}a_{k\downarrow} - a^+_{-k_j\uparrow}a_{-k\uparrow}-a^+_{-k_j\downarrow}a_{-k\downarrow})$ [b]
	$G_8=L(n_1\mathbf{a}_1)\times S\times M$	CDW	$H_8=\sum_j\sum_k \{Z^j(k)(a^+_{k_j\uparrow}a_{k\uparrow}+a^+_{k_j\downarrow}a_{k\downarrow}) + Z^j(k)^*(a^+_{-k_j\uparrow}a_{-k\uparrow}+a^+_{-k_j\downarrow}a_{-k\downarrow})\}$ [c]
$\check{G}_0^{q,0,-1}$	$G_9=(1+(\overline{n_1\mathbf{a}_1/2})t)P(n_1\mathbf{a}_1)S\Phi$ [d]	CCW	$H_9=\sum_j{}''\sum_k ix_1^j(k)(a^+_{k_j\uparrow}a_{k\uparrow}+a^+_{k_j\downarrow}a_{k\downarrow} + a^+_{-k_j\uparrow}a_{-k\uparrow}+a^+_{-k_j\downarrow}a_{-k\downarrow})$ [e]
	$G_{10}=(1+tI)(1+(\overline{n_1\mathbf{a}_1/2})I)L(n_1\mathbf{a}_1)S\Phi$	CCW $n_1=4n$	$H_{10}=\sum_j{}''\sum_k x_1^j(k)(a^+_{k_j\uparrow}a_{k\uparrow}+a^+_{k_j\downarrow}a_{k\downarrow} - a^+_{-k_j\uparrow}a_{-k\uparrow}-a^+_{-k_j\downarrow}a_{-k\downarrow})$
	$G_{11}=(1+(\overline{n_1\mathbf{a}_1/2})t)L(n_1\mathbf{a}_1)S\Phi$	CCW	$H_{11}=\sum_j{}''\sum_k z^j(k)(a^+_{k_j\uparrow}a_{k\uparrow}+a^+_{k_j\downarrow}a_{k\downarrow} - z^j(k)^*(a^+_{-k_j\uparrow}a_{-k\uparrow}+a^+_{-k_j\downarrow}a_{-k\downarrow})$
$\check{G}_0^{0,-1,+1}$	$G_{12}=(1+Iu_{2x})L(\mathbf{a}_1)A(\mathbf{e}_z)M$	ASCW	$H_{12}=\sum_k x_1(k)(a^+_{k\uparrow}a_{k\uparrow}-a^+_{k\downarrow}a_{k\downarrow})$, $x_1(-k)=-x_1(k)$
$\check{G}_0^{0,+1,-1}$	$G_{13}=(1+u_{2y}t)P(\mathbf{a}_1)A(\mathbf{e}_z)\Phi$	FM	$H_{13}=\sum_k x_1(k)(a^+_{k\uparrow}a_{k\uparrow}-a^+_{k\downarrow}a_{k\downarrow})$, $x_1(-k)=x_1(k)$
$\check{G}_0^{Q,+1,1}$	$G_{14}=(1+\bar{\mathbf{a}}_1u_{2x})P(2\mathbf{a}_1)A(\mathbf{e}_z)M$	ASCW	$H_{14}=\sum_k ix_1(k)(a^+_{k+Q\uparrow}a^+_{k\uparrow}-a^+_{k+Q\downarrow}a_{k\downarrow})$, $x_1(-k)=x_1(k)$
$\check{G}_0^{Q,-1,+1}$	$G_{15}=(1+\bar{\mathbf{a}}_1I)(1+Iu_{2x})L(2\mathbf{a}_1)A(\mathbf{e}_z)M$	ASCW	$H_{15}=\sum_k x_1(k)(a^+_{k+Q\uparrow}a_{k\uparrow}-a^+_{k+Q\downarrow}a_{k\downarrow})$, $x_1(-k)=-x_1(k)$
$\check{G}_0^{Q,+1,-1}$	$G_{16}=(1+\bar{\mathbf{a}}_1u_{2x})(1+tu_{2y})P(2\mathbf{a}_1)A(\mathbf{e}_z)\Phi$	ASDW	$H_{16}=\sum_k x_1(k)(a^+_{k+Q\uparrow}a_{k\uparrow}-a^+_{k+Q\downarrow}a_{k\downarrow})$, $x_1(-k)=x_1(k)$
$\check{G}_0^{Q,-1,-1}$	$G_{17}=(1+I\bar{\mathbf{a}}_1)(1+Iu_{2x})(1+tu_{2y})L(2a_1)A(\mathbf{e}_z)\Phi$	SBOW	$H_{17}=\sum_k ix_1(k)(a^+_{k+Q\uparrow}a_{k\uparrow}-a^+_{k+Q\downarrow}a_{k\downarrow})$, $x_1(-k)=x_1(k)$
$\check{G}_0^{q,1,+1}$	$G_{18}=(1+(\overline{n_1\mathbf{a}_1/2})u_{2x})P(n_1\mathbf{a}_1)A(\mathbf{e}_z)M$	ASCW	$H_{18}=\sum_j{}''\sum_k ix_1^j(k)(a^+_{k_j\uparrow}a_{k\uparrow}-a^+_{k_j\downarrow}a_{k\downarrow} + a^+_{-k_j\uparrow\,-k\uparrow}-a^+_{-k_j\downarrow\,-k\downarrow})$
	$G_{19}=(1+Iu_{2x})(1+(\overline{n_1\mathbf{a}_1/2})u_{2x})L(n_1\mathbf{a}_1)A(\mathbf{e}_z)M$	ASCW $n_1\neq 4n$	$H_{19}=\sum_j{}''\sum_k x_1^j(k)(a^+_{k_j\uparrow}a_{k\uparrow}-a^+_{k_j\downarrow}a_{k\downarrow} - a^+_{-k_j\uparrow}a_{-k\uparrow}+a^+_{-k_j\downarrow}a_{-k\downarrow}a_{-k\downarrow})$
	$G_{20}=(1+(\overline{n_1\mathbf{a}_1/2})u_{2x})L(n_1\mathbf{a}_1)A(\mathbf{e}_z)M$	ASCW	$H_{20}=\sum_j\sum_k \{Z^j(k)(a^+_{k_j\uparrow}a_{k\uparrow}-a^+_{k_j\downarrow}a_{k\downarrow}) - Z^j(k)^*(a^+_{-k_j\uparrow\,-k\uparrow}-a^+_{-k_j\downarrow}a_{-k\downarrow})\}$
	$G_{21}=(1+Iu_{2x})G_{TS}(q)M$ [f]	HSCW	$H_{21}=\sum_k ix_1(k)(a^+_{k+q\uparrow}a_{k\downarrow}+a^+_{-k-q\downarrow}Q_{-k\uparrow})$
$\check{G}_0^{q,1,-1}$	$G_{22}=(1+tu_{2y})(1+(\overline{n_1\mathbf{a}_1/2})u_{2x})P(n_1\mathbf{a}_1)A(\mathbf{e}_z)\Phi$	ASDW	$H_{22}=\sum_j{}''\sum_k x_1^j(k)(a^+_{k_j\uparrow}a_{k\uparrow}-a^+_{k_j\downarrow}a_{k\downarrow} + a^+_{-k_j\uparrow}a_{-k\uparrow}-a^+_{-k_j\downarrow}a_{-k\downarrow})$
	$G_{23}=(1+Iu_{2x})(1+tu_{2y})(1+(\overline{n_1\mathbf{a}_1/2})I) \times L(n_1\mathbf{a}_1)A(\mathbf{e}_z)\Phi$	ASDW	$H_{23}=\sum_j{}''\sum_k ix_1^j(k)(a^+_{k_j\uparrow}a_{k\uparrow}-a^+_{k_j\downarrow}a_{k\downarrow} - a^+_{-k_j\uparrow}a_{-k\uparrow}+a^+_{-k_j\downarrow}a_{-k\downarrow})$
	$G_{24}=(1+tu_{2y})L(n_1\mathbf{a}_1)A(e_z)\Phi$	ASDW	$H_{24}=\sum_j\sum_k Z^j(k)(a^+_{k_j\uparrow}a_{k\uparrow}-a^+_{k_j\downarrow}a_{k\downarrow} + Z^j(k)^*(a^+_{-k_j\uparrow}a_{-k\uparrow}-a^+_{-k_j\downarrow}a_{-k\downarrow})$
	$G_{25}=(1+Iu_{2x})(1+tu_{2y})G_{TS}(q)\Phi$	HSDW	$H_{25}=\sum_k x_1(k)(a^+_{k+q\uparrow}a_{k\downarrow}+a^+_{-k-q\downarrow}a_{-k\uparrow})$

TABLE II. (*Continued*).

$\check{G}^{0_+,0,2}$	$G_{26}=PSM_\pi$ [g]	$H_{26}=\sum_k y_0(k)(a^+_{k\uparrow}a^+_{-k\downarrow}-a^+_{k\downarrow}a^+_{-k\uparrow})$
	BCS	$y_0(-k)=y_0(k)$
$\check{G}^{Q_+,0,2}$	$G_{27}=(1+\bar{\mathbf{a}}_1\widetilde{(\pi/2)})P(2\mathbf{a}_1)SM_\pi$	$H_{27}=\sum_k y_1(k)(a^+_{k+Q\uparrow}a^+_{-k\downarrow}-a^+_{k+Q\downarrow}a^+_{-k\uparrow})$
	SSC	$y_1(-k)=y_1(k)$
$\check{G}^{Q_-,0,2}$	$G_{28}=(1+\bar{\mathbf{a}}_1 I)(1+I\widetilde{(\pi/2)})L(2\mathbf{a}_1)SM_\pi$	$H_{28}=\sum_k iy_1(k)(a^+_{k+Q\uparrow}a^+_{-k\downarrow}-a^+_{k+Q\downarrow}a^+_{-k\uparrow})$
	SSC	$y_1(k)=-y_1(k)$
$\check{G}^{q,0,2}_0$	$G_{29}=(1+(\overline{n_1\mathbf{a}_1/2})\widetilde{(\pi/2)})P(n_1\mathbf{a}_1)SM_\pi$	$H_{29}=\sum_j{}''\sum_k y^j_1(k)(a^+_{k_j\uparrow}a^+_{-k\downarrow}-a^+_{k_j\downarrow}a^+_{-k\uparrow}$
	SSC	$+a^+_{-k_j\uparrow}a^+_{k\downarrow}-a^+_{-k_j\downarrow}a^+_{k\uparrow})$
	$G_{30}=(1+(\overline{n_1\mathbf{a}_1/2})I)(1+I\cdot\widetilde{(\pi/2)})$	$H_{30}=\sum_j{}''\sum_k iy^j_1(k)(a^+_{k_j\uparrow}a^+_{-k\downarrow}-a^+_{k_j\downarrow}a^+_{-k\uparrow}$
	$\times L(n_1\mathbf{a})SM_\pi,\quad n_1\neq 4n$	$-a^+_{-k_j\uparrow}a^+_{k\downarrow}+a^+_{-k_j\downarrow}a^+_{k\uparrow})$
	SSC	
	$G_{31}=(1+(\overline{n_1\mathbf{a}_1/2})\widetilde{(\pi/2)})L(n_1\mathbf{a})SM_\pi$	$H_{31}=\sum_j{}''\sum_k\{w^j(k)(a^+_{k_j\uparrow}a^+_{-k\downarrow}-a^+_{k_j\downarrow}a^+_{-k\uparrow})$
	SSC	$+w^j(k)^*(a^+_{-k_j\uparrow}a^+_{k\downarrow}-a^+_{-k_j\downarrow}a^+_{k\uparrow})\}$
	$G_{32}=(1+tI)G_{TG}(q)S$ [h]	$H_{32}=\sum_k y_1(k)(a^+_{k+q\uparrow}a^+_{-k\downarrow}-a^+_{k+q\downarrow}a^+_{-k\uparrow})$
	SSC	
$\check{G}^{0_-,1,2}_0$	$G_{33}=(1+Iu_{2x})(1+u_{2x}\widetilde{(\pi/2)})L(\mathbf{a}_1)A(\mathbf{e}_z)M_\pi$	$H_{33}=\sum_k y_1(k)(a^+_{k\uparrow}a^+_{-k\downarrow}+a^+_{k\downarrow}a^+_{-k\uparrow})$
	Triplet super conducting state TSC	$y_1(-k)=-y_1(k)$
	$G_{34}=(1+Iu_{2z})(1+tu_{2x})(1+(\overline{n_1\mathbf{a}_1/2})\widetilde{(\pi/2)})$ [i]	$H_{34}=\sum_k y_1(k)a^+_{k\uparrow}a^+_{-k\uparrow}$
	$\times L(\mathbf{a}_1)G_{SG}$	$y_1(-k)=-y_1(k)$
	TSC	
$\check{G}^{Q_+,1,2}$	$G_{35}=(1+\bar{\mathbf{a}}_1u_{2x})(1+tu_{2y})(1+\bar{\mathbf{a}}_1\widetilde{(\pi/2)})$ [j]	$H_{35}=\sum_k y_1(k)(a^+_{k+Q\uparrow}a^+_{-k\downarrow}+a^+_{k+Q\downarrow}a^+_{-k\uparrow})$
	$\times P(2\mathbf{a}_1)A(\mathbf{e}_z)E_\pi$	$y_1(-k)=y_1(k)$
	TSC	
	$G_{36}=(1+\bar{\mathbf{a}}_1u_{2z})(1+tu_{2y})(1+\bar{\mathbf{a}}_1\widetilde{(\pi/2)})P(2\mathbf{a}_1)G_{SG}$	$H_{36}=\sum_k y_1(k)a^+_{k+Q\uparrow}a^+_{-k\uparrow}$
		$y_1(-k)=y_1(k)$
$\check{G}^{Q_-,1,2}_0$	$G_{37}=(1+I\bar{\mathbf{a}}_1)(1+Iu_{2x})(1+tu_{2y})(1+\bar{\mathbf{a}}_1\widetilde{(\pi/2)})$	$H_{37}=\sum_k iy_1(k)(a^+_{k+Q\uparrow}a^+_{-k\downarrow}+a^+_{k+Q\downarrow}a^+_{-k\uparrow})$
	$\times L(2\mathbf{a}_1)A(\mathbf{e}_z)E_\pi$	$y_1(-k)=-y_1(k)$
	TSC	
	$G_{38}=(1+I\bar{\mathbf{a}}_1)(1+\bar{\mathbf{a}}_1u_{2z})(1+tu_{2y})(1+\bar{\mathbf{a}}_1\widetilde{(\pi/2)})$	$H_{38}=\sum_k iy_1(k)a^+_{k+Q\uparrow}a^+_{-k\uparrow}$
	$\times L(2\mathbf{a}_1)G_{SG}$	$y_1(-k)=-y_1(k)$
	TSC	
$\check{G}^{q,1,2}_0$	$G_{39}=(1+(\overline{n_1\mathbf{a}_1/2})u_{2x})(1+tu_{2y})$	$H_{39}=\sum_j{}''\sum_k y^j_1(k)(a^+_{k_j\uparrow}a^+_{-k\downarrow}+a^+_{k_j\downarrow}a^+_{-k\uparrow}$
	$\times(1+u_{2x}\widetilde{(\pi/2)})P(n_1\mathbf{a}_1)A(\mathbf{e}_z)E_\pi$	$+a^+_{-k_j\uparrow}a^+_{k\downarrow}+a^+_{-k_j\downarrow}a^+_{k\uparrow})$
	TSC	
	$G_{40}=(1+(\overline{n_1\mathbf{a}_1/2})I)(1+Iu_{2x})(1+I\widetilde{(\pi/2)})$	$H_{40}=\sum_j{}''\sum_k iy^j_1(k)(a^+_{k_j\uparrow}a^+_{-k\downarrow}+a^+_{k_j\downarrow}a^+_{-k\uparrow}$
	$\times(1+tu_{2y})L(n_1\mathbf{a}_1)A(\mathbf{e}_z)E_\pi n_1\neq 4n$	$-a^+_{-k_j\uparrow}a^+_{k\downarrow}-a^+_{-k_j\downarrow}a^+_{k\uparrow})$
	TSC	
	$G_{41}=(1+(\overline{n_1\mathbf{a}_1/2})u_{2x})(1+u_{2x}\widetilde{(\pi/2)})(1+tu_{2y})$	$H_{41}=\sum_j{}''\sum_k\{w^j(k)(a^+_{k_j\uparrow}a^+_{-k\downarrow}+a^+_{k_j\downarrow}a^+_{-k\uparrow})$
	$\times L(n_1\mathbf{a}_1)A(\mathbf{e}_z)E_\pi$	$-w^j(k)^*(a^+_{-k_j\uparrow}a^+_{k\downarrow}+a^+_{-k_j\downarrow}a^+_{k\uparrow})\}$
	TSC	

$G_{42}=(1+Iu_{2y}t)(1+u_{2x}\widetilde{(\pi/2)})G_{TG}(q)A(\mathbf{e}_z)$	$H_{42}=\sum_k y_1(k)(a^+_{k+q\uparrow}a^+_{-k\downarrow}+a^+_{k+q\downarrow}a_{-k\uparrow})$
TSC	
$G_{43}=(1+Iu_{2x})(1+tu_{2z})(1+u_{2z}\widetilde{(\pi/2)})G_{TS}E_\pi$	$H_{43}=\sum_k y_1(k)(a^+_{k+q\uparrow}a^+_{-k\uparrow}-a^+_{-k-q\downarrow}a^+_{k\downarrow})$
TSC	
$G_{44}=(1+(\overline{n_1\mathbf{a}_1/2})u_{2z})(1+tu_{2y})(1+u_{2z}\widetilde{(\pi/2)})$ $\times P(n_1\mathbf{a}_1)G_{SG}$	$H_{44}=\sum_j{}''\sum_k y^j_1(k)(a^+_{k_j\uparrow}a^+_{-k\uparrow}+a^+_{-k_j\uparrow}a^+_{k\uparrow})$
TSC	
$G_{45}=(1+Iu_{2z})(1+I(\overline{n_1\mathbf{a}_1/2}))(1+tu_{2y})$ $\times(1+u_{2z}\widetilde{(\pi/2)})L(n_1\mathbf{a}_1)G_{SG}$	$H_{45}=\sum_j{}''\sum_k iy^j_1(k)(a^+_{k_j\uparrow}a^+_{-k\uparrow}-a^+_{-k_j\uparrow}a^+_{k\uparrow})$
TSC	
$G_{46}=(1+tu_{2y})(1+u_{2z}\widetilde{(\pi/2)})(1+u_{2z}(\overline{n_1\mathbf{a}_1/2}))$ $\times L(n_1\mathbf{a}_1)G_{SG}$	$H_{46}=\sum_j{}''\sum_k w^j(k)(a^+_{k_j\uparrow}a^+_{-k\uparrow}-a^+_{-k_j\uparrow}a^+_{k\uparrow})$
TSC	
$G_{47}=\{\overline{n\mathbf{a}_1}u_z(\theta)((\theta+2\pi(m_1/n_1)n)/2)\}L(n_1\mathbf{a}_1)$	$H_{47}=\sum_k y_1(k)(a^+_{k+q\uparrow}a^+_{-k\uparrow})$
TSC	

[a] $k_j=k+q_j,\quad q_j=(j/n_1)b_1,\ \sum_j=\sum_{j=1}^{[n_1/2]}$, where $[n_1/2]$ denotes the maximum integer $\leqslant n_1/2$.
[b] $\sum_j{}'$ means that j runs odd number $<[n_1/2]$.
[c] $Z^j(k)$ is a complex number.
[d] $(\overline{n_1\mathbf{a}_1/2})$ is put to zero when n_1 is odd.
[e] $\sum_j{}''=\sum_j{}'$, for even n_1, $=\sum_j$, for odd n_1.
[f] $G_{TS}(q)=\{\overline{n\mathbf{a}_1}u_z(-2\pi(m_1/n_1)n)|n=0,\pm 1,\ldots\}$.
[g] $M_\pi=(\tilde{0},\tilde{\pi})+t(\tilde{0},\tilde{\pi})$.
[h] $G_{TG}(q)=\{(\overline{n\mathbf{a}_1})\widetilde{(\pi(m_1/n_1)n)}|n=0,\pm 1,\ldots\}$.
[i] $G_{SG}=\{u_z(\phi)\widetilde{(\phi/2)}|0<\phi<4\pi\}$.
[j] $E_\pi=(\tilde{0},\tilde{\pi})$.

$$x_0(k')=\epsilon(k')+\frac{1}{2}\sum_k\{\langle a^+_{k\uparrow}a_{k\uparrow}\rangle+\langle a^+_{-k\uparrow}a_{-k\uparrow}\rangle\}^0W^0(k,k'),\tag{4.4}$$

$$x_1(k')=\frac{1}{2}\sum_k\{\langle a^+_{k\uparrow}a_{k\uparrow}\rangle-\langle a^+_{-k\uparrow}a_{-k\uparrow}\rangle\}^0W^0(k,k').$$

For G_{12},

$$x_0(k')=\epsilon(k')+\frac{1}{4}\sum_k\{\langle a^+_{k\uparrow}a_{k\uparrow}\rangle+\langle a^+_{k\downarrow}a_{k\downarrow}\rangle+\langle a^+_{-k\uparrow}a_{-k\uparrow}\rangle+\langle a^+_{-k\downarrow}a_{-k\downarrow}\rangle\}^0W^0(k,k'),\tag{4.5}$$

$$x_1(k')=\frac{1}{4}\sum_k\{\langle a^+_{k\uparrow}a_{k\uparrow}\rangle-\langle a^+_{k\downarrow}a_{k\downarrow}\rangle-\langle a^+_{-k\uparrow}a_{-k\uparrow}\rangle+\langle a^+_{-k\downarrow}a_{-k\downarrow}\rangle\}^1W^0(k,k').$$

For G_{13}, $x_0(k')$ has the same SC condition as (4.5), and

$$x_1(k')=\frac{1}{4}\sum_k\{\langle a^+_{k\uparrow}a_{k\uparrow}\rangle-\langle a^+_{k\downarrow}a_{k\downarrow}\rangle+\langle a^+_{-k\uparrow}a_{-k\uparrow}\rangle-\langle a^+_{-k\downarrow}a_{-k\downarrow}\rangle\}^1W^0(k,k').\tag{4.6}$$

For $G_i(2\leqslant i\leqslant 11$ or $14\leqslant i\leqslant 25)$, H_i is generally expressed by

$$H_i=\sum_j\sum_{ss'}\sum_k u^j(k)_{ss'}(a^+_{k+p_js}a_{ks'}\pm\cdots),\tag{4.7}$$

where the $u^j(k)_{ss'}$ are $x_1(k)$, $x^j_1(k)$, $ix_1(k)$, etc. The SC condition is given as follows. The SC condition on $x_0(k')$ is (4.3) and

$$u^j(k')_{ss'}=\sum_k\langle a^+_{ks'}a_{k+p_js}\rangle^xW^{p_j}(k,k'),\tag{4.8}$$

where $x=0$ for $i\leqslant 11$ and $x=1$ for $14\leqslant i\leqslant 25$. For G_i $(26\leqslant i\leqslant 47)$, H_i is generally expressed by

$$H_i=\sum_j\sum_{ss'}\sum_k v^j(k)_{ss'}(a^+_{k+p_js}a^+_{-ks'}\pm\cdots).\tag{4.9}$$

Then the SC condition is given as follows: $x_0(k')$ has the same form as (4.3) and

$$v^j(k')_{ss'}=\sum_k\langle a_{-ks'}a_{k+p_js}\rangle^2W^{p_j}(k,k').\tag{4.10}$$

Here we give a few comments on some classes of Table II: G_0 is the normal paramagnetic state (NP); G_2 corresponds to the charge density wave (CDW) state in the half-filled band case; and G_3 is the bond order wave (BOW) state (or off-diagonal CDW state) which is obtained by Čížek and Paldus.[10] Note that the CDW state is invariant to I and the BOW state is antisymmetric to I. We have that G_6 is a CDW state with inversion symmetry. The case with $\mathbf{q} = (1/3)\mathbf{b}_1$ especially has been discussed in detail in relation to the fractional charge soliton by Su and Schrieffer.[11] Also, G_8 is a CDW state-breaking inversion symmetry, which is a hitherto unknown type; G_{13} is the ferromagnetic (FM) state; G_{16} is the spin density wave (SDW) or antiferromagnetic (AF) state; and G_{17} is the spin bond order wave (SBOW) (see Ref. 10) (or off-diagonal SDW). Note that the SDW state is symmetric to I and SBOW is antisymmetric to I. We have that G_{25} is the Overhauser's[12] helical spin density wave (HSDW) state; and G_{26} is the usual BCS state.

$G_{TS}(q)$, $G_{TG}(q)$, and G_{SG} in Table II appearing in (G_{21}, G_{25}, G_{43}), G_{42}, and $(G_{34}, G_{36}, G_{38}, G_{41}, G_{45}, G_{46})$ are the relative translation-spin, the relative translation-gauge, and the relative spin-gauge symmetry in the usage of Liu.[13] In this usage G_{47} may be called the relative translation-spin-gauge symmetry.

Thus all the known ordered states are included in Table II. If other ordered states with a single-order parameter are found, these must be in Table II.

V. CLASSIFICATION OF THE BCS STATE COEXISTENT WITH A NONSUPERCONDUCTING ORDER

In this section we consider the states with multiple-order parameters such as the coexisting state of superconductivity and magnetism. For this purpose we consider the BCS superconducting state characterized by $G_s = G_{26} = P(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3) \times S \times M_\pi$ of Table II. The coexisting states are obtained from the invariance group $[\check{G}_s]$ of the R-rep $\check{G}_s$ of G_s. The R-reps of G_s are expressed by

TABLE III. The invariance group and its HFB Hamiltonian of BCS coexistent state with other nonsuperconducting orders.

$\check{G}_s$	$[\check{G}_s]$	H_{si}
$\check{G}_s^{0_+,0,+1}$	$G_{s0} = G_s = P(\mathbf{a}_1)SM_\pi$	
$\check{G}_s^{0_-,0,-1}$	$G_{s1} = (1 + tI)L(\mathbf{a}_1)SE_\pi$	H_1
$\check{G}_s^{Q_+,0,+1}$	$G_{s2} = P(2\mathbf{a}_1)SM_\pi$	$H_2 + H_{27}$
$\check{G}_s^{Q_-,0,+1}$	$G_{s3} = (1 + \bar{\mathbf{a}}_1 I)L(2\mathbf{a}_1)SM_\pi$	$H_3 + H_{28}$
$\check{G}_s^{Q_+,0,-1}$	$G_{s4} = (1 + \bar{\mathbf{a}}_1 t)P(2\mathbf{a}_1)SE_\pi$	$H_4 + iH_{27}$
$\check{G}_s^{Q_-,0,-1}$	$G_{s5} = (1 + I\bar{\mathbf{a}}_1)(1 + \bar{\mathbf{a}}_1 t)L(2\mathbf{a}_1)SE_\pi$	$H_5 + iH_{28}$
$\check{G}_s^{\mathbf{q},0,1}$	$G_{s6} = P(n_1\mathbf{a}_1)SM_\pi$	$H_6 + H_{29}$
	$G_{s7} = (1 + (\overline{n_1\mathbf{a}_1/2})I)L(n_1\mathbf{a}_1)SM_\pi$	$H_7 + H_{30}$
	$G_{s8} = L(n_1\mathbf{a}_1)SM_\pi$	$H_8 + H_{31}$
$\check{G}_s^{\mathbf{q},0,-1}$	$G_{s9} = (1 + (\overline{n_1\mathbf{a}_1/2})t)P(n_1\mathbf{a}_1)SE_\pi$	$H_9 + iH_{29}$
	$G_{s10} = (1 + (\overline{n_1\mathbf{a}_1/2})t)(1 + tI)L(n_1\mathbf{a}_1)SE_\pi$	$H_{10} + iH_{30}$
	$G_{s11} = (1 + (\overline{n_1\mathbf{a}_1/2})t)L(n_1\mathbf{a}_1)SE_\pi$	$H_{11} + iH_{31}$
$\check{G}_s^{0_-,1,1}$	$G_{s12} = (1 + Iu_{2x})L(\mathbf{a}_1)A(\mathbf{e}_z)M_\pi$	$H_{12} + H_{33}$
$\check{G}_s^{0_+,1,-1}$	$G_{s13} = (1 + tu_{2y})P(\mathbf{a}_1)A(\mathbf{e}_z)E_\pi$	H_{13}
$\check{G}_s^{0_-,1,-1}$	$G_{s14} = (1 + Iu_{2x})(1 + tu_{2y})L(\mathbf{a}_1)A(\mathbf{e}_z)E_\pi$	iH_{33} [a]
$\check{G}_s^{Q_+,1,+1}$	$G_{s15} = (1 + \bar{\mathbf{a}}_1 u_{2x})P(2\mathbf{a}_1)A(\mathbf{e}_z)M_\pi$	$H_{14} + iH_{35}$
$\check{G}_s^{Q_-,1,1}$	$G_{s16} = (1 + \bar{\mathbf{a}}_1 u_{2x})(1 + \bar{\mathbf{a}}_1 I)L(2\mathbf{a}_1)A(\mathbf{e}_z)M_\pi$	$H_{15} + iH_{37}$
$\check{G}_s^{Q_+,1,-1}$	$G_{s17} = (1 + \bar{\mathbf{a}}_1 u_{2x})(1 + tu_{2y})P(2\mathbf{a}_1)A(\mathbf{e}_z)E_\pi$	$H_{16} + H_{35}$
$\check{G}_s^{Q_-,1,-1}$	$G_{s18} = (1 + \bar{\mathbf{a}}_1 u_{2x})(1 + tu_{2y})(1 + Iu_{2x})L(2\mathbf{a}_1)A(\mathbf{e}_z)E_\pi$	$H_{17} + H_{37}$
$\check{G}_s^{\mathbf{q},1,1}$	$G_{s19} = (1 + (\overline{n_1\mathbf{a}_1/2})u_{2x})P(n_1\mathbf{a}_1)A(\mathbf{e}_z)M_\pi$	$H_{18} + iH_{39}$
	$G_{s20} = (1 + (\overline{n_1\mathbf{a}_1/2})u_{2x})(1 + Iu_{2x})L(n_1\mathbf{a}_1)A(\mathbf{e}_z)M_\pi$	$H_{19} + iH_{40}$
	$G_{s21} = (1 + (\overline{n_1\mathbf{a}_1/2})u_{2x})L(n_1\mathbf{a}_1)A(\mathbf{e}_z)M_\pi$	$H_{20} + iH_{41}$
	$G_{s22} = (1 + Iu_{2x})G_{TS}(q)M_\pi$	$H_{21} + iH_{43}$
$\check{G}_s^{\mathbf{q},1,-1}$	$G_{s23} = (1 + (\overline{n_1\mathbf{a}_1/2})u_{2x})(1 + tu_{2y})P(n_1\mathbf{a}_1)A(\mathbf{e}_z)E_\pi$	$H_{22} + H_{39}$
	$G_{s24} = (1 + (\overline{n_1\mathbf{a}_1/2})u_{2x})(1 + tu_{2y})(1 + Iu_{2x})L(n_1\mathbf{a}_1)A(\mathbf{e}_z)E_\pi$	$H_{23} + H_{40}$
	$G_{s25} = (1 + (\overline{n_1\mathbf{a}_1/2})u_{2x})(1 + ty_{2y})L(n_1\mathbf{a}_1)A(\mathbf{e}_z)E_\pi$	$H_{24} + H_{41}$
	$G_{s26} = (1 + Iu_{2x})(1 + tu_{2z})G_{TS}(q)E_\pi$	$H_{25} + H_{43}$

[a] This state does not have nonsuperconducting order.

$$\check{G}_s^{\lambda,x,\mu} = \check{P}^\lambda \otimes \check{S}^x \otimes \check{M}_\pi^\mu, \tag{5.1}$$

where $\lambda = 0_\pm, \mathbf{Q}_\pm, \mathbf{q}, x = 0, 1, \mu = \pm 1$, and

$$\check{M}_\pi^{\pm 1}(\tilde{0}) = \check{M}_\pi^{\pm}(\tilde{\pi}) = 1, \quad \check{M}_\pi^{\pm 1}(t) = \check{M}_\pi^{\pm 1}(t\tilde{\pi}) = \pm 1. \tag{5.2}$$

By the similar manner to Sec. IV we have obtained all invariance groups of $\check{G}_s$. These are listed in the second column of Table III. The HFB Hamiltonian corresponding to G_{si} is obtained by looking for the invariance vector space $V[\check{G}_s]$ of $G_{si} = [\check{G}_s]$. We can immediately obtain the HFB Hamiltonians of invariance groups $[\check{G}_s^{\lambda,0,+1}]$ and $[\check{G}_s^{\lambda,1,-1}]$ ($\lambda = 0_\pm, \mathbf{Q}_\pm, \mathbf{q}$) from Table II and $[\check{G}_s]$ of Table III. For example, for $G_{s2} = P(2\mathbf{a}_1)SM_\pi$, it is easy to see that H_2 and H_{27} are invariant on G_{s2} because $G_2 = P(2\mathbf{a}_1)SM \supset G_{s2}$ and $G_{27} = (1 + \bar{\mathbf{a}}_1(\widetilde{\pi/2}))P(2\mathbf{a}_1)SM_\pi \supset G_{s2}$.

But for $[\check{G}_s^{\lambda,0,-1}]$ and $[\check{G}_s^{\lambda,1,+1}]$ the HFB Hamiltonian cannot be directly obtained from Table II and $[\check{G}_s]$ of Table III. For example, for $G_{s4} = (1 + \bar{\mathbf{a}}_1 t)P(2a_1)SE_\pi$, it is easy to see that H_4 belongs to G_{s4} as $G_4 = (1 + \bar{\mathbf{a}}_1 t) \times P(2\mathbf{a}_1)S\Phi \supset G_{s4}$. It seems that there is no other H_i which belongs to G_{s4}. However, we can obtain G'_{27} from G_{27} by conjugation of $(\widetilde{\pi/4})$:

$$G_{27} = (\widetilde{\pi/4})G_{27}(\widetilde{\pi/4})^{-1}$$
$$= (1 + \bar{\mathbf{a}}_1(\widetilde{\pi/2}))(1 + \bar{\mathbf{a}}_1 t)P(2a_1)SE_\pi \supset G_{s4}. \tag{5.3}$$

Then $(\widetilde{\pi/4})H_{27} = iH_{27}$ is G'_{27} invariant, thus the iH_{27} belong to G_{s4}.

For other cases, in a similar manner we can obtain the HFB Hamiltonians. These HFB Hamiltonians for G_{si} ($i = 1,\ldots,26$) are listed in the third column of Table III. In this column the total HFB Hamiltonian for G_{si} should be understood as

$$H = H_0 + H_{26} + H_{26}^+ + H_{si} + (H_{si}^+ \text{ if } H_{si} \neq H_{si}^+). \tag{5.4}$$

Note that in each of the coexisting states except G_{s1} and G_{s13} there occurs another abnormal (non-BCS) Cooper pair. G_{s13} is the BCS state coexistent with ferromagnetism[6] and does not involve an abnormal Cooper pair because the spatially homogeneous triplet Cooper pair is inversion antisymmetric while G_{s13} is inversion symmetric.

G_{s2}, G_{s3}, and G_{s17} are the BCS coexistent states with CDW (see Ref. 3), BOW, and SDW (see Refs. 2 and 5).

The other coexistent states with multiple-order parameters, such as CDW and BOW, can be obtained by the same method as above.

VI. CONCLUSION

The main results of the present paper are Tables II and III, which summarize the group theoretical classifications of the state with a single-order parameter and the BCS state coexistent with other nonsuperconducting orders. The tables give also the standard HFB Hamiltonian for each class in the case of a single band model.

It is an interesting and important problem whether every class of the tables has physical significance or not and, if it has, under what conditions. However, this problem must be solved numerically and individually in the concrete case. Our tables are useful for the systematic search of realizable phases for the individual case.

These classifications work also for the electron system with multiple bands, though with slight modifications in the HFB Hamiltonians of the third column of the tables.

In the case of a crystal with the other space groups $\check{P}'$ than P, we must consider the R-rep $\check{G}_0 = \check{P}' \otimes \check{S} \otimes \check{M}$ and its invariance group $[\check{G}_0]$. Also in this case $[\check{G}_0]$ can be obtained in a manner similar to that in the previous paper.[8]

Finally we note that the present classification works similarly for the mean field theory of the coexistence of the BCS superconductivity and the magnetic order by the localized spins.[14]

ACKNOWLEDGMENTS

The author would like to express his sincere thanks to Professor Fukutome and Dr. Machida for helpful advice and discussions.

[1]C. A. Balseiro and L. M. Falicov, in *Superconductivity in d and f-band Metals*, edited by H. Suhl and M. B. Maple (Academic, New York, 1980), pp. 105–112; M. Ishikawa and P. Fisher, Solid State Commun. **24**, 747 (1977).

[2]K. Machida, J. Phys. Soc. Jpn. **50**, 2195 (1981); **50**, 3231 (1981).

[3]K. Machida, T. Koyama, and T. Matsubara, Phys. Rev. B **23**, 99 (1981).

[4]G. Balseiro and L. M. Falicov, Phys. Rev. B **20**, 4457 (1979).

[5]E. W. Fenton and G. C. Psaltakis, Solid State Commun. **45**, 5 (1982).

[6]H. Nakanishi, K. Machida, and T. Matsubara, Solid State Commun. **43**, 899 (1982).

[7]H. Fukutome, Prog. Theor. Phys. **52**, 115 (1974); **52**, 1766 (1974).

[8]M. Ozaki and H. Fukutome, Prog. Theor. Phys. **60**, 1322 (1978); M. Ozaki, *ibid.* **67**, 83 (1982); **67**, 415 (1982).

[9]M. Ozaki, J. Math. Phys. **26**, 1514 (1985).

[10]J. Čížek and J. Paldus, J. Chem. Phys. **47**, 3976 (1967); J. Paldus and J. Čížek, Phys. Rev. A **2**, 2268 (1970).

[11]W. P. Su and F. R. Schrieffer, Phys. Rev. Lett. **46**, 738 (1981).

[12]A. W. Overhauser, Phys. Rev. **128**, 1437 (1962).

[13]M. Liu, Physica B **109 & 110**, 1615 (1982).

[14]K. Machida, K. Nokura, and T. Matsubara, Phys. Rev. B **22**, 2307 (1982).

An alternate approach to finding and using the Lie group of the Vlasov equation

Dana Roberts
Department of Electrical Engineering, Washington University, St. Louis, Missouri 63130

(Received 29 October 1984; accepted for publication 29 March 1985)

Two recent methods for finding exact solutions to the Vlasov–Maxwell equations using Lie group theory are compared by the introduction of an "intermediate" approach. In the latter, the Lie group and general similarity solutions of the Vlasov equation are found through a method which treats independent and dependent variables as forms that are on an essentially equal footing. A Maxwell equation is then used to constrain the solutions further. The procedure is shown to illustrate a more general theorem that implies that the reduction of the number of variables in a set of equations through the use of canonical variables generated from the Lie group invariance of one equation in the set leads to the same solutions as are found by considering the invariance of the entire set.

I. INTRODUCTION

Recently a variety of methods have been used to find exact solutions to the Vlasov–Maxwell equations that are more general than those of the BGK[1] type. One set of methods has involved the use of Lie groups to construct solutions that exploit the invariance properties of the equations.[2–5] Abraham-Shrauner used Lie group methods presented in Cohen[6] to find the Lie point group of the Vlasov equation alone, and then subsequently required the solutions thus found to satisfy Maxwell's equations. These solutions were put in a form that could be easily compared to the solutions of Lewis and Symon [7] that provided part of the motivation for the Lie group approach. Roberts used the point of view of Bluman and Cole[8] to find the point group and general similarity solutions of the Vlasov–Maxwell equations considered simultaneously. Many of these solutions were found to be equivalent to those found by Abraham-Shrauner, but it was difficult to see if the methods were actually equivalent in results. The point of this work is to show that using the approach of Bluman and Cole in constructing the invariance group of the Vlasov equation alone both simplifies finding some of the results of the method based on Cohen's presentation and sheds light on the above question of the equivalence of methods for finding solutions to the Vlasov–Maxwell equations. The present method also generates the group and associated possible forms for the electric field for the Vlasov equation without the need for auxiliary differential equations; this both makes the explicit form of the group more transparent and allows the straightforward application of the Maxwell equation constraints on the solutions. In what follows, previous methods will be presented briefly; next, the group of the Vlasov equation and associated results will be developed by an alternate method, and these results will be compared to the results of previous methods. Finally, it will be shown that the present case illustrates a more general phenomenon: Lie group solutions to sets of equations based on the invariance of one of the equations will, under specific conditions, be equivalent to those found by considering the invariance of the entire set.

II. PREVIOUS METHODS

Abraham-Shrauner,[2,3] following Cohen,[6] uses the fact that the Lie group that leaves invariant the one-dimensional equation of motion of a particle of charge q_α, mass m_α, position x, and velocity v, in an electric field E,

$$\frac{dv}{dt}=\frac{q_\alpha}{m_\alpha}E, \tag{1a}$$

with

$$\frac{dx}{dt}=v, \tag{1b}$$

also leaves invariant the associated partial differential equation (the Vlasov equation)

$$\frac{\partial f_\alpha}{\partial t}+v\frac{\partial f_\alpha}{\partial x}+\left(\frac{q_\alpha}{m_\alpha}\right)E\frac{\partial f_\alpha}{\partial v}=0. \tag{2}$$

This is essentially due to the fact that the level sets of f_α solve (1), or equivalently that Eqs. (1) are the characteristic equations of (2). Roberts,[5] following Bluman and Cole,[8] starts directly with (2) and the two associated nontrivial Maxwell equations for E to find the Lie group that leaves all three invariant. This procedure leads to an initially more restricted set of solutions than the Abraham-Shrauner procedure, but in the latter case, applying Maxwell's equations as constraints similarly restricts the set of solutions. The differences in the results, as will be shown in detail below, are primarily (1) Abraham-Shrauner's method (as outlined above; see her paper[2] for an alternate approach) limits the solutions by implicitly taking f_α (which is not used in the procedure) to be invariant under the group action; (2) in the case of a plasma with a space- and time-varying background, Roberts' method only allows those solutions for which the background has specific group transformation properties, unlike Abraham-Shrauner's method; and (3) treating initially arbitrary functions in the equations (such as E in the Vlasov equation) as separate forms, rather than as explicit functions, may in some cases eliminate interesting solutions.

III. AN ALTERNATE APPROACH

This section develops an "intermediate" approach to solving the Vlasov–Maxwell equations. It differs from Roberts'[5] method in that only the Vlasov equation is required to be invariant, not the entire Vlasov–Maxwell set. On the other hand, unlike the Abraham-Shrauner method presented above, it uses the Vlasov equation rather than the character-

 0022-2488/85/071529-05$02.50

istic equations, thus allowing f to be transformed, and treats E as an independent variable, rather than as an initially arbitrary function of x and t.

The approach to finding the point group of the Vlasov equation that is suggested by following the methods of Bluman and Cole[8] consists of treating all dependent and independent variables as forms that are on an essentially equal footing, leading to a group generator given by

$$U=\xi_t\frac{\partial}{\partial t}+\xi_x\frac{\partial}{\partial x}+\xi_v\frac{\partial}{\partial v}+\eta^E\frac{\partial}{\partial E}+\sum_\alpha\eta^\alpha\frac{\partial}{\partial f_\alpha}. \tag{3}$$

A self-consistency condition (termed an "invariant surface condition" in Bluman and Cole) exists for the variation in the dependent variables; their transformation according the above generator must agree with their transformation by varying the independent variables in explicit solutions, giving, for example,

$$\xi_t\frac{\partial E}{\partial t}+\xi_x\frac{\partial E}{\partial x}=\eta^E. \tag{4}$$

The characteristic equations for this condition and the corresponding one for f generate the general similarity forms for the dependent variables, as well as new "natural" or "canonical" forms for the independent variables.

Associated with the above generator are generators for the derivatives $\partial f_\alpha/\partial t$, etc., which have coordinates given by

$$\eta_j^i=\frac{\partial\eta^i}{\partial x_j}+\sum_\mu\frac{\partial\eta^i}{\partial u_\mu}\frac{\partial u_\mu}{\partial x_j}-\sum_\nu\frac{\partial\xi_\nu}{\partial x_j}\frac{\partial u_i}{\partial x_\nu}-\sum_{\mu,\nu}\frac{\partial\xi_\nu}{\partial u_\mu}\frac{\partial u_\mu}{\partial x_j}\frac{\partial u_i}{\partial x_\nu}, \tag{5}$$

where the dependent and independent variables are denoted generically by u_i and x_j, respectively, and the sums are over all possibilities. An invariance group of the Vlasov equation results from finding a general form for U. This can be done by applying the "extended group generator" U', with

$$U'=U+\eta_j^i\frac{\partial}{\partial(\partial u_i/\partial x_j)}, \tag{6}$$

to the equation, and requiring this to result in at most the same equation multiplied by a function ρ of t, x, v, E, and f, but not of $\partial f/\partial t$, etc. The restriction on the form of ρ leads in general to a restriction on the generality of the group; in particular, it restricts the possibility of generating terms, through the last term in (5), that give the original equation multiplied by a derivative. This case only arises if the ξ_v depend on the dependent variables, and this possibility, which virtually never occurs for physically relevant partial differential equations,[8] will be neglected in the present development. Note that all other terms besides the last one in (5) are linear in the derivatives $\partial u_i/\partial x_j$, thus justifying the assumption that U' applied to the equation will at most generate a linear function of the original equation when ξ_v is independent of u_j.

Explicitly, the invariance condition just stated becomes, for the Vlasov equation (2),

$$\eta_t^\alpha+v\eta^\alpha{}_x+\xi_v\frac{\partial f_\alpha}{\partial x}+\frac{q_\alpha}{m_\alpha}\eta^E\frac{\partial f_\alpha}{\partial v}+\frac{q_\alpha}{m_\alpha}E\eta^\alpha{}_v$$

$$=\rho\left(\frac{\partial f_\alpha}{\partial t}+v\frac{\partial f_\alpha}{\partial x}+\frac{q_\alpha}{m_\alpha}E\frac{\partial f_\alpha}{\partial v}\right). \tag{7}$$

[In principle it is possible for η^α in the multispecies case to depend on $f_\beta(\beta\neq\alpha)$, thus requiring a sum on the right-hand side, but this turns out to be possible only for the artificial case where $q_\alpha/m_\alpha=q_\beta/m_\beta$, so this possibility is neglected here.] The coefficients of the various forms ($\partial f/\partial t$, etc.) in (7) must be equated individually for the equation to be satisfied, since (7) is a statement on the equivalence of forms, and not a differential equation. A first result of this is that η^α is independent of E, since the second term in (5) generates unique terms such as that involving $\partial E/\partial t$.

The remaining constraints in (7) come from equating the coefficients of 1, $\partial f/\partial t$, $\partial f/\partial x$, and $\partial f/\partial v$ on both sides of the equation. The results are, respectively,

$$\frac{\partial\eta^\alpha}{\partial t}+\frac{\partial\eta^\alpha}{\partial x}+\frac{q_\alpha}{m_\alpha}E\frac{\partial\eta^\alpha}{\partial v}=0, \tag{8}$$

$$\frac{\partial\xi_t}{\partial t}+v\frac{\partial\xi_t}{\partial x}+\frac{q_\alpha}{m_\alpha}E\frac{\partial\xi_t}{\partial v}=-\rho', \tag{9}$$

$$\frac{\partial\xi_x}{\partial t}+v\frac{\partial\xi_x}{\partial x}+\frac{q_\alpha}{m_\alpha}E\frac{\partial\xi_x}{\partial v}-\xi_v=-\rho' v, \tag{10}$$

and

$$\frac{\partial\xi_v}{\partial t}+v\frac{\partial\xi_v}{\partial x}+\frac{q_\alpha}{m_\alpha}E\frac{\partial\xi_v}{\partial v}-\frac{q_\alpha}{m_\alpha}\eta^E=-\rho'\frac{q_\alpha}{m_\alpha}E, \tag{11}$$

where $\rho'=\rho-\partial\eta^\alpha/\partial f_\alpha$.

Since η^α is independent of E, (8) shows that it is also independent of v; still using (8) this latter fact shows that η^α is independent of x, and subsequently of t. Thus the form for η^α is simply

$$\eta^\alpha=F(f_\alpha), \tag{12}$$

where F is an arbitrary differentiable function. To satisfy (9), ρ must then be chosen so that ρ' is independent of f_α.

To proceed further, we note that since E is required to be independent of v, it should be independent of v under an infinitesimal transformation. Using this requirement and (4) shows that ξ_t, ξ_x and η^E are all independent of v. Given this, (9) shows that ρ' is independent of E, and thus, from (11), that η^E is at most linear in E. Also from (9) we see that ρ' is at most linear in v, so from (10), ξ_v is at most quadratic in v. Thus the forms of ρ' and the coordinates of the generator are, at this point,

$$\begin{aligned}&\xi_t=\xi_t(t,x),\quad \xi_x=\xi_x(t,x),\\&\xi_v=v^2g_2(t,x)+vg_1(t,x)+g_0(t,x),\\&\eta^E=Eh_1(t,x)+h_0(t,x),\quad \eta^\alpha=c_1f_\alpha,\\&\rho'=v\rho_1(t,x)+\rho_0(t,x).\end{aligned} \tag{13}$$

Putting these forms into (9)–(11) and equating coefficients of E and powers of v gives a set of equations involving the new functions. These equations are, from (9),

$$\frac{\partial\xi_t}{\partial t}=-\rho_0 \tag{14}$$

and

$$\frac{\partial \xi_t}{\partial x} = -\rho_1; \tag{15}$$

from (10),

$$\frac{\partial \xi_x}{\partial t} - g_0 = 0, \tag{16}$$

$$\frac{\partial \xi_x}{\partial x} - g_1 = -\rho_0, \tag{17}$$

and

$$g_2 = \rho_1; \tag{18}$$

and from (11),

$$\frac{\partial g_0}{\partial t} - \frac{q_\alpha}{m_\alpha} h_0 = 0, \tag{19}$$

$$\frac{\partial g_1}{\partial t} + \frac{\partial g_0}{\partial x} = 0, \tag{20}$$

$$\frac{\partial g_2}{\partial t} + \frac{\partial g_1}{\partial x} = 0, \tag{21}$$

$$\frac{\partial g_2}{\partial x} = 0, \tag{22}$$

$$g_1 - h_1 = -\rho_0, \tag{23}$$

and

$$2g_2 = -\rho_1. \tag{24}$$

From (18) and (24) we see that $g_2 = \rho_1 = 0$, which implies, using (15), that ξ_t is independent of x, and using (21) that g_1 is independent of x. Solving (20) for g_0 gives

$$g_0 = x \frac{dg_1}{dt} + g_3(t), \tag{25}$$

where g_3 is an arbitrary function of time. Now, using this and (16) gives

$$\frac{\partial^2 \xi_x}{\partial x\, \partial t} = -\frac{dg_1}{dt}, \tag{26}$$

while (17) and (14) give

$$\frac{\partial^2 \xi_x}{\partial x\, \partial t} = \frac{dg_1}{dt} + \frac{d^2 \xi_t}{dt^2}. \tag{27}$$

Combining (26) and (27) and integrating with respect to t gives

$$g_1 = -\frac{1}{2}\frac{d\xi_t}{dt} + c_2, \tag{28}$$

where c_2 is a constant. Integrating (17) now yields

$$\xi_x = \left(\frac{1}{2}\frac{d\xi_t}{dt} + c_2\right)x + k(t), \tag{29}$$

where k is another arbitrary function of t.

At this point, all remaining unknowns can be determined in terms of ξ_t, k, and c_2. First, from (23)

$$h_1 = -\frac{3}{2}\frac{d\xi_t}{dt} + c_2. \tag{30}$$

Next, combining (16), (29), (25), and (28) gives

$$g_3 = \frac{dk}{dt} + x\frac{d^2\xi_t}{dt^2}. \tag{31}$$

This combined with (19), (28), and (25) gives

$$h_0 = \frac{m_\alpha}{q_\alpha}\left(\frac{x}{2}\frac{d^3\xi_t}{dt^3} + \frac{d^2 k}{dt^2}\right). \tag{32}$$

Collecting everything and recalling the forms (13), we have

$$\begin{aligned}
&\xi_t = \xi_t(t), \qquad \xi_x = \left(\frac{1}{2}\frac{d\xi_t}{dt} + c_2\right)x + k(t),\\
&\xi_v = \left(-\frac{1}{2}\frac{d\xi_t}{dt} + c_2\right)v + \frac{x}{2}\left(\frac{d^2\xi_t}{dt^2}\right) + \frac{dk}{dt},\\
&\eta^E = \left(-\frac{3}{2}\frac{d\xi_t}{dt} + c_2\right)E + \frac{m_\alpha}{q_\alpha}\left(\frac{x}{2}\frac{d^3\xi_t}{dt^3} + \frac{d^2k}{dt^2}\right),\\
&\eta^\alpha = F(f_\alpha).
\end{aligned} \tag{33}$$

These expressions give the general form of the generator (3) for the Lie group that leaves the Vlasov equation invariant.

The results for ξ_t and ξ_x in (33) are exactly the same as those in Abraham-Shrauner's work,[2] except that in the present development, there is no possibility of an ax term appearing in ξ_t. The origin of this is that making E a separate dependent variable imposes more constraints; in (11), for instance, if E is considered to be an explicit function of x and t, we would not be able to equate its coefficients. The solutions neglected here are not of great interest; the form required for E by them is completely determined by the group parameters, leaving little freedom to satisfy Maxwell's equations. The other solutions involve an arbitrary function [$\tilde{E}$ in (41) below] that allows nontrivial solutions. This shows, however, that in solving a single equation by Lie group methods, making free functions into separate dependent variables might possibly eliminate interesting solutions.

The first benefit in using the forms (33) is seen by considering the expression for η^E; the distinction between the multispecies and single species cases becomes immediately apparent. Since E cannot depend on species, the m_α/q_α term in η^E must be zero for the multispecies case, except again in the artificial case of two different species of equal q/m ratios. This means that the multispecies case corresponds to ξ_t at most quadratic in t and k at most linear in t. Abraham-Shrauner[3] arrives at similar conclusions, but more indirectly.

The second advantage of the present approach is that it allows us to reduce the forms of E and f to quadrature, with no auxiliary differential equations except for the Vlasov equation reduced by one dimension. To find the form of E, we return to (4), using the forms found in (33). It is interesting to note that the equation now given by (4) is also stated by Abraham-Shrauner,[2] but the present approach more clearly suggests the present method of solution. (The method of solution used there was motivated by comparison to solutions found by Lewis and Symon.[7]) The form for f will be derived using the analog of (4) given by

$$\xi_t\frac{\partial f}{\partial t} + \xi_x\frac{\partial f_\alpha}{\partial x} + \xi_v\frac{\partial f_\alpha}{\partial v} = \eta^\alpha. \tag{34}$$

This equation is new, the difference between this and the Abraham-Shrauner case outlined above being that in the latter development η^α is assumed implicitly to be zero.

To simplify the notation, let $\xi_x = k_1 x + k, \xi_v = g_1 v + g_4 x + g_3$, and $\eta^E = h_1 E + h_4 x + h_3$, where (33) shows that the functions $k_1, k, g_1, g_3, g_4, h_1, h_4$, and h_3 depend only on time. Then the characteristic equations for (4) and (34) are the following (with F linear for an integrable example):

$$\frac{dt}{\xi_t} = \frac{dx}{k_1 x + k} = \frac{dE}{h_1 E + h_4 x + h_3} \tag{35}$$

and

$$\frac{dt}{\xi_t} = \frac{dx}{k_1 x + k} = \frac{dv}{g_1 v + g_4 x + g_3} = \frac{df_\alpha}{c_1 f_\alpha}. \tag{36}$$

The solutions to pairs of these equalities that involve only independent variables yield new "natural" or "canonical" variables. With the forms for E and f found using the remaining equalities, the Vlasov equation will be reduced from a three-dimensional equation to a two-dimensional one.

The first equality in the above equations may be written

$$\frac{dx}{dt} - \frac{k_1}{\xi_t} x = \frac{k}{\xi_t}, \tag{37}$$

which has an integrating factor I_1, given by

$$I_1 = \exp\left(-\int \frac{k_1}{\xi_t} dt\right), \tag{38}$$

so the first natural variable ζ_1 can be taken to be the integration constant in the solution to (37)

$$\zeta_1 = x I_1 - \int I_1 \frac{k}{\xi_t} dt. \tag{39}$$

The form for E can then be found by equating the first and third terms in (35). The solution of the resulting equation requires a second integrating factor I_2 given by

$$I_2 = \exp\left(-\int \frac{h_1}{\xi_t} dt\right). \tag{40}$$

Calling the integration constant $\widetilde{E}(\zeta_1)$, the form for E is

$$E = \widetilde{E} I_2^{-1} + I_2^{-1} \int I_2 \frac{h_4 x + h_3}{\xi_t} dt, \tag{41}$$

where in performing the last integration, we must use (39) to give

$$x = \zeta_1 I_1^{-1} + I_1^{-1} \int I_1 \frac{k}{\xi_t} dt, \tag{42}$$

and ζ_1 is held fixed for the integration.

A similar procedure, based on (34), may be used to find the second natural variable ζ_2 and the form for f. The results of all these calculations, simplified somewhat by performing some integrations, are as follows:

$$\zeta_1 = x \xi_t^{-1/2} C^{-1} - \int \xi_t^{-3/2} C^{-1} k \, dt, \tag{43}$$

$$\begin{aligned} \zeta_2 = {} & v \xi_t^{1/2} C^{-1} - x \xi_t^{-1/2} C^{-1} \int \frac{1}{2} \frac{d^2 \xi_t}{dt^2} dt \\ & + \left(\int \xi_t^{-3/2} C^{-1} k \, dt\right) \int \frac{1}{2} \frac{d^2 \xi_t}{dt^2} dt - \frac{1}{2} \int \frac{d^2 \xi_t}{dt^2} \\ & \times \left(\int \xi_t^{-3/2} C^{-1} k \, dt\right) dt - \int \xi_t^{-1/2} C^{-1} \frac{dk}{dt} dt, \end{aligned} \tag{44}$$

$$\begin{aligned} E = {} & \widetilde{E}(\zeta_1) \xi_t^{-3/2} C + \frac{m}{q} x \xi_t^{-2} \int \frac{1}{2} \frac{d^3 \xi_t}{dt^3} \xi_t \, dt \\ & + \frac{m}{q} \xi_t^{-3/2} C \left[\int \xi_t^{1/2} \frac{d^2 k}{dt^2} C^{-1} dt \right. \\ & \left. - \int \xi_t^{-3/2} C^{-1} k \left(\int \frac{1}{2} \frac{d^3 \xi_t}{dt^3} \xi_t dt \right) dt \right], \end{aligned} \tag{45}$$

and

$$f_\alpha = \tilde{f}_\alpha(\zeta_1, \zeta_2) \exp\left(\int \frac{C_1}{\xi_t} dt\right), \tag{46}$$

where C is given by

$$C = \exp\left(\int \frac{c_2}{\xi_t} dt\right), \tag{47}$$

and $\tilde{f}$ is a new initially arbitrary function. Substituting these variables into the Vlasov equation leads, after considerable algebra and many integrations by parts, to the "reduced" equation

$$c_1 \tilde{f} + (\zeta_2 - c_2 \zeta_1) \frac{\partial \tilde{f}}{\partial \zeta_1} + \left(-c_2 \zeta_2 + \frac{q \widetilde{E}}{m}\right) \frac{\partial \tilde{f}}{\partial \zeta_2} = 0. \tag{48}$$

Comparing these results to those of Abraham-Shrauner,[2] we see that, apart from some relabeling, ζ_1, ζ_2 and the reduced Vlasov equation are exactly the same in both cases, and E has the same general form. The difference in the latter case is that here E has been reduced to quadrature, whereas Abraham-Shrauner introduces a function ρ (unrelated to ρ above) by

$$\xi_t = \rho^2, \tag{49}$$

and gives the form of E as

$$E = F(t) - \Omega^2(t) x - \frac{1}{\rho^3} \exp\left(\int \frac{c_2}{\rho^2} dt\right) \frac{dU_e(\zeta_1)}{d\zeta_1}, \tag{50}$$

where U_e is a nonquadratic potential, $\Omega(t)$ satisfies the equation

$$\ddot{\rho} + (q/m) \Omega^2 = k'/\rho^3, \tag{51}$$

and $F(t)$ also satisfies an auxiliary differential equation. (Here, k' is a constant, the notation has been changed somewhat for easy comparison, and species subscripts have been dropped for simplicity.) Explicitly comparing forms, noting that the linear part $(k'\zeta_1)$ of E must be taken out to isolate linear terms in x, we find

$$\Omega^2(t) = -\xi_t^{-2} \int \xi_t \left(\frac{1}{2} \frac{m}{q} \frac{d^3 \xi_t}{dt^3}\right) dt - \frac{mk'}{q} \xi_t^{-2}. \tag{52}$$

This form satisfies (51), as it should. Further detailed comparisons can be made, but these are not enlightening. Some of the utility of the quadrature solution for E is illustrated by the development in the next section, which is more lengthy without it (cf. Abraham-Shrauner, Ref. 3).

IV. THE EFFECT OF CONSTRAINING EQUATIONS AND A GENERALIZATION TO OTHER CASES

Maxwell's equations must be solved along with the Vlasov equation, and this puts further constraints on the solutions. Poisson's equation with a background of density n_b and charge per particle q_b,

$$\frac{\partial E}{\partial x} - 4\pi q_b n_b - 4\pi q \int f\, dv = 0, \tag{53}$$

becomes, in the new variables,

$$\frac{\partial \widetilde{E}}{\partial \zeta_1} + \frac{m}{q} \int \frac{1}{2} \frac{d^3 \xi_t}{dt^3} \xi_t \, dt - 4\pi q_b n_b \xi_t^2 - 4\pi q \xi_t^{3/2} \exp\left(\int \frac{c_1 + c_2}{\xi_t} dt \right) \tilde{f}\, d\zeta_2 = 0. \tag{54}$$

This can be solved as is, without further restrictions, for n_b, but in a case with multiple species and no background, a complete reduction to the new variables requires (apart from an uninteresting multiplicative constant)

$$c_1 + c_2 = -\frac{3}{2} \frac{d\xi_t}{dt}. \tag{55}$$

This makes ξ_t linear in t, and relates c_1 and c_2. The results now become equivalent to those found by looking for the invariance group of the one-dimensional Vlasov–Maxwell equations considered simultaneously (Roberts, Ref. 5.).

The reduction of the group for the Vlasov equation alone to that for the Vlasov–Maxwell equations through the application of Maxwell's equations as constraints is a special case of a more general theorem: The reduction of a set of equations to a set involving one fewer independent variables by the method of finding the Lie invariance group of one equation in the set, and using the resulting canonical variables to reduce the other equations, will never result in anything more general than is found using the canonical variables found from the Lie invariance of the complete set to perform a similar reduction. That this is true can be seen as follows.

Suppose we have two equations involving the same dependent and independent variables $u_1, \ldots, u_n$ and $x_1, \ldots, x_m$. Finding the most general one-parameter point group of the first equation allows us to find canonical independent variables $\zeta_1, \ldots, \zeta_m$ and general similarity forms for the dependent variables such that one independent variable (say, ζ_m) does not appear in the equation in terms of the new variables. (Note that the one-parameter group may involve arbitrary constants and functions in its only generator.) In terms of the new independent variables, the group of the first equation can always be made to be the translation group in ζ_m. (See Bluman and Cole.[8]) We now require that the substitution of the new variables in the second equation results in an equation involving only $\zeta_1, \ldots, \zeta_{m-1}$. In general, this restricts the forms allowed for the generator of the group of the first equation, by restricting the allowable forms of $\zeta_1, \ldots, \zeta_m$. On the other hand, the fact that the new second equation is independent of the restricted form of ζ_m means that it is invariant under translations in that variable. Since the latter group is a special case of the group of the first equation, both equations are invariant under translations in the restricted form of ζ_m. This means that the reduced form of both equations is at most as general as that found by considering the most general invariance group of both equations simultaneously. The procedure outlined above can be extended to more equations, and in fact should provide one (perhaps cumbersome) method for finding the general invariance group of a set of equations.

An essential part of the above reasoning is the assumption that we wish to put all the equations in terms of a reduced set of variables. Generally speaking, it is this requirement that simplifies the work in finding solutions. However, as is illustrated above, this will not always be the case; the introduction of the new variable n_b made it possible to find solutions to (53) corresponding to any solution of the Vlasov equation, thus eliminating the need to state (53) in terms of a reduced set of variables. In this case, a method based on the invariance of a single equation gives more general results than one involving the invariance of the set.

V. DISCUSSION

The development above shows some of the advantages and disadvantages of various approaches to finding Lie point group solutions to both the Vlasov–Maxwell and other sets of equations. It will sometimes be the case that solutions generated by any point group method will be limited in utility, in which case more general transformations (contact or Lie–Bäcklund, for example; see Anderson and Ibragimov[9]) may be useful. The disadvantage of the latter methods is that they are often much more involved and less systematic. Especially in the case of nonlinear systems of equations, where any exact solution is often a useful step, Lie point group methods may provide a good starting point, and if so, the above discussion should help in sorting out alternative approaches.

ACKNOWLEDGMENTS

I thank B. Abraham-Shrauner for useful discussions.

I also thank the National Science Foundation and NASA for financial support.

[1] I. B. Bernstein, J. M. Green, and M. D. Kruskal, Phys. Rev. **108**, 546 (1957).
[2] B. Abraham-Shrauner, J. Math. Phys. **26**, 1428 (1985).
[3] B. Abraham-Shrauner, J. Plasma Phys. **32**, 197 (1984).
[4] V. B. Baranov, Sov. Phys. Tech. Phys. **21**, 720 (1976).
[5] D. Roberts, "The general Lie group and similarity solutions for the one-dimensional Vlasov–Maxwell equations," J. Plasma Phys. (in press).
[6] A. Cohen, *An Introduction to the Lie Group Theory of One-Parameter Groups* (Heath, New York, 1911).
[7] H. R. Lewis and K. R. Symon, Phys. Fluids **27**, 192 (1984).
[8] G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer, Berlin, 1974).
[9] R. L. Anderson and N. H. Ibragimov, *Lie-Backlund Transformations in Applications* (SIAM, Philadelphia, 1979).

Polynomial tensors for the space groups pm and $p4m$

D. Phaneuf and R. T. Sharp
Physics Department, McGill University, Montreal, Quebec, H3A 2T8, Canada

(Received 29 November 1984; accepted for publication 8 February 1985)

Generating functions are calculated for polynomial tensors in the components of tensors of the one- and two-dimensional space groups pm and $p4m$. For tensors whose k vectors lie at rational points (denominator q) of the Brillouin zone, the problem is imaged by the corresponding problem for an appropriate finite group, and known methods are used for its solution. For tensors with continuous k the solution is obtained by letting q become infinite.

I. INTRODUCTION

Polynomial space group tensors are needed for renormalization group calculations and for applications of Landau theory to phase transitions in crystals.[1] Even when only space group invariants are required, these are often based on reducible representations; then it is simplest to construct all space group tensors based on the irreducible components and combine them to make composite scalars.

A great deal of effort has gone into hand or computer calculation of polynomial space group tensors of low degree.[2] This paper proposes analytic methods for finding such tensors of all degrees.

By an (l,m) tensor of a group G we mean one whose components transform by the representational l of G and are homogeneous polynomials in the components of a tensor transforming by the representation m. The problem of determining (l,m) tensors has been solved for point groups,[3,4] and the method is applicable to any finite group; as we shall show, it may equally well be used to solve the same problem for space groups.

The first step in enumerating (l,m) tensors of a finite group G is to calculate[3,4] the generating function

$$B_{lm}(\lambda) = N^{-1} \sum_c \frac{N_c \chi^*_{cl}}{\det(1-\lambda A_{cm})} = \sum_{n=0}^{\infty} C^{(n)}_{lm} \lambda^n \tag{1.1}$$

for the pair (l,m) of irreducible representations (IR's). In (1.1), N is the order of G, N_c is the number of elements in the class c, χ_{cl} is the character of the class c for the IR l, and A_{cm} is the matrix that represents an element of the class c in the IR m. The coefficient $C^{(n)}_{lm}$ in the power series expansion of $B_{lm}(\lambda)$ is the number of linearly independent (l,m) tensors of degree n. It turns out that $B_{lm}(\lambda)$ is a rational function

$$B_{lm}(\lambda) = \sum_u \frac{h^{(u)}_{lm}\lambda^u}{\Pi_t(1-\lambda^t)}; \tag{1.2}$$

the sum over u and the product over t are both finite. The coefficients $h^{(u)}_{lm}$ are non-negative integers; the denominator factors are equal in number to f_m, the dimension of the IR m, and are the same for all generating functions with the same m. The denominator factors correspond to functionally independent scalars of degrees t. The numerator terms correspond to (l,m) tensors, of degrees u, which are linearly independent when their coefficients belong to the ring of denominator scalars. The properties of the generating functions $B_{lm}(\lambda)$ are discussed more fully in Refs. 3 and 4. We note that a generating function for which m is reducible can be expressed in terms of those for which m is irreducible. The generating functions with fixed m satisfy a dimensionality condition[4]

$$\sum_l f_l B_{lm}(\lambda) = (1-\lambda)^{-f_m}, \tag{1.3}$$

which provides a useful check on their correctness. The l sum in (1.3) is over all IR's.

The (unitary) IR's of a space group G are labeled $(\mathbf{k})_s$. The vector $\mathbf{k}$ lies in or on the boundary of a sector of reciprocal space comprising $1/g$ of the Brillouin zone, where g is the order of the point group of G; the integer s takes a finite number of values 1,2,..., and distinguishes inequivalent IR's belonging to $\mathbf{k}$. We write $\mathbf{k}$ as a linear combination of the primitive reciprocal lattice vectors $\mathbf{K}_i$ and suppose the coefficients are rational fractions

$$\mathbf{k} = q^{-1} \sum_i p_i \mathbf{K}_i . \tag{1.4}$$

(The $\mathbf{K}_i$ are defined in terms of the primitive lattice translations $\mathbf{a}_j$ by $\mathbf{K}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij}$; the Brillouin zone is the region of $\mathbf{k}$ space closer to the origin than to any other integer linear combination of $\mathbf{K}_i$.)

For fixed denominator q there is a finite number of IR's; direct products of IR's with the same q, of which polynomial tensors are a special case, can lead only to tensors with $\mathbf{k}=0$, or with $q' = q$ (or a factor of q); we are dealing with a finite image, or matrix, group which is in fact the quotient of our space group G by the same group with the translations increased by a factor q. The continuous $\mathbf{k}$ case is obtained by letting $q \to \infty$.

The matrices which represent the group operations are obtained by a method due to Raghavachacharyulu.[5,6] Starting with a one-dimensional IR of the translation group, one induces IR's of higher groups until one arrives at the group of $\mathbf{k}$, and, in a final step, at G.

The approach (but not the results) of this paper is described in Ref. 7.

II. THE ONE-DIMENSIONAL GROUP pm

The one-dimensional space group pm (see Ref. 8) consists of pure translations $(\epsilon|n)$ and reflection in the origin followed by translation $(P|n)$. The translations ($n = 0, \pm 1, \pm 2,...$) are measured in units of a, the fundamental lattice displacement; we measure k in units of $2\pi/a$.

 0022-2488/85/071534-06$02.50

The order of the point group is 2; hence k, as an IR label, ranges over half the Brillouin zone (BZ), $0 \leqslant k \leqslant \frac{1}{2}$.

It turns out that at each of the two points of symmetry $k = 0, \frac{1}{2}$, there are two one-dimensional IR's of $\wp m$, while at an interior point, $0 < k < \frac{1}{2}$, there is one two-dimensional IR. The representation matrices are given in Table I. The characters (traces) are easily read from the table.

It is easy to see, for given q, which operators are distinct, which of the distinct ones belong in the same class, and which representations are those of the corresponding image group. For $k = 0$ the translations have no effect and the image group is the point group with the representations $(0)_1$, $(0)_2$. For $q = 2$ four operators can be distinguished, $(\epsilon|e)$, $(\epsilon|o)$, $(P|e)$, $(P|o)$, where $e(o)$ signifies that n is even (odd). The image group has representations $(0)_1$, $(0)_2$, $(\frac{1}{2})_1$, $(\frac{1}{2})_2$. For $q > 2$, n may be restricted to the range $0 \leqslant n \leqslant q-1$. The representations of the image group are $(0)_1$, $(0)_2$, (p/q), $p = 1,2,\ldots,[\frac{1}{2}(q-1)]$ and, in addition, if q is even, $(\frac{1}{2})_1$, $(\frac{1}{2})_2$; $[a]$ means the integer part of a.

Formula (1.1) could now be used to construct generating functions for polynomial tensors; however, we can save work by recognizing that the image group of q is the point group D_q, the symmetry group of a regular q-sided polygon, including rotations π about the symmetry axes in the plane of the polygon. The primitive translation a corresponds to the rotation $2\pi/q$ and a reflection translation is imaged by one of the rotations π about an axis in the plane of the polygon.

The desired generating functions are given in Ref. 4. To transcribe them into our present notation we must make the identifications $n \to q$, $\Gamma_1 \to (0)_1$, $\Gamma_2 \to (0)_2$; for q odd, there is also $\Gamma_{p+2} \to (p/q)$, $1 \leqslant p \leqslant (q-1)/2$, and for q even, $\Gamma_3 \to (\frac{1}{2})_1$, $\Gamma_4 \to (\frac{1}{2})_2$, $\Gamma_{p+4} \to (p/q)$, $1 \leqslant p \leqslant q/2 - 1$. We find, for all $q \geqslant 2$,

$$
\begin{aligned}
&B_{(0)_1(0)_1} = (1-\lambda)^{-1}, \quad B_{(0)_1(0)_2} = (1-\lambda^2)^{-1},\\
&B_{(0)_2(0)_2} = \lambda(1-\lambda^2)^{-1},\\
&B_{(0)_1(p/q)} = [(1-\lambda^2)(1-\lambda^q)]^{-1},\\
&B_{(0)_2(p/q)} = \lambda^q[(1-\lambda^2)(1-\lambda^q)]^{-1},\\
&B_{(p'/q,\,p/q)} = (\lambda^r + \lambda^{q-r})[(1-\lambda^2)(1-\lambda^q)]^{-1},
\end{aligned} \tag{2.1a}
$$

and, in addition, for q even,

$$
\begin{aligned}
&B_{(0)_1(1/2)_1} = B_{(0)_1(1/2)_2} = (1-\lambda^2)^{-1},\\
&B_{(1/2)_1(1/2)_1} = B_{(1/2)_2(1/2)_2} = \lambda(1-\lambda^2)^{-1},\\
&B_{(1/2)_1(p/q)} = B_{(1/2)_2(p/q)} = \lambda^{q/2}[(1-\lambda^2)(1-\lambda^q)]^{-1}.
\end{aligned} \tag{2.1b}
$$

TABLE I. Representation matrices for the one-dimensional space group $\wp m$; $q = 3,4,\ldots$; $p = 1,2,\ldots,[(q-1)/2]$; $A = \exp(-2\pi i n p/q)$; $k = p/q$ could also be irrational, $0 < k < \frac{1}{2}$.

Op \ IR	$(0)_1$	$(0)_2$	$(\frac{1}{2})_1$	$(\frac{1}{2})_2$	(p/q)
$(\epsilon\|n)$	1	1	$(-1)^n$	$(-1)^n$	$\begin{pmatrix} A & 0 \\ 0 & A^{-1} \end{pmatrix}$
$(P\|n)$	1	-1	$(-1)^n$	$-(-1)^n$	$\begin{pmatrix} 0 & A \\ A^{-1} & 0 \end{pmatrix}$

It is understood in (2.1) that p (but not necessarily p') is prime relative to q and that $1 \leqslant p \leqslant [(q-1)/2]$; p' is given in terms of r, p, q by $p' = q/2 - |(rp)_{\mathrm{mod}\, q} - q/2|$; $r = 1,2,\ldots,[(q-1)/2]$.

To obtain the irrational k case we let $q \to \infty$ and set $\lambda^q = 0$ (since we may suppose $|\lambda| < 1$). Then the last three equations of (2.1a) are replaced by

$$
\begin{aligned}
&B_{(0)_1(k)} = (1-\lambda^2)^{-1},\\
&B_{(k')(k)} = \lambda^r(1-\lambda^2)^{-1}, \quad k' = \tfrac{1}{2} - |rk - [rk] - \tfrac{1}{2}|,\\
&r = 1,2,3\ldots, \quad 0 < k < \tfrac{1}{2}, \quad k \text{ irrational}.
\end{aligned} \tag{2.2}
$$

In (2.1) and (2.2) generating functions which vanish have been omitted.

To see that nothing is lost when k is irrational we use the dimensionality check (1.3). It reads

$$
(1-\lambda^2)^{-1}\left[1 + 2\sum_{r=1}^{\infty} \lambda^r\right] = (1-\lambda)^{-2}, \tag{2.3}
$$

as required.

The problem of polynomial tensors for the other one-dimensional space group $\wp 1$ (without reflections) is imaged in a similar way by the point group C_q.

III. THE TWO-DIMENSIONAL GROUP $p4m$

The space group $p4m$ (see Ref. 8) is the symmetry group of a square lattice, including reflection symmetries. Its operations are pure translations $(\epsilon|n_x n_y)$, rotations followed by translations $(\pi/2|n_x n_y)$, $(\pi|n_x n_z)$, $(3\pi/2|n_x n_y)$, and reflections with translations $(P_x|n_x n_y)$, $(P_y|n_x n_y)$, $(P_{\pi/4}|n_x n_y)$, $(P_{3\pi/4}|n_x n_y)$. The rotations are about the origin through angles $\pi/2$, π, $3\pi/2$ and the reflections are in lines through the origin making angles 0, $\pi/2$, $\pi/4$, $3\pi/4$ with the x axis; the translations n_x, n_y are measured in units of the primitive translation a. The Brillouin zone is the square $-\frac{1}{2} < k_x \leqslant \frac{1}{2}$, $-\frac{1}{2} < k_y \leqslant \frac{1}{2}$; k_x, k_y are measured in units $2\pi/a$. The order of the point group is 8 and the range for k_x, k_y as representation labels is $0 \leqslant k_y \leqslant k_x \leqslant \frac{1}{2}$, one eighth of the BZ.

The representation matrices for given $\mathbf{k}$ are found by following the algorithms of Refs. 5 and 6. It turns out that at $\mathbf{k} = 0$ and at $\mathbf{k} = (\frac{1}{2}, \frac{1}{2})$, there are four one-dimensional and one two-dimensional IR's; at $\mathbf{k} = (\frac{1}{2},0)$, there are four two-dimensional IR's; at $\mathbf{k} = (k,0)$, $(\frac{1}{2},k)$, and (k,k), there are two four-dimensional IR's; and at an interior generic point $0 < k_y < k_x < \frac{1}{2}$, there is one eight-dimensional IR.

The image group for $\mathbf{k} = 0$ is the point group of $p4m$ and is isomorphic to the point group D_4; when representations at $\mathbf{k} = (\frac{1}{2}, \frac{1}{2})$ are included the image group is isomorphic to $D_4 \times P$ (P means inversion). When representations at $\mathbf{k} = (\frac{1}{2},0)$ are added we have the image group of $q = 2$. It has 32 ($8q^2$) elements, and 14 each of classes and IR's.

According to (1.1) we need the eigenvalues of one matrix from each class for each IR in order to construct the generating functions for polynomial tensors [$\det(1 - \lambda A_{cm}) = \Pi_{i=1}^{f_m}(1 - \lambda s_{cm}^{(i)})$, where the $s_{cm}^{(i)}$ are the eigenvalues for class c, IR m]. Their eigenvalues are given in Table II. The character may also be read from the table—it is just the sum of the eigenvalues.

TABLE II. Eigenvalues of representation matrices of $p4m$ at the symmetry points $(k_x, k_y) = (0,0), (\frac{1}{2},\frac{1}{2}), (\frac{1}{2},0)$; a superscript indicates the eigenvalue's multiplicity.

order		1	2	1	1	2	1	4	4	2	2	2	2	4	4
IR \ class		1	2	3	4	5	6	7	8	9	10	11	12	13	14
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
(00)	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	5	1^2	1^2	1^2	1^2	1^2	1^2	$\pm i$	$\pm i$	$\pm i$	$\pm i$	$\pm i$	$\pm i$	$\pm i$	$\pm i$
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$(\frac{1}{2}\frac{1}{2})$	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	5	1^2	1^2	1^2	1^2	1^2	1^2	$\pm i$	$\pm i$	$\pm i$	$\pm i$	$\pm i$	$\pm i$	$\pm i$	$\pm i$
	1	1^2	± 1	1^2	1^2	± 1	1^2	± 1	$\pm i$	1^2	± 1	± 1	1^2	± 1	$\pm i$
$(\frac{1}{2}0)$	2	1^2	± 1	1^2	1^2	± 1	1^2	± 1	$\pm i$	1^2	± 1	± 1	1^2	± 1	$\pm i$
	3	1^2	± 1	1^2	1^2	± 1	1^2	$\pm i$	± 1	± 1	1^2	1^2	± 1	± 1	$\pm i$
	4	1^2	± 1	1^2	1^2	± 1	1^2	$\pm i$	± 1	± 1	1^2	1^2	± 1	± 1	$\pm i$

The translations need be specified only modulo 2; e (or o) means that n_x (or n_y) is even or odd. The classes turn out to be

$$\begin{aligned}
&1{:}\ (\epsilon|ee);\quad 2{:}\ (\epsilon|eo),(\epsilon|oe);\quad 3{:}\ (\epsilon|oo);\\
&4{:}\ (\pi|ee);\quad 5{:}\ (\pi|eo),(\pi|oe);\quad 6{:}\ (\pi|oo);\\
&7{:}\ (\pi/2|ee),\ (\pi/2|oo),\ (3\pi/2|ee),\ (3\pi/2|oo);\\
&8{:}\ (\pi/2|eo),\ (\pi/2|oe),\ (3\pi/2|eo),\ (3\pi/2|oe);\\
&9{:}\ (P_x|ee),\ (P_y|ee);\quad 10{:}\ (P_x|eo),\ (P_y|oe);\\
&11{:}\ (P_x|oe),\ (P_y|eo);\quad 12{:}\ (P_x|oo),\ (P_y|oo);\\
&13{:}\ (P_{\pi/4}|ee),\ (P_{\pi/4}|oo),\ (P_{3\pi/4}|ee),\ (P_{3\pi/4}|oo);\\
&14{:}\ (P_{\pi/4}|eo),\ (P_{\pi/4}|oe),\ (P_{3\pi/4}|eo),\ (P_{3\pi/4}|oe)\,.
\end{aligned} \tag{3.1}$$

It happens that there are just seven distinct generating functions $B_{lm}(\lambda)$ when m (and hence l) refer to representations belonging to $q=2$ $[k=(0,0), (\frac{1}{2},\frac{1}{2}), (\frac{1}{2},0)]$. We have $B_{lm}(\lambda) = (1-\lambda)^{-1}$, when l and m are both $(00)_1$; $B_{lm} = (1-\lambda^2)^{-1}$, when $l=(00)_1$ and $m=(00)_{2,3,4}$, $(\frac{1}{2}\frac{1}{2})_{1,2,3,4}$; $B_{mm} = \lambda(1-\lambda^2)^{-1}$, when $m=(00)_{2,3,4}$, $(\frac{1}{2}\frac{1}{2})_{1,2,3,4}$; $B_{lm} = [(1-\lambda^2)(1-\lambda^4)]^{-1}$, when $l=(00)_1$ and $m=(00)_5$, $(\frac{1}{2}\frac{1}{2})_5$, $(\frac{1}{2}0)_{1,2,3,4}$; $B_{lm} = \lambda^2[(1-\lambda^2)(1-\lambda^4)]^{-1}$, when $l=(00)_3$ and $m=(00)_5$, $(\frac{1}{2}\frac{1}{2})_5$, $(\frac{1}{2}0)_{1,2,3,4}$, when $l=(00)_4$ and $m=(00)_5$, $(\frac{1}{2}\frac{1}{2})_5$, when $l=(\frac{1}{2}\frac{1}{2})_1$ and $m=(\frac{1}{2}0)_{1,2}$, when $l=(\frac{1}{2}\frac{1}{2})_4$ and $m=(\frac{1}{2}0)_{3,4}$; $B_{lm} = \lambda^4[(1-\lambda^2)(1-\lambda^4)]^{-1}$, when $l=(00)_2$ and $m=(00)_5$, $(\frac{1}{2}\frac{1}{2})_5$, when $l=(\frac{1}{2}\frac{1}{2})_5$ and $m=(\frac{1}{2}0)_{3,4}$, when $l=(\frac{1}{2}\frac{1}{2})_3$ and $m=(\frac{1}{2}0)_{1,2}$; $B_{mm} = (\lambda+\lambda^3)\times[(1-\lambda^2)(1-\lambda^4)]^{-1}$, when $m=(00)_5$, $(\frac{1}{2}\frac{1}{2})_5$, $(\frac{1}{2}0)_{1,2,3,4}$; all other generating functions vanish.

We now turn to representations belonging to general $q \geqslant 3$. The distinct translations (n_x, n_y) are restricted by $0 \leqslant n_x, n_y \leqslant q-1$. The IR's, other than those at $\mathbf{k}=(0,0)$, and $(\frac{1}{2},\frac{1}{2})$, $(\frac{1}{2},0)$ for q even, are $(p/q\,0)_{1,2}$, $(p/q\,p/q)_{1,2}$, and $(\frac{1}{2}p/q)_{1,2}$ for q even, where $1 \leqslant p \leqslant [(q-1)/2]$, and $(p_x/q\,p_y/q)$, where $1 \leqslant p_y < p_x \leqslant [(q-1)/2]$. The eigenvalues of the matrices for these IR's are given in Table III.

The image group of $p4m$ for IR's with denominator q is the symmetry group of a four-dimensional figure consisting of two regular q-sided polygons centered at the origin, one lying in the 1-2 plane (1 a symmetry axis), the other oriented similarly in the 3-4 plane. Then the primitive x translation a is imaged by a rotation $2\pi/q$ of the 1-2 or x polygon. A rotation $\pi/2$ of the lattice is imaged by a reflection which

TABLE III. Eigenvalues of representation matrices of $p4m$. If two or more $\pm$ signs occur in the same expression they may be chosen independently. $\eta = e^{2\pi i p/q}$, $\xi = e^{2\pi i p_x/q}$, $\zeta = e^{2\pi i p_y/q}$. A superscript of 2 or 4 indicates the multiplicity of an eigenvalue. The eigenvalues of the representation $(\mathbf{k})_2$, where $\mathbf{k} = (p/q, 0)$, $(p/q, p/q)$, or $(\frac{1}{2}, p/q)$ are obtained from those of $(\mathbf{k})_1$ by reversing the sign of the entries in the last four columns. For irrational points, p/q, p_x/q, p_y/q may be replaced by the continuous variables k, k_x, k_y.

IR \ element	(ϵ/ab)	$(\pi/2\|ab)$ $(3\pi/2\|ab)$	$(\pi\|ab)$	$(P_x\|ab)$	$(P_y\|ab)$	$(P_{\pi/4}\|ab)$	$(P_{3\pi/4}\|ab)$
$(p/q\,0)_1$	$\eta^{\pm a}, \eta^{\pm b}$	$\pm 1, \pm i$	$(\pm 1)^2$	$\pm 1, \eta^{\pm b}$	$\pm 1, \eta^{\pm a}$	$\pm\eta^{\pm(1/2)(a+b)}$	$\pm\eta^{\pm(1/2)(a-b)}$
$(p/q\,p/q)_1$	$\eta^{\pm(a+b)}$, $\eta^{\pm(a-b)}$	$\pm 1, \pm i$	$(\pm 1)^2$	$\eta^{\pm b}$	$\pm\eta^{\pm a}$	± 1, $\eta^{\pm(a+b)}$	± 1, $\eta^{\pm(a-b)}$
$(\frac{1}{2}p/q)_1$	$(-1)^a\eta^{\pm b}$, $(-1)^b\eta^{\pm a}$	$\pm 1, \pm i$	$(\pm 1)^2$	± 1, $(-1)^a\eta^{\pm b}$	± 1, $(-1)^b\eta^{\pm a}$	$\pm i^{a+b}$, $\times\eta^{\pm(1/2)(a+b)}$	$\pm i^{a+b}$, $\times\eta^{\pm(1/2)(a-b)}$
$(p_x/q\,p_y/q)$	$\xi^{\pm a}\zeta^{\pm b}$, $\xi^{\pm b}\zeta^{\pm a}$	$(\pm 1)^2, (\pm i)^2$	$(\pm 1)^4$	$\pm\xi^{\pm b}$, $\pm\zeta^{\pm b}$	$\pm\xi^{\pm a}$, $\pm\zeta^{\pm a}$	$\pm\xi^{\pm(1/2)(a+b)}$ $\times\zeta^{\pm(1/2)(a+b)}$	$\pm\xi^{\pm(1/2)(a-b)}$ $\times\zeta^{\pm(1/2)(a-b)}$

TABLE IV. Generating function $B_{lm}(\lambda)$ with $m=(p/q\,0)_{1,2}$, $l=(00)_{1\text{–}5}$. The numerator is tabulated; the denominator, the same for all generating functions in the same column, is shown in the bottom row.

$l \backslash m$	$(p/q\,0)_2$, q even or $(p/q\,0)_1$	$(p/q\,0)_2$, q odd
$(00)_1$	$1+\lambda^{q+2}$	$1+\lambda^{2q+2}+\lambda^{3q}+\lambda^{3q+2}$
$(00)_2$	$\lambda^{2q+2}+\lambda^{3q}$	$\lambda^{q}\lambda^{q+2}+\lambda^{2q+2}+\lambda^{4q}$
$(00)_3$	$\lambda^{2}+\lambda^{q}$	$\lambda^{2}+\lambda^{2q}+\lambda^{3q}+\lambda^{3q+2}$
$(00)_4$	$\lambda^{2q}+\lambda^{3q+2}$	$\lambda^{q}+\lambda^{q+2}+\lambda^{2q}+\lambda^{4q+2}$
$(00)_5$	$(1+\lambda^2)(1+\lambda^q)\lambda^q$	$(1+\lambda^2)(1+\lambda^q)^2\lambda^q$
Denom	$(1-\lambda^2)(1-\lambda^4)(1-\lambda^q)(1-\lambda^{2q})$	$(1-\lambda^2)(1-\lambda^4)(1-\lambda^{2q})^2$

TABLE VI. Numerator of B_{lm}, $m=(p/q\,0)_1$, q even, $(p/q\,0)_2$, $q/2$ even, $l=(\frac{1}{2}0)_{1\text{–}4}$. For $(p/q\,0)_2$, $q/2$ odd, interchange the entries in the two rows; for q odd the generating functions vanish.

$l \backslash m$	$(p/q\,0)_1$, q even or $(p/q\,0)_2$, $q/2$ even
$(\frac{1}{2}\,0)_{1,4}$	$(1+\lambda^2)(1+\lambda^q)\lambda^{q/2}$
$(\frac{1}{2}\,0)_{2,3}$	$(1+\lambda^2)(1+\lambda^q)\lambda^{3q/2}$

TABLE V. Numerator of B_{lm}, $m=(p/q\,0)_{1,2}$ q even, $l=(\frac{1}{2}\frac{1}{2})_{1\text{–}5}$. For q odd the generating functions vanish.

$l \backslash m$	q even $(p/q\,0)_{1,2}$
$(\frac{1}{2}\frac{1}{2})_{1,4}$	$\lambda^{q}+\lambda^{2q+2}$
$(\frac{1}{2}\frac{1}{2})_{2,3}$	$\lambda^{q+2}+\lambda^{2q}$
$(\frac{1}{2}\frac{1}{2})_5$	$(1+\lambda^2)(1+\lambda^q)\lambda^q$

TABLE VII. Numerators of B_{lm}, $m=(p/q\,0)_1$, $l=(p'/q\,0)_{1,2}$.

$l \backslash m$	$(p/q\,0)_1$
$(p'/q\,0)_1$	$(1+\lambda^2)(1+\lambda^q)(\lambda^{r}+\lambda^{q-r})$
$(p'/q\,o)_2$	$(1+\lambda^2)(1+\lambda^q)(\lambda^{q+r}+\lambda^{2q-r})$

TABLE VIII. Numerators of B_{lm}, $m=(p/q\,0)_2$, $(p'/q\,0)_{1,2}$.

$l \backslash m$	$(p/q\,0)_2$, q even	$(p/q\,0)_2$, q odd
$(p'/q\,0)_1$, p' even or $(p/q\,0)_2$, p' odd	$(1+\lambda^2)(1+\lambda^q)(\lambda^{r}+\lambda^{q-r})$	$(1+\lambda^2)(1+\lambda^q)^3\lambda^{r}$
$(p'/q\,0)_1$, odd or $(p'/q\,0)_2$, p' even	$(1+\lambda^2)(1+\lambda^q)(\lambda^{q+r}+\lambda^{2q-r})$	$(1+\lambda^2)(1+\lambda^q)^3\lambda^{q-r}$

TABLE IX. Numerator of B_{lm}, $m=(p/q\,0)_2$, q even, $(p/q\,0)_1$, $l=(p/q\;p/q)_{1,2}$. For $m=(p/q\,0)_2$, q odd, multiply by $1+\lambda^q$.

$l \backslash m$	$(p/q\,0)_2$, q even, or $(p/q\,0)_1$
$(p'/q\,p'/q)_1$	$(1+\lambda^2)(\lambda^{2r}+\lambda^{q}+\lambda^{2q-2r}+\lambda^{2q})$
$(p'/q\,p'/q)_2$	$(1+\lambda^2)(\lambda^{q}+\lambda^{q+2r}+\lambda^{2q}+\lambda^{3q-2r})$

TABLE XI. Numerator of B_{lm} for $m=(p/q\,0)_2$, q even, or $(p/q\,0)_1$. For $m=(p/q\,0)_2$, q odd, multiply by $1+\lambda^q$.

$l \backslash m$	$(p/q\,0)_2$ q even or $(p/q\,0)_1$
$(p_x/q\,p_y/q)$	$(1+\lambda^2)(1+\lambda^q)(\lambda^{r_x+r_y}+\lambda^{q+r_y-r_x}+\lambda^{q+r_x-r_y}+\lambda^{2q-r_x-r_y})$

TABLE X. Numerator of B_{lm}, $m=(p/q\,0)_{1,2}$, q even, $l=(\frac{1}{2}p/q)_{1,2}$. For q odd the generating functions vanish.

$l \backslash m$	q even $(p/q\,0)_{1,2}$
$(\frac{1}{2}p'/q)_{1,2}$	$(1+\lambda^2)(1+\lambda^q)(\lambda^{r}+\lambda^{3q/2-r})$

interchanges 1 and 3 components of a vector together with a rotation $\pi/2$ in the 2-4 plane. The reflection P_x is imaged by a reflection through the origin in the 2 direction. All other operations may be written as products of these generating elements; symbolically we may write $(a\sim n_x,\, b\sim n_y)$

$$
\begin{aligned}
&(\epsilon|ab)\sim\begin{pmatrix}R_a & 0\\ 0 & R_b\end{pmatrix},\quad (\pi|ab)\sim\begin{pmatrix}R_aP & 0\\ 0 & R_bP\end{pmatrix},\\
&\left(\frac{\pi}{2}\Big|ab\right)\sim\begin{pmatrix}0 & R_aP\\ R_b & 0\end{pmatrix},\quad \left(\frac{3\pi}{2}\Big|ab\right)\sim\begin{pmatrix}0 & R_a\\ R_bP & 0\end{pmatrix},\\
&(P_x|ab)\sim\begin{pmatrix}R_aP & 0\\ 0 & R_b\end{pmatrix},\quad (P_y|ab)\sim\begin{pmatrix}R_a & 0\\ 0 & R_bP\end{pmatrix},\\
&(P_{\pi/4}|ab)=\begin{pmatrix}0 & R_a\\ R_b & 0\end{pmatrix},\quad (P_{3\pi/4}|ab)\sim\begin{pmatrix}0 & R_aP\\ R_bP & 0\end{pmatrix};
\end{aligned}
\tag{3.2}
$$

R_a means a rotation $2\pi a/q$,

$$R_a=\begin{pmatrix}\cos 2\pi a/q & -\sin 2\pi a/q\\ \sin 2\pi a/q & \cos 2\pi a/q\end{pmatrix},$$

and P stands for

$$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},$$

a reflection of the second component. It may be verified that this realization has the same multiplication table as the space group elements. It carries the representation $(1/q\ 0)_1$.

There are $\frac{1}{8}(q+3)(q+9)$ classes, or IR's, for q odd, and $\frac{1}{8}(q+6)(q+12)$, for q even. The number of elements is $8\,q^2$; we do not attempt to sort them into classes, but evaluate the sum in (1.1) by summing directly over elements:

$$B_{lm}(\lambda)=(8q^2)^{-1}\sum_\alpha \frac{\chi^*_{\alpha l}}{\Pi_{i=1}^{f_m}(1-\lambda s^{(i)}_{\alpha m})};$$

the element α is $(\beta\,|n_xn_y)$, where β is one of the eight elements of the point group and $0\leqslant n_x, n_y\leqslant q-1$. The $s^{(i)}_{\alpha m}$ are the eigenvalues of α in the IR m, $\chi_{\alpha l}$ is their sum in the IR l.

A typical sum, which arises in computing $B_{(00)_1(p/q\,0)_1}$, is

$$
\begin{aligned}
q^{-1}&\sum_{n=0}^{q-1}[(1-\lambda e^{2\pi i n\,p/q})(1-\lambda e^{-2\pi i n\,p/q})]^{-1}\\
&=q^{-1}\sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\lambda^{a+b}\sum_{n=0}^{q-1}e^{2\pi i n(a-b)p/q}.
\end{aligned}
$$

The n sum vanishes unless $a-b=0$ modulo q (we suppose p and q are relatively prime), i.e., $a=b+cq$, $c=0,1,2,\ldots$, or $b=a+cq$, $c=1,2,\ldots$. We get

$$
\begin{aligned}
&\sum_{b=0}^{\infty}\sum_{c=0}^{\infty}\lambda^{2b+cq}+\sum_{a=0}^{\infty}\sum_{c=1}^{\infty}\lambda^{2a+cq}\\
&\quad=(1+\lambda^q)[(1-\lambda^2)(1-\lambda^q)]^{-1}.
\end{aligned}
$$

After a great deal of this sort of algebra we arrive at the contents of Tables IV–XI. They give all the generating functions $B_{lm}(\lambda)$ where $m=(p/q\ 0)_1$, $(p/q\ 0)_2$, $1\leqslant p\leqslant[(q-1)/2]$. The denominator is independent of m and is found in Table IV. We have that p' is related to r by $p'=q/2-|(rp)_{\mathrm{mod}\,q}-q/2|$, $r=1,2,\ldots,[(q-1)/2]$. The same formula holds with $p'\to p_x(p_y)$ and $r\to r_x(r_y)$. It is assumed that p/q (but not necessarily $p'/q, p_x/q, p_y/q$) is in its lowest terms.

We have applied the dimension check (1.3) to these generating functions. We have also examined the irrational k case, by letting $q\to\infty$; the dimension check is still satisfied.

Generating functions for polynomial tensors based on the IR's $(p/q\,p/q)_{1,2}$, $(\frac{1}{2}p/q)_{1,2}$, $(p_x/q\,p_y/q)$ could be found by similar methods.

To construct the polynomial tensors described by the generating functions, it is only necessary to construct the elements of the integrity basis. They are the denominator scalars and numerator tensors. Just assume that the components are polynomials of the prescribed degree and impose that they transform correctly when the components of the tensor on which they are based are transformed by each of the generating elements. For examples of the procedure (for point groups) see Ref. 4.

IV. CLOSING REMARKS

The work in this paper was inspired by a seminar given in Montreal by J. Birman and by the papers of Jarić and Birman,[9] in which are given generating functions for polynomial scalars based on IR's at special symmetry points of the BZ for the three-dimensional space group $pm3n$. We have developed an approach which is simpler to apply and also more general (it works for tensors transforming by any IR); we have rederived all their results and plan to publish further results on $pm3n$ in a future publication.

It would be advisable to find such generating functions for other space groups. They would be useful for attacking problems of Landau phase transitions or making renormalization group calculations. The results would be analytic and valid to all degrees. We have no plans to pursue such a project at present, however.

Our results are for polynomial tensors only; they correspond to completely symmetric plethysms. To obtain the generating function for plethysms of all exchange symmetries based on an IR m of a space group G it is only necessary to embed the corresponding image group of G in SU(f) (f is the dimension of the IR m) so that m spans the defining IR of SU(f). The generating function for branching rules SU(f)$\supset I$, where I is the image group, then enumerates plethysms of all symmetries. See Ref. 10 for application of this approach to other finite groups.

ACKNOWLEDGMENTS

We thank M. V. Jarić for helpful discussions.

This work was supported by the Natural Sciences and Engineering Council of Canada and by the Ministère de l'Education du Québec.

[1]L. Michel, J-C. Toledano, and P. Toledano, *Proceedings of the Colloque P. Curie*, Paris, Sept. 1980 (IDSET, Paris, 1981), p. 263; L. Michel, *Proceedings of the 6th International Conference on Mathematical Physics*, Berlin, August 1981 (Springer, Berlin, 1982), p. 336.

[2]B. L. Davies and A. P. Cracknell, *Kronecker Product Tables, Vol. 4; Symmetrized Powers of IR's of Space Groups* (Plenum, New York, 1980).

[3]A. G. McLellan, J. Phys. C 7, 3326 (1974); L. Michel, in *Group Theoretical*

Methods in Physics, edited by R. T. Sharp and B. Kolman (Academic, New York, 1977); V. Kopsky, J. Phys. A: Math. Gen. **12**, 429, 943 (1979).
[4]J. Patera, R. T. Sharp, and P. Winternitz, J. Math. Phys. **19**, 2362 (1978).
[5]I. V. V. Raghavachacharyulu, Can. J. Phys. **39**, 830 (1961).
[6]S. C. Miller and W. F. Love, *Tables of IR 's of Space Groups and Corepresentations of Magnetic Space Groups* (Pruett, Boulder, CO, 1967).
[7]D. Phaneuf and R. T. Sharp, *Lecture Notes in Physics, 135* (Springer, Berlin, 1980), p. 517.
[8]N. F. M. Henry and K. Lonsdale, *International Tables for X-Ray Crystallography* (Kynoch, Birmingham, U.K., 1952), Vol. 1, p. 45.
[9]M. V. Jarić and J. L. Birman, J. Math. Phys. **18**, 1456, 1459 (1977).
[10]J. Patera and R. T. Sharp, J. Phys. A: Math. Gen. **13**, 397 (1980).

Operator Racah algebra and Laplace-type expansions of irreducible spherical tensors

A. W. Niukkanen
Vernadsky Institute for Geochemistry and Analytical Chemistry, Academy of Sciences of the USSR, Moscow, USSR

(Received 20 July 1984; accepted for publication 30 November 1984)

Differential formulas for coefficients in the Laplace-type series of an arbitrary spherical tensor $f_{LM}(\mathbf{r}+\mathbf{R})$ are given in terms of an operator N applied to the radial part $\varphi(r)$ of $f_{LM}(\mathbf{r})$. Very compact and convenient expressions for N in terms of operator Pochhammer symbols are established. A special representation of the coefficients of the Laplace-type series, in terms of the operator Gauss function ${}_2F_1$, is given, which, in turn, provides a remarkably short proof of two earlier Sack expansions. More general gradient formulas are introduced and numerous particular cases of the Laplace-type expansions are considered in detail.

I. INTRODUCTION

Expansions of a given function $f(\mathbf{r}+\mathbf{R})$ in terms of a complete set of functions $\psi_i(\mathbf{r})$ (translational transformation of the function f) are frequently used in various physical applications. The Laplace-type series represents the most important example of such expansions. Many particular cases of the Laplace-type series have been studied in detail and several attempts[1-3] have been made to give a general method of obtaining coefficients in such expansions, for the case of an arbitrary function f. Though none of these attempts has resulted, apparently, in a comprehensive and sufficiently general method, it was shown,[4] however, that the Bayman approach[3] can be readily reformulated to give relatively simple and transparent formulas for the coefficients. The resulting expansions bear some general resemblance to those given by Dr. Brook Taylor in that the coefficients are represented by differential operators which are applicable to any arbitrary function f to be expanded.

Inasmuch as the functions $f(\mathbf{r})$ encountered in applications have, usually, the form of an irreducible spherical tensor (IST),

$$f_{LM}(\mathbf{r}) = f(r)Y_{LM}(\mathbf{r}) = \varphi(r)\,\mathcal{Y}^0_{LM}(\mathbf{r}), \tag{1}$$

where $Y_{LM}(\mathbf{r})$ is the spherical function and $\mathcal{Y}^0_{LM}(\mathbf{r}) = r^L \times Y_{LM}(\mathbf{r})$ is the regular solid harmonic, therefore one can easily anticipate that a sort of *operator Racah algebra* should be relevant to Laplace-type expansions of an IST f_{LM}.

A typical example is given by the irreducible tensor derivative (ITD) of the function (1) (see Ref. 4):

$$\begin{aligned} f^{nlL}_{\lambda\mu}(\mathbf{r}) &\equiv \{\mathcal{Y}^n_l(\nabla)\otimes f_L(\mathbf{r})\}_{\lambda\mu} \\ &= \{\mathcal{Y}^n_l(\nabla(\mathbf{r}))\otimes\varphi(r)\,\mathcal{Y}^0_L(\mathbf{r})\}_{\lambda\mu}, \end{aligned} \tag{2}$$

where $\nabla(\mathbf{r})$ is the gradient operator;

$$\mathcal{Y}^n_{lm}(\mathbf{r}) = r^{l+2n}Y_{lM}(\mathbf{r}) = r^{2n}\mathcal{Y}^0_{lm}(\mathbf{r}); \tag{3}$$

and $\{\chi_l\otimes f_L\}_{\lambda\mu}$ is the irreducible tensor product (ITP)[4,5] of the IST's χ_{lm} and f_{LM}. The ITD's (2) occur, for example, in the expansion of $f_{LM}(\mathbf{r}+\mathbf{R})$ in terms of the functions $\mathcal{Y}^n_{lm}(\mathbf{r})$,

$$\begin{aligned} f_{LM}(\mathbf{r}+\mathbf{R}) = \sum_{nl\lambda}(-1)^l\frac{\pi^{3/2}\pi(\lambda,\bar{L})2^{-l-2n+1}}{\Gamma(l+\frac{3}{2})n!(l+\frac{3}{2})_n} \\ \times\{\mathcal{Y}^n_l(\mathbf{R})\otimes\{\mathcal{Y}^n_l(\nabla(\mathbf{r}))\otimes f_L(\mathbf{r})\}_\lambda\}_{LM}, \end{aligned} \tag{4}$$

where $\pi(a,b,\bar{c},\ldots) = [(2a+1)(2b+1)(2c+1)^{-1}\cdots]^{1/2}$ and $(a)_n$ is the Pochhammer symbol which will be, occasionally, referred to as a P symbol or P function of order n.

It has been shown[4] that the ITD (2) has the form of an IST:

$$\{\mathcal{Y}^n_l(\nabla(\mathbf{r}))\otimes\varphi(r)\,\mathcal{Y}^0_L(\mathbf{r})\}_{\lambda\mu} = \Phi^{(nlL\lambda)}(r)\,\mathcal{Y}^0_{\lambda\mu}(\mathbf{r}), \tag{5}$$

whose radial part is

$$\Phi^{(nlL\lambda)}(r) = H(l,L,\lambda)2^\nu N_{nlL\lambda}(u)\,\varphi(r), \tag{6}$$

where the constant $H(l,L,\lambda)$ is related to the Clebsch–Gordan coefficient $\langle l0L0|\lambda 0\rangle$ by

$$H(l,L,\lambda) = (4\pi)^{1/2}\pi(l,L,\bar{\lambda})\langle l0L0|\lambda 0\rangle, \tag{7}$$

and $N(u)$ is the differential operator

$$N_{nlL\lambda}(u) = \nu!L^{\lambda+1/2}_\nu\Big(-\overset{2}{u}\,\overset{1}{d}(u)\Big)d^{\,\omega}(u), \tag{8}$$

where $u = r^2/2$, $d(u) = \partial/\partial u$, $L^\alpha_\nu(x)$ is the Laguerre polynomial and

$$\nu = n + [lL\bar{\lambda}], \tag{9}$$

$$\omega = n + [l\bar{L}\lambda], \tag{10}$$

$$[ab\bar{c}\cdots] = (a+b-c+\cdots)/2. \tag{11}$$

The encircled indices in Eq. (8) mark, according to the Maslov and Feynmann notations,[6] the operator positions. In any algebraic transformations the quantities $\overset{2}{u}$ or their functions should stand on the left of the expressions depending on $\overset{1}{d}(u)$. For example, $\big(\overset{2}{u}\,\overset{1}{d}(u)\big)^i = u^id^{\,i}(u)$, etc.

If $n=0$, then Eq. (5) is equivalent to the Bayman gradient formula Eq. (2) in Ref. 3. Equation (8) gives in this case a compact analytical expression for the Bayman differential operator which is written in Eq. (11) in Ref. 3 as a cumbersome "unidentified" sum. If $n\neq 0$, then Eqs. (5), (6), and (8) give an essential generalization of the Bayman result (which, in turn, is a generalization of the well-known Darwin gradient formula), since they allow us to represent, in a uniform and compact form, *all* the tensor derivatives occurring in Eq. (4), rather than those ITD's which correspond to $n=0$. Equations (5) and (6) will be called, therefore, the *generalized gradient formula*.

 0022-2488/85/071540-07$02.50

We derive here alternative expressions for the operator $N(u)$, some of which prove to be simpler compared to Eq. (8) and, in many cases, more expedient for calculations and analytical transformations. This leads to convenient expressions for coefficients in many physically important Laplace-type expansions of functions whose radial part is represented by generalized hypergeometric series.

II. OPERATOR ANALOGS OF THE NEWTON BINOMIAL THEOREM

Expansion of an operator binomial $(d(x)+f(x))^n$, where $d(x)=d/dx$, in terms of the operators $d^i(x)$, standing—relative to coefficients of the expansion—in the right position (operator analog of the binomial theorem), is, generally, applicable to various analytical problems. In particular, one of the expansions obtained in this section provides an expedient starting point for establishing alternative forms of the operator $N(u)$ in Eq. (8).

Using an *operator* identity

$$\varphi^{-1}(x)d^n(x)\,\varphi(x)=(\varphi^{-1}(x)d(x)\,\varphi(x))^n\,, \tag{12}$$

and the commutation relation

$$d(x)\,\varphi(x)=\varphi(x)d(x)+\varphi'(x)\,, \tag{13}$$

and expressing $\varphi^{-1}\varphi'$ as $(\ln\varphi)'$, we have

$$\left(d(x)+\frac{d}{dx}\ln\varphi(x)\right)^n=\varphi^{-1}(x)d^n(x)\,\varphi(x)\,. \tag{14}$$

Operating with the right-hand side of Eq. (14) on a "test" function $\psi(x)$, doing the necessary differentiation with the help of the Leibniz rule, and converting the resulting expression into an operator form by omitting $\psi(x)$, we obtain the following operator paraphrase of the binomial theorem:

$$\left(d(x)+\frac{d}{dx}\ln\varphi(x)\right)^n=\sum_{i=0}^{n}C_n^i\langle\varphi^{-1}d^{(n-i)}(x)\,\varphi\rangle d^i(x)\,, \tag{15}$$

where C_n^i is the binomial coefficient and the angular brackets are used to indicate that the expression inside should be interpreted as a function rather than an operator, opposite, say, to the right-hand side of Eq. (14). In the case of the operator binomial $(d+f)^n$, the function $\varphi(x)$, on the right in Eq. (15), should be replaced by

$$\varphi(x)=\exp\left(\int f(x)dx\right). \tag{16}$$

In many particular cases the coefficients $\varphi^{-1}\varphi^{(n-i)}$ in Eq. (15) can be given an algebraic form. For example, if φ is a weight function for an orthogonal polynomial, then the quantities $\varphi^{-1}\varphi^{(n-i)}$ can be easily calculated by means of the Rodrigues formulas.[7] If $\varphi=\exp(-x^2/2)$, then[7(a)]

$$(d(x)-x)^n=\sum_{i=0}^{n}C_n^i(-1)^{n-i}H_{n-i}(x)d^i(x)\,, \tag{17}$$

where $H_k(x)$ is the Hermite polynomial.[8] In the case $\varphi=(1-x)^A(1+x)^B$ we get

$$\left(d(x)+\frac{B}{1+x}-\frac{A}{1-x}\right)^n=\sum_{i=0}^{n}C_n^i(-2)^{n-i}(n-i)!(1-x^2)^{-n+i}\times P_{n-i}^{(A-n+i,B-n+i)}(x)d^i(x)\,, \tag{18}$$

where $P_k^{(\alpha,\beta)}(x)$ is the Jacobi polynomial.[7(b)] If $\varphi=x^A\times\exp(-tx)$, then

$$\left(d(x)+\frac{A}{x}-t\right)^n=\sum_{i=0}^{n}C_n^i x^{-n+i}(n-i)!L_n^{A-n+i}(x)d^i(x)\,, \tag{19}$$

where $L_k^\alpha(x)$ is the Laguerre polynomial.[7(c)]. A number of "nonclassical" polynomials[7] may be treated in the same manner, as well.

Applying the operators (17)–(19) to the unit function results in differential representations for orthogonal polynomials which prove to be useful modifications of the Rodrigues formulas. Applying these operators to more complicated functions, one would obtain various addition theorems and "sum rules." For instance, operating with Eqs. (17)–(19) on the corresponding (to H, P, and L polynomials, respectively) weight functions and realizing that the weight functions' product is, again, a weight function belonging to the same class, that is,

$$(1-x)^A(1+x)^B\times(1-x)^C(1+x)^D=(1-x)^{A+C}(1+x)^{B+D}\,,$$

etc., then by using Eq. (14) one could easily obtain some evident bilinear expansions in terms of polynomials with changing weight indices. Note that such polynomials arise in some important physical applications.[9]

In the following we restrict ourselves to the particular version of Eq. (19) corresponding to $t=0$. Since $L_n^\alpha(0)=(-1)^n(-n-\alpha)_n/n!$, by transforming the resulting sum to hypergeometric form we obtain

$$(d(x)+A/x)^n=(A-n+1)_n x^{-n}\times\Phi(-n,A-n+1;-xd(x))\,, \tag{20}$$

where $\Phi(a,c;x)$ is the Kummer function. A similar relation results if we put $A=0$ in Eq. (18).

III. ALTERNATIVE EXPRESSIONS FOR THE OPERATOR $N(u)$

Since

$$L_n^\alpha(z)=(\alpha+1)_n/n!\,\Phi(-n,\alpha+1;z)\,, \tag{21}$$

then, combining Eqs. (20) and (8), we obtain

$$N_{nlL\lambda}(u)=u^\nu(d(u)+(\lambda+\nu+\tfrac{1}{2})/u)^\nu d^\omega(u)\,. \tag{22}$$

Using Eq. (14) we have

$$N_{nlL\lambda}(u)=u^{-\lambda-1/2}d^\nu(u)u^{\lambda+\nu+1/2}d^\omega(u)\,. \tag{23}$$

Although Eqs. (22) and (23) look, aparently, simpler compared to Eq. (8), a further transformation of $N(u)$ such that the operators $d(u)$ would be involved in $N(u)$ through P functions of the argument $\delta(u)=u\,d(u)$, exclusively, proves to be possible.

Consider some important elementary properties of the operator $\delta(u)$ which plays an important part in formal transformations of hypergeometric series.[10(a),(b),(c)] By the use of the commutation relations

$$d(x)x^\alpha-x^\alpha d(x)=\alpha x^{\alpha-1}\,, \tag{24a}$$

$$x\,d^\alpha(x)-d^\alpha(x)x=-\alpha\,d^{\alpha-1}(x)\,, \tag{24b}$$

the operator identities

$$f[\delta(x)]x^{\alpha} = x^{\alpha} f[\delta(x) + \alpha] , \tag{25a}$$

$$x^{\alpha} f[\delta(x)] = f[\delta(x) - \alpha] x^{\alpha} , \tag{25b}$$

$$f[\delta(x)] d^{\alpha}(x) = d^{\alpha}(x) f[\delta(x) - \alpha] , \tag{26a}$$

$$d^{\alpha}(x) f[\delta(\alpha)] = f[\delta(x) + \alpha] d^{\alpha}(x) \tag{26b}$$

are readily verified, which hold for any analytical function $f(z)$. Though Eqs. (25) and (26) are rather trivial, they are, nevertheless, very expedient in deriving various properties of operator functions $f(\delta)$. For instance, they immediately lead, by induction, to the standard relations

$$x^n d^n \equiv (\underset{②①}{xd})^n = (-1)^n (-\delta(x))_n , \tag{27}$$

$$d^n x^n \equiv (\underset{②①}{dx})^n = (\delta(x) + 1)_n , \tag{28}$$

which show, in particular, that the product of *two* operators, x^n and d^n, as well as the operator $d^n x^n$, are functions of one operator δ (see Ref. 11). Note that the right-hand sides of Eqs. (27) and (28) can be transformed with the help of the well-known identity

$$(a)_n = (-1)^n (-a - n + 1)_n . \tag{29}$$

Though none of the equations (25)–(28), separately, is novel [see, for example, Ref. 10(a)], the system of these equations, as a whole, provides a complete set of elementary steps which are sufficient to transform a product $x^{\alpha} d^{\beta} x^{\gamma} d^{\omega} \cdots$ to any prescribed form.

For example, transforming the first three multipliers on the right of (23) with the help of Eqs. (28) and (25a) we have, consecutively,

$$u^{-\lambda - 1/2} d^{\nu}(u) u^{\lambda + \nu + 1/2} = u^{-\lambda - 1/2} (d^{\nu}(u) u^{\nu}) u^{\lambda + 1/2}$$
$$= u^{-\lambda - 1/2} (\delta(u) + 1)_{\nu} u^{\lambda + 1/2} = (\delta(u) + \lambda + \tfrac{3}{2})_{\nu} . \tag{30}$$

With the aid of Eq. (30), Eq. (23) becomes

$$N_{nlL\lambda}(u) = (\delta(u) + \lambda + \tfrac{3}{2})_{\nu} d^{\omega}(u) , \tag{31}$$

or, taking account of Eq. (26b),

$$N_{nlL\lambda}(u) = d^{\omega}(u) (\delta(u) + L - \nu + \tfrac{3}{2})_{\nu} . \tag{32}$$

The connection between Eqs. (8) and (31) is also implied by the following curious observation. Writing the Laguerre polynomial in Eq. (8) as the Kummer function Φ, using an explicit expression for Φ as a series in powers of the operator $-\underset{②①}{ud}(u)$, and transforming the latter with the aid of Eq. (27), one can express $N(u)$ as a Gauss function ${}_2F_1$ of the unit argument,

$$N_{nlL\lambda}(u) = (\lambda + \tfrac{3}{2})_{\nu}\ {}_2F_1(-\nu, -\delta(u); \lambda + \tfrac{3}{2}; 1) d^{\omega}(u) , \tag{33}$$

This expression can be easily shown to be consistent with Eq. (31), if use is made of the Gauss summation theorem.[10(d)]

Expressing $d^{\omega}(u)$ in Eq. (31) with the aid of Eq. (27), as $(-u)^{-\omega} (-\delta(u))_{\omega}$ and pulling the multiplier $u^{-\omega}$ to the left position with the aid of Eq. (25a), we obtain the expression

$$N_{nlL\lambda}(u) = (-u)^{-\omega} (\delta(u) + L - \nu + \tfrac{3}{2})_{\nu} (-\delta(u))_{\omega} , \tag{34}$$

with the operators $d(u)$ being involved through P functions of the argument $\delta(u)$, exclusively. In each of the equations (31), (32), and (34) the sign of the operator $\delta(u)$ in the P symbols' arguments can be changed with the aid of the identity (29). For example, applying Eq. (29) to the second P symbol on the right-hand side of Eq. (34), we obtain

$$N_{nlL\lambda}(u) = u^{-\omega} (\delta(u) + L - \nu + \tfrac{3}{2})_{\nu} (\delta(u) - \omega + 1)_{\omega} . \tag{35}$$

If the same transformation is applied to the first P symbol, then

$$N_{nlL\lambda}(u) = (-1)^{\nu + \omega} u^{-\omega} (-\delta(u) - L - \tfrac{1}{2})_{\nu} (-\delta(u))_{\omega} . \tag{36}$$

IV. GENERAL REMARKS

Equations (35) and (36) serve as a basis for all transformations which are used in the sequel. We shall make here some precursory remarks featuring the possible applications of these equations.

The operator $N(u)$ may occur in expressions of two different types. The first type expressions are those which involve summation over n [see, for example, the right-hand side of Eq. (4)]. Since these summations are, as a rule, of hypergeometric type, it is expedient to express the operator $N_{nlL\lambda}(u)$ through P symbols $(a)_n$ of order n (with a independent of n). To this end one should use Eq. (36). Indeed, since the variables ν and ω differ from n in additive contributions [see Eqs. (9) and (10)], with the aid of

$$(a)_{n+k} = (a)_k (a + k)_n , \tag{37}$$

the operator $N(u)$ in Eq. (36) can be easily transformed to the necessary form

$$N_{nlL\lambda}(u) = (-1)^l u^{-[l\lambda\bar{L}]} u^{-n} (-\delta(u) - [L\lambda\bar{l}] - \tfrac{1}{2})_n$$
$$\times (-\delta(u) + [l\lambda\bar{L}])_n (-\delta(u) - L - \tfrac{1}{2})_{[lL\bar{\lambda}]}$$
$$\times (-\delta(u))_{[l\lambda\bar{L}]} . \tag{38}$$

The second type expressions are those in which the operator $N(u)$ is to be applied to a hypergeometric series $F(r^{\alpha})$. In term-by-term application of $N(u)$ to the series, the following simple properties of the operator $\delta(x)$ should be taken into account. First, since $\delta(x) x^i = i x^i$, then

$$f[\delta(x)] x^i = f[i] x^i . \tag{39}$$

Second, if c is a constant, then

$$\delta(cx) = \delta(x) . \tag{40}$$

Third,

$$\delta(x) = x (\ln v(x))' \delta(v(x)) . \tag{41}$$

If we put $v(x) = x^{\alpha}$ in Eq. (41), then

$$\delta(x^{\alpha}) = \alpha^{-1} \delta(x) . \tag{42}$$

Since $u = r^2/2$, these equations imply that $\delta(u) = \delta(r)/2$, from which we have $f[\delta(u)] r^{\alpha i} = f[\alpha i/2] r^{\alpha i}$, where i is the summation index involved in explicit representation of the series $F(r^{\alpha})$. If α is a rational fraction of the form $\alpha = 2S/M$, then, expressing i as $i = kM + t$, where $0 \leqslant t < M$, we have

$$f[\delta(u)] r^{\alpha i} = r^{2St/M} f[kS + St/M] (r^{2S})^k . \tag{43}$$

If α is a rational fraction of the form $\alpha = (2S + 1)/M$, then, using the representation $i = 2kM + t$, where $0 \leqslant t < 2M$, we have

$$f[\delta(u)]r^{\alpha i} = r^{(2S+1)t/M} f\left[k(2S+1) + \frac{(2S+1)t}{2M}\right](r^{4S+2})^k. \tag{44}$$

Setting $f[\delta(u)] = N(u)$ and using Eq. (35), the resulting P symbols can be written, with the aid of Eqs. (43) and (44), as the products of P symbols of order k, if use is made of the Gauss–Legendre multiplication theorem[10(e)] for $\Gamma(Nk)$, where $N = S$ or $N = 2S + 1$. Thus the initial hypergeometric series $F(r^\alpha)$ splits into a sum of $M - 1$ hypergeometric series, in the case of Eq. (43), and into a sum of $2M - 1$ hypergeometric series, in the case of Eq. (44). Therefore, Eq. (35), rather than Eq. (36), should be used in the case of the second-type expressions.

Two examples of the first-type expressions are discussed in Secs. V–VII. Some simple examples of the second-type expressions, corresponding to the cases $\alpha = 2$ and $\alpha = 1$, are considered in Sec. VIII.

V. REPRESENTATION OF THE COEFFICIENTS OF THE LAPLACE-TYPE EXPANSIONS IN TERMS OF THE OPERATOR GAUSS FUNCTIONS

Substituting Eqs. (5) and (6) in Eq. (4) and using Eq. (38), one can easily show that summation over n gives rise to the Gauss function ${}_2F_1$. Taking account of the operator ordering, we obtain

$$\begin{aligned} f_{LM}(\mathbf{r}+\mathbf{R}) = \sum_{l,\lambda} 2\pi^{3/2} \frac{\pi(\lambda,\bar{L})H(l,L,\lambda)}{\Gamma(l+\frac{3}{2})} \\ \times\{\mathcal{Y}_l^0(\mathbf{R}) \otimes \mathcal{Y}_\lambda^0(\mathbf{r})\}_{LM} r^{-2[l\lambda\bar{L}]} \\ \times {}_2F_1(-\underset{①}{\delta}(r^2) - [L\lambda\bar{l}] - \tfrac{1}{2}, \\ -\underset{①}{\delta}(r^2) + [l\lambda\bar{L}]; l + \tfrac{3}{2}; R^2 \underset{②}{r}{}^{-2}) \\ \times(-\delta(r^2) - L - \tfrac{1}{2})_{[lL\bar{\lambda}]}(-\delta(r^2))_{[l\lambda\bar{L}]}\varphi(r). \end{aligned} \tag{45}$$

In the scalar case $L = M = 0$ we have $\lambda = l$. Since $Y_{00} = 1/\sqrt{4\pi}$ and $H(l,0,l) = 1/\sqrt{4\pi}$, Eq. (45) converts to

$$\begin{aligned} f_{00}(\mathbf{r}+\mathbf{R}) &= (4\pi)^{-1/2}\varphi(|\mathbf{r}+\mathbf{R}|) \\ &= \sum_l \sqrt{\pi} \frac{\pi(l)}{\Gamma(l+\frac{3}{2})}\{\mathcal{Y}_l^0(\mathbf{R}) \otimes \mathcal{Y}_l^0(\mathbf{r})\}_{00} r^{-2l} \\ &\quad \times {}_2F_1(-\underset{①}{\delta}(r^2) - \tfrac{1}{2}, -\underset{①}{\delta}(r^2) + l; l + \tfrac{3}{2}; R^2 \underset{②}{r}{}^{-2}) \\ &\quad \times(-\delta(r^2))_l \varphi(r), \end{aligned} \tag{46}$$

Note that the parameters of the series ${}_2F_1(a,b;c;z)$ in Eq. (46) are related by the equality $c = b - a + 1$. This implies that the function ${}_2F_1$ satisfies, in the scalar case, numerous quadratic transformations[10(f)] which can be used to give the argument of ${}_2F_1$ an appropriate analytical form. Some of these forms prove to be symmetric in R and r.

VI. THE SACK EXPANSIONS

If

$$f_{LM}(\mathbf{r}) = \mathcal{Y}_{LM}^N(\mathbf{r}) \equiv r^{2N} \mathcal{Y}_{LM}^0(\mathbf{r}), \tag{47}$$

then $\varphi(r) = r^{2N}$. Expansions of $\mathcal{Y}_{LM}^N(\mathbf{r}+\mathbf{R})$ in terms of bipolar harmonics, in the general ($L \neq 0$) and in the scalar ($L = 0$) cases, are given, immediately, by Eqs. (45) and (46), respectively, if we put $\varphi = r^{2N}$ in their right-hand sides and replace, formally, $\delta(r^2) \to N$. Besides that, the ordering indices 1 and 2 should be omitted. As a result, we get a remarkably short proof of two Sack expansions for the scalar[12] and general[1] cases. For exact correspondence with Sack's formulas, besides correlating with the notations of Refs. 1 and 12 [$N \to (N-1)/2$, etc.], $\Gamma(1 + 3/2)$ should be written as a double factorial and all P symbols should be expressed through Γ functions, provided that some of P symbols are, preliminarily, transformed with the aid of Eq. (29). In addition, use should be made of the symmetry relations

$$\begin{aligned} (-1)^\lambda \pi(\lambda)H(l,L,\lambda) &= (-1)^L \pi(L)H(l,\lambda,L) \\ &= (-1)^l \pi(l)H(L,\lambda,l). \end{aligned} \tag{48}$$

It is worth noting that the expansion of the $\mathcal{Y}_{00}^N(\mathbf{r}+\mathbf{R})$ (the scalar case) can be reformulated in terms of four-dimensional sphererical harmonics corresponding to *hyperbolic* rotations in four-dimensional *pseudo-Euclidean* space. Indeed, if $\delta(r^2) \to N$, then the function ${}_2F_1$ in Eq. (46) can be written in terms of the Gegenbauer polynomial[7(d)]

$$\begin{aligned} {}_2F_1\left(-N + l, N - \tfrac{1}{2}; l + \tfrac{3}{2}; \frac{R^2}{r^2}\right) \\ = \frac{(N-l)!}{(2l+2)_{N-l}}\left(\frac{r^2 - R^2}{r^2}\right)^{N-l} C_{N-l}^{l+1}\left(\frac{r^2+R^2}{r^2-R^2}\right). \end{aligned} \tag{49}$$

If $R < r$, then the argument φ of C_n^λ in Eq. (49) satisfies inequality $1 \leqslant \varphi < \infty$, which makes possible the following parametrization:

$$\varphi \equiv (R^2 + r^2)/(r^2 - R^2) = \cosh\gamma. \tag{50}$$

Then

$$2Rr/(r^2 - R^2) = \sinh\gamma.$$

Taking into account that the four-dimensional harmonic, corresponding to rotations in four-dimensional Euclidean space, is[7(e),13]

$$Y_{Nlm}^{(4)}(\epsilon,\theta,\varphi) = (\sin\epsilon)^l C_{N-l}^{l+1}(\cos\epsilon) Y_{lm}(\theta,\varphi), \tag{51}$$

and using the relations $\cosh\gamma = \cos i\gamma$, $\sinh\gamma = -i\sin i\gamma$, after some minor algebraic transformations, we obtain

$$\begin{aligned} \mathcal{Y}_{00}^N(\mathbf{r}+\mathbf{R}) = 2N! \sum_l \frac{(2i)^l \pi(l)}{(l+1)_{N+1}} \\ \times\{Y_l(\mathbf{R}) \otimes \rho^{2N} Y_{Nl}^{(4)}(i\gamma,\theta,\varphi)\}_{00}, \end{aligned} \tag{52}$$

where $\rho^2 = \boldsymbol{\rho}\cdot\boldsymbol{\rho}$ and $\boldsymbol{\rho}$ is the four-dimensional vector with the components (x,y,z,iR). Evidently, the case of $R < r$ corresponds to the inner points of the cone $R = (x^2 + y^2 + z^2)$.$^{1/2}$ In the case $r < R$ the formal replacement $\mathbf{r} \rightleftarrows \mathbf{R}$ should be made in Eq. (52). In this latter case the four-dimensional harmonics refer to a space whose points are specified by the vectors $\boldsymbol{\rho} = (X,Y,Z,ir)$.

VII. MORE GENERAL GRADIENT FORMULAS

Equation (4) gives us the expansion of an IST $f_{LM}(\mathbf{r}+\mathbf{R})$ in terms of the functions $\mathcal{Y}_{lm}^n(\mathbf{R})$. In the case of more general expansions the ITD (2) in the right-hand side of Eq. (4) is replaced by more complicated ITD's. In many cases, however, such ITD's prove to be particular cases of the general expression

$$\left\{ {}_pF_q\begin{bmatrix}\mathbf{a};t\Delta(\mathbf{r})\\ \mathbf{c}\end{bmatrix} \mathscr{Y}_l^0(\boldsymbol{\nabla}(\mathbf{r}))\otimes f_L(\mathbf{r})\right\}_{\lambda\mu}, \tag{53}$$

where ${}_pF_q$ is generalized hypergeometric series (GHS)[10(g)] in powers of the Laplace operator $\Delta(\mathbf{r})$. The sets of parameters in ${}_pF_q$ are denoted, for brevity, as

$$\mathbf{a}=(a_1,\ldots,a_p),\quad \mathbf{c}=(c_1,\ldots,c_q).$$

Representing the GHS ${}_pF_q$ in Eq. (53) in explicit form[10(g)] and taking account of Eqs. (5), (6), and (38), we obtain the generalized analog of the gradient formula (5)

$$\left\{ {}_pF_q\begin{bmatrix}\mathbf{a};t\Delta(\mathbf{r})\\ \mathbf{c}\end{bmatrix} \mathscr{Y}_l^0(\boldsymbol{\nabla}(\mathbf{r}))\otimes f_L(\mathbf{r})\right\}_{\lambda\mu}$$
$$= {}_p\Phi_q^{(lL\lambda)}(\mathbf{a},\mathbf{c};t,r)\mathscr{Y}^0_{\lambda\mu}(r), \tag{54}$$

where

$${}_p\Phi_q^{(lL\lambda)}(\mathbf{a},\mathbf{c};t,r)=H(l,L,\lambda)2^{[lL\bar{\lambda}]}\,{}^pN^q{}_{lL\lambda}(\mathbf{a},\mathbf{c};t,u)\,\varphi(r), \tag{55}$$

$${}^pN^q_{lL\lambda}(\mathbf{a},\mathbf{c};t,u)=(-1)^l(r^2/2)^{-[l\lambda\bar{L}]}\,{}_{p+2}F_q\begin{bmatrix}\mathbf{a};-\delta(\underset{\textcircled{1}}{r}{}^2)-[L\lambda\bar{l}]-\frac{1}{2},-\delta(\underset{\textcircled{1}}{r}{}^2)+[l\lambda\bar{L}];4t^2\underset{\textcircled{2}}{r}{}^{-2}\\ \mathbf{c}\end{bmatrix}$$
$$\times(-\delta(r^2)-L-\tfrac{1}{2})_{[lL\bar{\lambda}]}(-\delta(r^2))_{[l\lambda\bar{L}]}. \tag{56}$$

It is easily verified, by carrying out the summation over n, that the right-hand side of Eq. (4) involves, implicitly, the gradient operator ${}_0F_1$ $(l+3/2;\ R^2\Delta(\mathbf{r})/4)\ \mathscr{Y}_l^0(\boldsymbol{\nabla}(\mathbf{r}))$; therefore, Eqs. (45) and (46) become, on examination, particular cases of the general relation (56).

Evidently, the transformations considered here give an example of the "first-type problem" if the use of classification adopted in Sec. IV is made.

VIII. EXAMPLES OF THE LAPLACE-TYPE EXPANSIONS

Now consider some examples of expansions described by Eq. (4) which lead to the "second-type problems."

The radial part of the tensors $f_{LM}(\mathbf{r})$ occurring in applications is representable, for the most interesting physical cases, as a product of GHS ${}_pF_q$, whose argument is proportional to r^2 or r, and a power of distance r. First, consider the case

$$\varphi(r)=r^{\beta}\,{}_pF_q\begin{bmatrix}\mathbf{a};vr^2\\ \mathbf{c}\end{bmatrix}. \tag{57}$$

To calculate the ITD (2) occurring on the right-hand side of Eq. (4), we use Eqs. (5) and (6). Denoting

$$\chi(r)=N_{nlL\lambda}(r^2/2)r^{\beta}\,{}_pF_q\begin{bmatrix}\mathbf{a};vr^2\\ \mathbf{c}\end{bmatrix}, \tag{58}$$

expressing $N(u)$ with the help of Eq. (35), commuting the operator P symbols with the power r^{β} with the aid of Eq. (25a) and representing the GHS ${}_pF_q$ in explicit form,[10(g)] by term-by-term application of Eq. (39) we obtain

$$\chi(r)=2^{\omega}r^{\beta-2\omega}\sum_i\frac{(\mathbf{a})_i}{(\mathbf{c})_i}P_{\nu}P_{\omega}(vr^2)^i/i!, \tag{59}$$

where

$$P_{\nu}=\Gamma(\beta/2+L+\tfrac{3}{2}+i)/\Gamma(\beta/2+L+\tfrac{3}{2}-\nu+i), \tag{60}$$

$$P_{\omega}=\Gamma(\beta/2+1+i)/\Gamma(\beta/2+1-\omega+i). \tag{61}$$

If β is a noninteger, then, expressing P_{ν} and P_{ω} through P symbols of order i, we have

$$N(u)\,r^{\beta}\,{}_pF_q\begin{bmatrix}\mathbf{a};vr^2\\ \mathbf{c}\end{bmatrix}$$
$$=2^{\omega}\mathrm{r}^{\beta-2\omega}(-1)^{\nu+\omega}(-\beta/2-\mathrm{L}-1/2)_{\nu}(-\beta/2)_{\omega}$$
$$\times{}_{p+2}F_{q+2}\begin{bmatrix}\mathbf{a},\beta/2+L+\frac{3}{2},\beta/2+1;vr^2\\ \mathbf{c},\beta/2+\mathrm{L}+\frac{3}{2}-\nu,\beta/2+1-\omega\end{bmatrix}. \tag{62}$$

In the case of integral β denominator parameters in ${}_{p+2}F_{q+2}$ in Eq. (62) may assume nonpositive integral values. Therefore, the formal expression for the series ${}_{p+2}F_{q+2}$ may involve divergent contributions corresponding to zero P symbols in denominators. Evidently, we should find such transformations of Eqs. (59) and (62) which would prevent appearance of divergent terms. Note that the special cases are related either to the quantity P_{ν}, if $\beta=2m$, or to the quantity P_{ω}, if $\beta=2m+1$, with $m=0,\pm1,\pm2,\ldots$ in both cases. Any of these quantities can be represented as

$$R=\Gamma(s+i)/\Gamma(s-n+i), \tag{63}$$

where

$$s=m+1,\quad n=\omega,\quad \text{if}\quad \beta=2m;$$
$$s=m+L+2,\quad n=\nu,\quad \text{if}\quad \beta=2m+1;$$

with $n\geqslant0$ and $i\geqslant0$.

On contemplation of Eqs. (59)–(61) and (63), the following three cases, $s\geqslant n+1$, $1\leqslant s\leqslant n$, and $s\leqslant0$, can be shown to arise, each giving a number of subcases for the summation variable i.

If $n-s+1\leqslant0$, then, for any $i\geqslant0$, the both Γ functions in Eq. (63) are not singular. This allows us to use Eq. (62), directly.

If $n-s+1\geqslant1$ and $-s+1\leqslant0$, then, for $0\leqslant i\leqslant-s+n$, $\Gamma(s-n+i)$ is singular; therefore $R\equiv0$. For $i\geqslant n-s+1$, both Γ functions in Eq. (63) are regular. In this case we replace the summation variable i by i', setting $i=i'+n-s+1$. With the aid of Eq. (37) all terms in Eq. (59) can be written as P symbols of order i' $(i'=0,1,2,\ldots)$.

If $s\leqslant0$ then there are contributions of three different types. For $0\leqslant i\leqslant-s$ both Γ functions in Eq. (63) have poles, that is, the quantity R becomes indefinite. In this case we use the transformation (29) which leads to a nonsingular expression for R. For $-s+1\leqslant i\leqslant n-s$ we have $R\equiv 0$. Finally, for $i\geqslant -s+n+1$, both of the Γ's in (63) are regular. In this case we introduce, again, the new summation variable i', putting $i=i'+n-s+1$, and use Eq. (37) in all terms on the right-hand side of (59).

The final results of these transformations are as follows.

If $\beta=2m$ and $m\geqslant\omega$, then

$$N(r^2/2)r^{2m}\,{}_pF_q\begin{bmatrix}\mathbf{a};vr^2\\ \mathbf{c}\end{bmatrix}={}^{m}_{p+2}S_{q+2}(\mathbf{a},\mathbf{c};v,r^2), \tag{64}$$

where

$${}^{m}_{p+2}S_{q+2}=2^{\omega}r^{2(m-\omega)}(\lambda+3/2+m-\omega)_v(-1)^{\omega}(-m)_{\omega}\times{}_{p+2}F_{q+2}\begin{bmatrix}\mathbf{a},m+L+3/2,m+1;vr^2\\ \mathbf{c},m+L+3/2-v,m+1-\omega\end{bmatrix}. \tag{65}$$

If $\beta=2m$ and $0\leqslant m<\omega$, then

$$N(r^2/2)r^{2m}\,{}_pF_q\begin{bmatrix}\mathbf{a};vr^2\\ \mathbf{c}\end{bmatrix}={}^{m}_{p+2}T_{q+2}(\mathbf{a},\mathbf{c};v,r^2), \tag{66}$$

where

$${}^{m}_{p+2}T_{q+2}=2^{\omega}v^{\omega-m}(\lambda+\tfrac{3}{2})_v\,[\omega!(\mathbf{a})_{\omega-m}/(\omega-m)!(\mathbf{c})_{\omega-m}]\times{}_{p+2}F_{q+2}\begin{bmatrix}\mathbf{a}+\omega-m,\omega+L+\tfrac{3}{2},\omega+1;vr^2\\ \mathbf{c}+\omega-m,\lambda+\tfrac{3}{2},\omega+1-m\end{bmatrix}, \tag{67}$$

where the symbol $\mathbf{a}+k$ denotes the set of numbers a_1+k, $a_2+k,\ldots,a_p+k$.

If $\beta=2m$ and $m\leqslant -1$, then

$$N(r^2/2)r^{2m}\,{}_pF_q\begin{bmatrix}\mathbf{a};vr^2\\ \mathbf{c}\end{bmatrix}={}^{m}_{p+2}S_{q+2}+{}^{m}_{p+2}T_{q+2}. \tag{68}$$

If $\beta=2m+1$ and $m\geqslant v-L-1$, then

$$N(r^2/2)r^{2m+1}\,{}_pF_q\begin{bmatrix}\mathbf{a};vr^2\\ \mathbf{c}\end{bmatrix}={}^{m}_{p+2}U_{q+2}(\mathbf{a},\mathbf{c};v,r^2), \tag{69}$$

where

$${}^{m}_{p+2}U_{q+2}=2^{\omega}r^{2m-2\omega+1}\times(m+\tfrac{3}{2}-\omega)_{\omega}(-1)^{v}(-m-L-1)_v\times{}_{p+2}F_{q+2}\begin{bmatrix}\mathbf{a},m+L+2,m+\tfrac{3}{2};vr^2\\ \mathbf{c},m+L+2-v,m+\tfrac{3}{2}-\omega\end{bmatrix}. \tag{70}$$

If $\beta=2m+1$ and $-L-1\leqslant m\leqslant v-L-2$, then

$$N(r^2/2)r^{2m+1}\,{}_pF_q\begin{bmatrix}\mathbf{a};vr^2\\ \mathbf{c}\end{bmatrix}={}^{m}_{p+2}V_{q+2}(\mathbf{a},\mathbf{c};v,r^2), \tag{71}$$

where

$${}^{m}_{p+2}V_{q+2}=2^{\omega}v^{v-m-L-1}r^{-2\lambda-1}\frac{v!(-\lambda+\frac{1}{2})_{\omega}(\mathbf{a})_{v-m-L-1}}{(v-m-L-1)!(\mathbf{c})_{v-m-L-1}}\,{}_{p+2}F_{q+2}\begin{bmatrix}\mathbf{a}+v-m-L-1,v+1,v-L+\tfrac{1}{2};vr^2\\ \mathbf{c}+v-m-L-1,v-m-L,-\lambda+\tfrac{1}{2}\end{bmatrix}, \tag{72}$$

If $\beta=2m+1$ and $m\leqslant -L-2$, then

$$N(r^2/2)r^{2m+1}\,{}_pF_q\begin{bmatrix}\mathbf{a};vr^2\\ \mathbf{c}\end{bmatrix}={}^{m}_{p+2}U_{q+2}+{}^{m}_{p+2}V_{q+2}, \tag{73}$$

The more complicated case of GHS ${}_pF_q\,(vr)$ can be reduced to the above equations. To this end, taking account of the remarks exposed in Sec. IV, we shall write the series ${}_pF_q(vr)$ in the following form[14]:

$${}_pF_q\begin{bmatrix}\mathbf{a};vr\\ \mathbf{c}\end{bmatrix}=\sum_{t=0}^{1}\frac{(\mathbf{a})_t}{(\mathbf{c})_t}\frac{(vr)^t}{t}\,{}_{2p}F_{2q+1}\begin{bmatrix}\mathbf{A}(t);Vr^2\\ \mathbf{C}(t)\end{bmatrix}, \tag{74}$$

where

$$\mathbf{A}(t)=\{(\mathbf{a}+t)/2,(\mathbf{a}+t+1)/2\}, \tag{75}$$

$$\mathbf{C}(t)=\{(\mathbf{c}+t)/2,(\mathbf{c}+t+1)/2,t+\tfrac{1}{2}\}, \tag{76}$$

$$V=4^{p-q+1}v^2. \tag{77}$$

Then,

$$N(r^2/2)r^{\beta}\,{}_pF_q\begin{bmatrix}\mathbf{a};vr\\ \mathbf{c}\end{bmatrix}=N(r^2/2)r^{\beta}\,{}_{2p}F_{2q+1}\begin{bmatrix}\mathbf{A}(0);Vr^2\\ \mathbf{C}(0)\end{bmatrix}+v\frac{(\mathbf{a})_1}{(\mathbf{c})_1}N(r^2/2)r^{\beta+1}\,{}_{2p}F_{2q+1}\begin{bmatrix}\mathbf{A}(1);Vr^2\\ \mathbf{C}(1)\end{bmatrix}. \tag{78}$$

For any integral β both even and odd powers of r are present on the right-hand side of Eq. (78). Therefore, Eq. (78) may assume more special forms compared to the expression (58). We shall not present them explicitly here since their derivation can be easily performed by the direct use of Eqs. (64)–(73).

Note that in many important particular cases the series ${}_{p+2}F_{q+2}$ on the right-hand sides of Eqs. (65), (67), (70), and (72) reduce to simpler form. This may take place, for example, due to cancellation of numerator and denominator parameters. In the particular case $m=0$ the series ${}_{p+2}F_{q+2}$ in Eq. (67) reduces to ${}_{p+1}F_{q+1}$ which is equivalent to the result given in Ref. 4.

IX. CONCLUSIONS

It is shown that the operator $N(r^2/2)$ [Eq. (8)] occurring in the general gradient formula (6) allows simpler expressions in terms of the Pochhammer symbols depending on operators $\delta(r^2)$. This result seems to be important for the "hypergeometric" form of $N(r^2/2)$ proves to be consistent with those structures which have to be dealt with, usually, in physical and quantum chemical applications. Two kinds of problems have been found to be the most typical, the operators N being either involved with summation of the hypergeometric type or applied to a certain hypergeometric sum. In both these cases the operator N is easily transformed in such a way that explicit algebraic expressions follow immediately for each of these two cases.

In particular, this gives a very compact expression for the Laplace-type series coefficients through the operator Gauss functions ${}_2F_1$ [see Eq. (45)]. In turn, remarkably short proofs of two important Sack's expansions[1,12] are implied by this observation.

Another important corollary is that the gradient formula (5), involving $\mathscr{Y}^n_{lm}((r))$, can be readily generalized for the case of the more complicated operator ${}_pF_q(t\Delta(\mathbf{r})) \times \mathscr{Y}^0_{lm}(\nabla(\mathbf{r}))$ [see Eq. (54)] which one may expect to be involved in a more complicated Laplace-type series compared to that in Eq. (4).

Explicit expressions for $N(r^2/2)\varphi(r)$ have been obtained for the two cases: $\varphi(r) = r^\beta {}_pF_q(vr^2)$ and $\varphi(r) = r^\beta \times {}_pF_q(vr)$. With due respect to Eqs. (4)–(6), these formulas give the Laplace-type expansions for a wide class of spherical tensors $f_{LM}(\mathbf{r}+\mathbf{R})$ whose radial part is represented by generalized hypergeometric series.

One may easily anticipate the possibility of generalizing the approach for the case of an IST depending on N vector arguments. In the case $N=2$ this should give, apparently, an alternative (differential) representation for the Talmi–Smirnov coefficients.[15]

One should realize that convergence properties are to be investigated prior to (or, at least, posterior to) using the gradient formulas. Only terminating series have to be dealt with, however, in a number of physically important cases. The direct use of the "gradient" expressions for coefficients is, therefore, possible, the convergence problem being, evidently, eliminated for these cases.

Note also that the Laplace-type series, which are implied by the formulas given in Sec. VIII, have a formal character in case of a negative power exponent β. Indeed, additional generalized function terms should be incorporated in the right-hand parts of the corresponding expansions, by analogy with the case of $\mathscr{Y}^n_{lm}(\mathbf{r}+\mathbf{R})$ (see Ref. 16), provided that $\beta \leqslant -3$.

Note added in proof: An interesting and suggestive account of connections between expansion theorems and nonclassical integrals of Bessel functions, interpreted as derivatives of generalized functions, has been given by S. N. Stuart [J. Austral. Math. Soc., Ser. B **22**, 968 (1981)]. He also presented an ancillary theorem which expresses the gradient $\mathscr{Y}^n_{lm}(\nabla)$ of an IST $f_{LM}(\mathbf{r})$ in a form that separates angular and radial variables. The gradient formulas given in our paper seem to provide more flexible alternatives to Stuart's result, and these also seem to be more suitable for generalizations (see Secs. V–VIII). These remarks are likely to hold true with respect to several interesting differential formulas obtained by E. J. Weniger and E. O. Steinborn [J. Math. Phys. **24**, 2533 (1983)].

ACKNOWLEDGMENTS

The author acknowledges Professor L. A. Gribov's comments which contributed to a more transparent presentation of the two types of problems exposed in Sec. IV. The author is very indebted to Professor Yu. F. Smirnov for stimulating discussions of possible applications of the formulas given in the paper.

[1]R. Sack, J. Math. Phys. **5**, 245 (1964).
[2]T. Sawaguri and W. Tobocman, J. Math. Phys. **8**, 2223 (1967).
[3]B. F. Bayman, J. Math. Phys. **19**, 2558 (1978).
[4]A. W. Niukkanen, J. Math. Phys. **24**, 1989 (1983).
[5]D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonsky, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975) (in Russian).
[6]V. P. Maslov, *Operator Methods* (Nauka, Moscow, 1973) (in Russian).
[7]A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. II: (a) Sec. 10.13, Eq. (7); (b) Sec. 10.8, Eq. (10); (c) Sec. 10.12, Eq. (5); (d) Sec. 10.9, Eq. (20); (e) Sec. 11.2, Eq. (23); (f) Sec. 7.2, Eq. (3).
[8]Due to a different choice of generating functions, the definition of $H_n(x)$ in Ref. 7(a) is equivalent to $2^{n/2} H_n(\sqrt{2}x)$ in our notations.
[9]See, for instance, Eqs. (3.1) and (3.2) in E. Filter and E. O. Steinborn, J. Math. Phys. **19**, 79 (1978).
[10]See Ref. 7, Vol. I: (a) Sec. 2.1.2; (b) Sec. 4.2, Eq. (2); (c) Sec. 5.12, Eq. (1); (d) Sec. 2.1.3, Eq. (14); (e) Sec. 1.2, Eq. (11); (f) Sec. 2.11, Eqs. (32)–(36); (g) Sec. 4.1, Eq. (1).
[11]This implies that $x^n d^n$ commutes with $x^m d^m$.
[12]R. Sack, J. Math. Phys. **5**, 252 (1964).
[13]Equation (51) differs from the definition in Ref. 7(e) in normalization factor.
[14]Y. L. Luke, *Mathematical Functions and Their Approximations* (Academic, New York, 1975), Sec. 5.2.1, Eq. (15).
[15]Yu. F. Smirnov, Nucl. Phys. **27**, 177 (1961); **39**, 346 (1962).
[16]K. G. Kay, H. D. Todd, and H. J. Silverstone, J. Chem. Phys. **51**, 2359 (1969).

On the polynomial solutions of ordinary differential equations of the fourth order

J. S. Dehesa, E. Buendia, and M. A. Sanchez-Buendia
Departamento de Física Nuclear, Facultad de Ciencias, Universidad de Granada, Granada, Spain

(Received 12 October 1983; accepted for publication 23 November 1984)

The density of zeros of polynomial solutions of ordinary differential equations of the fourth order with coefficients depending only on the independent variable is analyzed. The first four moments of such a density are given directly in terms of the coefficients which characterize the differential operator. Application to the nonclassical orthogonal polynomials corresponding to the names Krall–Legendre, Krall–Laguerre, and Krall–Jacobi is done. Global asymptotic properties of the zeros of these polynomials are also obtained.

I. INTRODUCTION

The determination of the distribution of zeros of (orthogonal) polynomials is a very relevant problem in theoretical physics. One reason is that often the Hamiltonian of a physical system may be transformed by means of a Lanczos-like method[1,2] into a Jacobi matrix (i.e., a real, symmetric tridiagonal matrix). The characteristic polynomials of the principal submatrices of a Jacobi matrix form[3] a system of orthogonal polynomials which satisfy a three-term recursion relation whose coefficients are the entries of the matrix. Then the calculation of the eigenvalue density of the physical Hamiltonian reduces to the problem of determining the density of zeros of such polynomials starting from the coefficients of the above-mentioned recurrent relation. This problem has been solved by Dyson[4] and Dean,[5] and by Dehesa[6] and Nevai[7] using very different methods.[8]

Another reason is that in a large number of physical problems the eigenfunctions of ordinary differential operators (e.g., Schrödinger operators) turn out to be orthogonal polynomials once one has separated out their behavior at the infinite and at singular points. Then the determination of the nodes of the eigenfunctions reduces to the problem of calculating the zeros of certain polynomials. Here we are interested in knowing the density of the zeros of these polynomials starting from the coefficients which characterize the corresponding differential equation.

Recently, Case[9] has found by very elementary means a method to obtain sum rules for the powers of the zeros of polynomials satisfying ordinary differential equations and subject to the two following conditions: (a) the zeros are simple and (b) the coefficient to the ith derivative is itself a polynomial of degree not greater than i. These restrictions are fulfilled by a large class of families of polynomials which play an important role in physics, e.g., the classical orthogonal polynomials.

Here we will use Case's method to investigate the distribution of zeros of all the systems of orthogonal polynomials, except the classical ones, which are solutions of an ordinary differential operator of the fourth order and with the above-mentioned restrictions. We consider the normalized density distribution function of zeros $\rho_N(x) = (1/N)\Sigma_{n=1}^{N}\delta(x - x_n)$ of the polynomial $P_N(x)$. The moments about the origin of this function are

$$\mu_r' = \frac{1}{N}\sum_{i=1}^{N} x_i^r \equiv \frac{1}{N} y_r.$$

Our purpose is the evaluation of the first few moments (μ_r'; $r = 0,1,2,3,4$) in terms of the coefficients characterizing the differential equation fulfilled by the polynomials. It is known that these quantities give the behavior of the density around the mean and often allow a good approximate description of this function all over the interval of definition by means of one of the various available parametrizations[10–12] (see Charlier, Weibull, etc.).

The structure of the paper is as follows: In Sec. II we briefly summarize the main result, appropriately corrected, of Ref. 9 for the case where the differential equation is of the fourth order. Other nontrivial results not found by Case but obtained with this method are given since they are needed later on. For the sake of clarity the proofs of these new results are postponed to the Appendices. The remaining sections are devoted to applying the previous results to the nonclassical orthogonal polynomials[13,14] identified with the names Krall–Legendre, Krall–Laguerre, and Krall–Jacobi. These polynomials, which are closely connected with the Legendre, Laguerre, and Jacobi polynomials, are orthogonal[14,15] with respect to Stieltjes weight functions, which are absolutely continuous on $[-1,1]$, $[0,\infty]$, and $[0,1]$, respectively, but which have jumps at some of the intervals' ends. Apart from the classical orthogonal polynomials which satisfy the square of their second-order equation, these polynomials are[16] the only Chebyshev sets which fulfill a fourth-order differential equation of the type (1). Finally some concluding remarks are given. Case's notation is used throughout the paper.

II. METHOD

Let us assume that our polynomials $P_N(x)$ satisfy the differential equation

$$\sum_{i=0}^{4} g_i(x) P_N^{(i)}(x) = 0, \tag{1}$$

where

$$g_i(x) = \sum_{j=0}^{i} a_j^{(i)} x^j, \quad i = 0,1,2,3,4, \tag{2}$$

with constant coefficients $a_j^{(i)}$. It is important to point out

 0022-2488/85/071547-06$02.50

that all the coefficients $a_j^{(i)}$, except $a_0^{(0)}$, are independent of N.

Assuming further that all the zeros of $P_N(x)$ are simple, then the following basic relation has been shown[9]:

$$\sum_{i=2}^{4} i \sum_{j=0}^{i} a_j^{(i)} J_{r+j}^{(i)} = -a_0^{(1)} y_r - a_1^{(1)} y_{r+1}, \tag{3}$$

for $r = 0,1,2,\ldots$, and

$$J_r^{(i)} = \sum_{\neq} \frac{x_{l_1}^r}{(x_{l_1} - x_{l_2})(x_{l_1} - x_{l_3})\cdots(x_{l_1} - x_{l_i})}, \tag{4}$$

where $\Sigma_{\neq}$ means to sum over all l's subject to none of them being equal. Here we avoid the erroneous factorial $i!$ of Eq. (5) of Ref. 9. In spite of this error the sum rules of zeros of classical orthogonal polynomials calculated by Case[9] are correct since such polynomials satisfy a differential equation of order $i = 2$.

It can be shown that the quantities $J_r^{(i)}$ can be expressed in terms of the y_s. Thus, ultimately, Eq. (3) is a recurrence relation which permits us to evaluate the y_s, hence the moments μ_s' of the density distribution of zeros of the polynomial $P_N(x)$. We remark that $a_0^{(0)}$ is the only coefficient which does not play any role for the calculation of the moments.

Therefore the first step here is to calculate the $J_r^{(i)}$, at least for $i = 2,3,4$ according to Eq. (3), in terms of the y_s. In Ref. 9 some results relevant for our purpose were found. They are as follows:

$$J_r^{(2)} = \left(N - \frac{r}{2}\right) y_{r-1} + \frac{1}{2} \sum_{s=1}^{r-2} y_{r-1-s}\, y_s, \tag{5}$$

$$J_r^{(i)} = 0, \quad 0 \leqslant r \leqslant i-2, \tag{6a}$$

$$J_{i-1}^{(i)} = \frac{1}{i} \prod_{t=0}^{i-1} (N-t), \tag{6b}$$

$$J_i^{(i)} = y_1 \prod_{t=1}^{i-1} (N-t), \tag{6c}$$

$$J_{i+1}^{(i)} = \left[\prod_{t=2}^{i-1}(N-t)\right]\left[\left(N - \frac{i+1}{2}\right) y_2 + \frac{i-1}{2} y_1^2\right], \tag{6d}$$

and that for $r \geqslant i$, $J_r^{(i)}$ depends only on the y_s for $s \leqslant r+1-i$. The dependence of y_{r+1-i} is linear.

With these data in mind and following the lines of that reference, we have proved that for $i = 3$ (see Appendix A)

$$J_r^{(3)} = \left(N^2 - rN + \frac{r(r-1)}{3}\right) y_{r-2} + \sum_{t=2}^{r-2} (N-t)\, y_{t-1}\, y_{r-1-t}(1-\delta_{r,3}) + \frac{1}{3} \sum_{s=3}^{r-2} \sum_{t=1}^{s-2} y_{r-1-s}\, y_{s-1-t}\, y_t (1-\delta_{r,3})(1-\delta_{r,4}), \tag{7}$$

valid for $r = 3,4,5,\ldots$. In particular one has

$$J_5^{(3)} = (N^2 - 5N + 20/3) y_3 + (2N-5) y_2 y_1 + \tfrac{1}{3} y_1^3, \tag{8a}$$

$$J_6^{(3)} = (N^2 - 6N + 10) y_4 + (2N-6) y_3 y_1 + (N-3) y_2^2 + y_2 y_1^2. \tag{8b}$$

(Here we give only the explicit expressions of the $J_r^{(3)}$ not given by Ref. 9 but needed here later on.)

Also, in Appendix B it is proved that for $i = 4$,

$$\begin{aligned} J_r^{(4)} = \frac{1}{4} \Bigg\{ &[(4N^3 - 6rN^2 + \langle 4[(r-1)(r-2)+2] + 6(r-2)\rangle N - r(r-1)(r-2)]\, y_{r-3} + \sum_{t=3}^{r-2} [6N^2 - 6(2t-1)N \\ &+ \langle 4t(t-1) + 3(t-1)(r-t)\rangle]\, y_{r-1-t}\, y_{t-2}(1-\delta_{r,4}) + \sum_{t=4}^{r-2} \sum_{s=2}^{t-2} (4N-6s)\, y_{r-1-t}\, y_{t-1-s}\, y_{s-1}(1-\delta_{r,4})(1-\delta_{r,5}) \\ &+ \sum_{t=5}^{r-2} \sum_{s=3}^{t-2} \sum_{u=1}^{s-2} y_{r-1-t}\, y_{t-1-s}\, y_{s-1-u}\, y_u (1-\delta_{r,4})(1-\delta_{r,5})(1-\delta_{r,6}) \Bigg\}, \end{aligned} \tag{9}$$

valid for $r = 4,5,6,\ldots$. In particular one has

$$J_6^{(4)} = (N-3)(N^2 - 6N + 10) y_3 + 3(N^2 - 6N + 9) y_2 y_1 + (N-3) y_1^3, \tag{10a}$$

$$J_7^{(4)} = \tfrac{1}{4}[2(2N-7)(N^2 - 7N + 15) y_4 + 4(3N^2 - 21N + 38) y_3 y_1 + 3(2N^2 - 14N + 25) y_2^2 + 6(2N-7) y_2 y_1^2 + y_1^4]. \tag{10b}$$

(Again, here we write the needed $J_r^{(4)}$ not found in Ref. 9.)

III. THE KRALL–LEGENDRE ORTHOGONAL POLYNOMIALS

These polynomials satisfy the following fourth-order equation[13,14]:

$$\begin{aligned} &(x^2-1)^2 P_N^{\mathrm{iv}}(x) + 8x(x^2-1) P_N^{\mathrm{iii}}(x) + (4\alpha+12)(x^2-1) P_N^{\mathrm{ii}}(x) + 8\alpha x P_N^{\mathrm{i}}(x) \\ &\quad - [8\alpha N + (4\alpha+12) N(N-1) + (8 + (N-3)) N(N-1)(N-2)] P_N(x) = 0, \end{aligned} \tag{11}$$

which is a differential equation of the type (1) and (2) with coefficients

$$\begin{array}{lllll} a_1^{(1)} = 8\alpha, & a_0^{(1)} = 0, & & & \\ a_2^{(2)} = 4\alpha + 12, & a_1^{(2)} = 0, & a_0^{(2)} = -(4\alpha+12), & & \\ a_3^{(3)} = 8, & a_2^{(3)} = 0, & a_1^{(3)} = -8, & a_0^{(3)} = 0, & \\ a_4^{(4)} = 1, & a_3^{(4)} = 0, & a_2^{(4)} = -2, & a_1^{(4)} = 0, & a_0^{(4)} = 1. \end{array} \tag{12}$$

The Krall–Legendre polynomials are orthogonal[14–15] on $[-1, +1]$ with respect to the Stieltjes weight function

$$w(x) = \tfrac{1}{2}(\delta(x-1) + \delta(x+1)) + \alpha/2.$$

Taking the values (12) to Eq. (3) one obtains

$$\begin{aligned} y_{r+1} &= [(\alpha+3)/\alpha](J_r^{(2)} - J_{r+2}^{(2)}) + (3/\alpha)(J_{r+1}^{(3)} - J_{r+3}^{(3)}) \\ &\quad - (1/2\alpha)(J_r^{(4)} - 2J_{r+2}^{(4)} + J_{r+4}^{(4)}), \end{aligned} \tag{13}$$

which for $r = 0,1,2,\ldots$ allows us to calculate the moments $\mu_s' = y_s/N$ of the density of zeros of the polynomial $P_N(x)$ directly in terms of N and α.

Indeed, with $r = 0$ this equation reduces as follows:

$$y_1 = -\frac{\alpha+3}{\alpha} J_2^{(2)} - \frac{3}{\alpha} J_3^{(3)} - \frac{1}{2\alpha} J_4^{(4)},$$

where the property (6a) was already used. Further, the property (6c) says that $J_i^{(i)}$ is proportional to y_1, hence it is clear that

$$\mu_1' = y_1/N = 0.$$

This should not surprise us since the Krall–Legendre polynomials are either odd or even. Therefore any moment of odd order vanishes. With $r = 1$, Eq. (13) gives

$$\begin{aligned} y_2 &= [(\alpha+3)/\alpha](J_1^{(2)} - J_3^{(2)}) + (3/\alpha)(J_2^{(3)} - J_4^{(3)}) \\ &\quad + (1/2\alpha)(2J_3^{(4)} - J_5^{(4)}). \end{aligned}$$

By means of Eq. (6b) one knows that

$$J_1^{(2)} = N(N-1)/2,$$
$$J_2^{(3)} = N(N-1)(N-2)/3,$$
$$J_3^{(4)} = N(N-1)(N-2)(N-3)/4,$$

and from Eq. (6d)

$$J_3^{(2)} = (N - \tfrac{3}{2})\, y_2 + y_1^2/2,$$
$$J_4^{(3)} = (N-2)[(N-2)\, y_2 + y_1^2],$$
$$J_5^{(4)} = (N-2)(N-3)[(N - \tfrac{5}{2})\, y_2 + 3y_1^2/2].$$

With this help, it is trivial to find that

$$\mu_2' = \frac{1}{N} y_2 = \frac{N^3 - 2N^2 + (2\alpha+5)N - (2\alpha+4)}{2N^3 - 3N^2 + (4\alpha+1)N - 2\alpha}. \tag{14}$$

To calculate μ_4', we work in a similar way. Now, the value $r = 2$ in Eq. (13) allows us to obtain the following:

$$\begin{aligned} y_4 &= [(\alpha+3)/\alpha](J_3^{(2)} - J_5^{(2)}) + (3/\alpha)(J_4^{(3)} - J_6^{(3)}) \\ &\quad - (1/2\alpha)(J_3^{(4)} - 2J_5^{(4)} + J_7^{(4)}). \end{aligned}$$

Taking into account that $J_5^{(2)} = (N - \tfrac{5}{2})\, y_4 + y_2^2/2$ due to Eq. (5) and the values (8b) and (10b) for $J_6^{(3)}$ and $J_7^{(4)}$, one has after regrouping terms that

$$\mu_4' = N^{-1} y_4 = A(N,\alpha)/B(N,\alpha), \tag{15}$$

where

$$\begin{aligned} A(N,\alpha) &= -\tfrac{1}{4}(N^3 - 6N^2 + 11N - 6) \\ &\quad + 2N^3 - 9N^2 + (19 + 2\alpha)N \\ &\quad - (15 + 3\alpha)\mu_2' - [\tfrac{3}{2}N^3 - \tfrac{9}{2}N^2 + (\tfrac{15}{4} + \alpha)N]\mu_2'^2, \end{aligned}$$

$$B(N,\alpha) = N^3 - \tfrac{9}{2}N^2 + (\tfrac{19}{2} + 2\alpha)N - (\tfrac{15}{2} + 3\alpha).$$

For very large N, Eqs. (14) and (15) say that $\mu_2' = \tfrac{1}{2} + O(N^{-1})$ and $\mu_4' = \tfrac{3}{8} + O(N^{-1})$, respectively. Furthermore it is possible to prove[17] by means of the Nevai–Dehesa theorem,[7] which starts from the three-term recurrence relation on the polynomials, that

$$\mu_{2k}' = \frac{1}{2^{2k}}\binom{2k}{k} + O(N^{-1}),$$

for $k = 1,2,3,\ldots$. Of course $\mu_{2k-1}' = 0$. This indicates that the asymptotic $(N \to \infty)$ density of zeros of the Krall–Legendre polynomials is an inverted semicircular function on $[-1, +1]$.

IV. THE KRALL–LAGUERRE POLYNOMIALS

These polynomials are defined by a differential equation of the form (1) and (2) with the following coefficients[16]:

$$\begin{aligned} &a_0^{(0)} = -[(2R+2)N + N(N-1)], \\ &a_1^{(1)} = 2R+2, \quad a_0^{(1)} = -2R, \\ &a_2^{(2)} = 1, \quad a_1^{(2)} = -2R-6, \quad a_0^{(2)} = 0, \\ &a_3^{(3)} = 0, \quad a_2^{(3)} = -2, \quad a_1^{(3)} = 4, \quad a_0^{(3)} = 0, \\ &a_4^{(4)} = 0, \quad a_3^{(4)} = 0, \quad a_2^{(4)} = 1, \quad a_1^{(4)} = 0, \quad a_0^{(4)} = 0. \end{aligned} \tag{16}$$

The Krall–Laguerre polynomials are orthogonal[16] on $[0, \infty)$ with respect to the following Stieltjes weight function:

$$w(x) = (1/R)\delta(x) + e^{-x}.$$

Now, we want to calculate the moments of the density of zeros of the Krall polynomial of Laguerre type of degree N as a function of R. First of all, we take the values (16) of the coefficients $a_j^{(i)}$ to Eq. (3). For $r = 0,1,2,\ldots$, the result is the following recursion relation:

$$(2R+2)\, y_{r+1} = 2R\, y_r + 2[(2R+6)J_{r+1}^{(2)} - J_{r+2}^{(2)}] - 6(2J_{r+1}^{(3)} - J_{r+2}^{(3)}) - 4J_{r+2}^{(4)}. \tag{17}$$

With $r = 0$, this equation reduces as follows:

$$(2R+2)\, y_1 = 2RN + 2[(2R+6)J_1^{(2)} - J_2^{(2)}] + 6J_2^{(3)},$$

where we have already used the fact that $y_0 = N$ and $J_1^{(3)} = 0$ and $J_2^{(4)} = 0$ due to Eq. (6a). Since the values of the other J's are known, one obtains

$$\mu_1' = (N^2 + RN - 1)/(N + R). \tag{18}$$

Hence, $\mu_1' = N + O(N^0)$ for large values of N.

Now, Eq. (17) with $r = 1$ gives

$$(2R+2)y_2 = 2Ry_1 + 2[(2R+6)J_2^{(2)} - J_3^{(2)}] - 6(2J_2^{(3)} - J_3^{(3)}) - 4J_3^{(4)}.$$

Operating here with the known values of J's, we obtain

$$\mu_2' = \frac{-(N^3 - 2N^2 - N + 2) + [6N^2 + (4R-6)N - 2R]\mu_1' - N\mu_1'^2}{2(N+R) - 1}. \tag{19}$$

We remark that $\mu_2' = 2N^2 + O(N)$ for large values of N.

With $r = 2$, Eq. (17) gives

$$(2R+2)y_3 = 2Ry_2 + 2[(2R+6)J_3^{(2)} - J_4^{(2)}] - 6(2J_3^{(3)} - J_4^{(3)}) - 4J_4^{(4)},$$

which, with the value (18) for y_2 and taking into account the relations (6b)–(6d), reduces as follows:

$$\mu_3' = [-N\mu_1'\mu_2' + [3N^2 + (2R-6)N - 2R + 3]\mu_2' + N(3N + R - 3)\mu_1'^2 - (2N^3 - 6N^2 + 4N)\mu_1'](N + R - 1)^{-1}, \tag{20}$$

which allows us to evaluate μ_2' in terms of N and R since μ_1' and μ_2' are already given in Eqs. (18) and (19), respectively. From this equation and the asymptotic values of μ_1' and μ_2', it is possible to obtain the relation $\mu_3' = 5N^3 + O(N^2)$ when N is very large.

In a similar way but with $r = 3$ in Eq. (17) and the help of the relation (8a), we obtain the following recurrence relation for μ_4':

$$\mu_4' = D(N,R)/[2(N+R) - 3], \tag{21}$$

with

$$\begin{aligned} D(N,R) = {} & -N\mu_2'^2 - 2N\mu_1'\mu_3' + (6N^2 + (4R-18)N + 16 - 6R)\mu_3' \\ & + (12N + 4R - 18)N\mu_1'\mu_2' - (4N^3 - 18N^2 + 26N - 12)\mu_2' + 2N^2\mu_1'^3 - (6N^2 - 18N + 12)N\mu_1'^2 \end{aligned}$$

which produces the result $\mu_4' = 14N^4 + O(N^3)$ for very large N. In general it can be proved[17] that

$$\mu_r' = \frac{1}{r+1}\binom{2r}{r}N^r + O(N^{r-1}).$$

V. THE KRALL–JACOBI POLYNOMIALS

These polynomials[16] verify the differential equation (1) and (2) with the coefficients

$$\begin{aligned} & a_0^{(0)} = -[(\alpha+2)(2\alpha+2+2M)N + (\alpha^2 + 9\alpha + 14 + 2M)N(N-1) + 2(\alpha+4)N(N-1)(N-2) + N(N-1)(N-2)(N-3)], \\ & a_1^{(1)} = (\alpha+2)(2\alpha+2+2M), \quad a_0^{(1)} = -2M, \\ & a_2^{(2)} = \alpha^2 + 9\alpha + 14 + 2M, \quad a_1^{(2)} = -6\alpha - 12 - 2M, \quad a_0^{(2)} = 0, \\ & a_3^{(3)} = 2\alpha + 8, \quad a_2^{(3)} = -2\alpha - 12, \quad a_1^{(3)} = 4, \quad a_0^{(3)} = 0, \\ & a_4^{(4)} = 1, \quad a_3^{(4)} = -2, \quad a_2^{(4)} = 1, \quad a_1^{(4)} = 0, \quad a_0^{(4)} = 0. \end{aligned} \tag{22}$$

The Krall–Jacobi polynomials are orthogonal[16] on [0,1] with respect to the Stieltjes weight function

$w(x) = 1/M\delta(x) + (1-x)^\alpha; \alpha > -1.$

Our purpose is to calculate the first four moments of the density of zeros of the Krall–Jacobi polynomial of degree N directly in terms of the two parameters α and M which characterize it. As in the previous two sections we will use Eq. (3), which for this case reduces as follows:

$$\begin{aligned} 2(\alpha+2)(\alpha+M+1)y_{r+1} = {} & 2My_r + 2[(6\alpha + 12 + 2M)J_{r+1}^{(2)} - (\alpha^2 + 9 + 14 + 2M)J_{r+2}^{(2)}] \\ & - 3[4J_{r+1}^{(3)} - (2\alpha+12)J_{r+2}^{(3)} + (2\alpha+8)J_{r+3}^{(3)}] - 4[J_{r+2}^{(4)} - 2J_{r+3}^{(4)} + J_{r+4}^{(4)}], \end{aligned} \tag{23}$$

for $r = 0,1,2,\ldots$. With $r = 0$ and the relations (6a)–(6d), this equation gives the moment about the origin of the first order

$$\mu_1' = [N^3 + \alpha N^2 + (M-1)N - \alpha]/[2N^3 + 3\alpha N^2 + (\alpha^2 + 2M)N + \alpha M]. \tag{24}$$

With $r = 1$ and the same relations (6a)–(6d), Eq. (23) reduces to the following recurrence relation for the second moment:

$$\mu_2' = F(\alpha,M,N)/G(\alpha,M,N), \tag{25}$$

with

$$\begin{aligned} F(\alpha,M,N) = {} & -(N^3 - 2N^2 - N + 2) + [8N^3 + (6\alpha - 12)N^2 + (4 - 6\alpha + 4M)N - 2M]\mu_1' \\ & - N[6N^2 + (6\alpha - 6)N + 2 - 3\alpha + \alpha^2 + 2M]\mu_1'^2, \\ G(\alpha,M,N) = {} & 4N^3 + 6(\alpha - 1)N^2 + (6 - 6\alpha + 2\alpha^2 + 4M)N + (2\alpha M - 2M - \alpha^2 + 3\alpha - 2). \end{aligned}$$

Putting $r = 2$ and $r = 3$ in Eq. (23), we obtain the expressions of the third and fourth moments, respectively, as functions of N, α, and M. Their evaluation is straightforward but a bit more involved and cumbersome since we have to play also with the values (8a) and (8b) for the quantities $J_5^{(3)}$ and $J_6^{(3)}$, respectively, and (10a) and (10b) for $J_6^{(4)}$ and $J_7^{(4)}$, also respectively. For this reason, we give only the recurrence expressions for these moments. The third moment is given as follows:

$$\mu_3' = P(\alpha,M,N)/Q(\alpha,M,N), \tag{26}$$

with

$$\begin{aligned} P(\alpha,M,N) = & -[4N^3 - 12N^2 + 8N]\mu_1' + [12N^2 + (6\alpha - 24)N + 12 - 6\alpha + 2M]N\mu_1'^2 \\ & + [8N^3 + (6\alpha - 24)N^2 + (28 - 12\alpha + 4M)N - 12 + 6\alpha - 4M]\mu_2' \\ & - [12N^2 + (12\alpha - 24)N + 16 - 12\alpha + 2\alpha^2 + 4M]N\mu_1'\mu_2' - [4N + 2\alpha - 4]N^2\mu_1'^3, \end{aligned}$$

$$Q(\alpha,M,N) = 4N^3 + 6(\alpha - 2)N^2 + (20 - 2\alpha^2 - 12\alpha + 4M)N + 12 - 2\alpha^2 + 10\alpha + 2\alpha M - 4M,$$

and the fourth moment is expressed by

$$\mu_4' = R(\alpha,M,N)/S(\alpha,M,N),$$

with

$$\begin{aligned} R(\alpha,M,N) \\ = & -[6N^2 - 18N + 12]N\mu_1'^2 - [4N^3 - 18N^2 + 26N - 12]\mu_2' + [8N + 2\alpha - 12]N^2\mu_1'^3 + [24N^2 + (12\alpha \\ & - 72)N + 60 - 18\alpha + 4M]N\mu_1'\mu_2' - [12N + 6\alpha - 18]N^2\mu_1'^2\mu_2' + [8N^3 + (6\alpha - 36)N^2 + (68 - 18\alpha \\ & + 4M)N - 48 + 16\alpha - 6M]\mu_3' - [12N^2 + (12\alpha - 36)N + 36 - 18\alpha + 2\alpha^2 + 4M]N\mu_1'\mu_3' - [6N^2 + (6\alpha \\ & - 18)N + 17 - 9\alpha + \alpha^2 + 2M]N\mu_2'^2 - N^3\mu_1'^4, \end{aligned}$$

$$S(\alpha,M,N) = 4N^3 + 6(\alpha - 3)N^2 + (42 - 18\alpha + 2\alpha^2 + 4M)N - (36 - 21\alpha + 3\alpha^2 - 2\alpha M + 6M).$$

From Eqs. (24)–(27) we can obtain the values of the first four moments for very large values of N in a straightforward manner. They are as follows:

$$\mu_1' = \tfrac{1}{2} + O(N^{-1}), \quad \mu_2' = \tfrac{3}{8} + O(N^{-1}),$$
$$\mu_3' = \tfrac{5}{16} + O(N^{-1}), \quad \mu_4' = \tfrac{35}{128} + O(N^{-1}).$$

These values seem to indicate that

$$\mu_r' = \frac{1}{2^{2r}}\binom{2r}{r} + O(N^{-1}); \quad r = 1,2,3,\ldots,$$

which are the moments of an inverted semicircular density.

The comparison of these values with the corresponding asymptotic quantities of other orthogonal polynomials permits us to make some interesting observations.

(a) The first two moments of the zeros of the Krall–Jacobi polynomials are the same as the first two even moments of the zeros of the Krall–Legendre polynomials.

(b) The first four moments of the zeros of the Krall–Jacobi polynomials are equal to the first four even moments of the zeros of the classical Jacobi polynomials.

It can be shown that these two statements are verified by moments of arbitrary order.[17] Then it is natural to think that there must be deeper relations between the asymptotical distribution of zeros of the various Krall polynomials and the corresponding quantity of the classical orthogonal polynomials.

VI. CONCLUSION

There have been found explicit expressions for the first four moments of the normalized density of zeros of polynomials satisfying a fourth-order differential equation. Also, the results have been applied to all the orthogonal polynomials, except the classical ones, verifying such a differential equation. To the best of our information, this is the first time that any property of the zeros of these polynomials is found.

Especially stricking are the moments of the density of zeros of the Krall–Jacobi polynomials and their relation with the corresponding quantities of the Krall–Legendre polynomials and of the classical Jacobi polynomials.

ACKNOWLEDGMENTS

One of us (J.S.D.) would like to thank Dr. Lance Littlejohn for supplying useful information about the Krall polynomials and for encouragement. A useful discussion with F. J. Galvez is also acknowledged.

This work was supported in part by the Comisión Asesora de Investigación Cientifica y Técnica (CAICYT) (Spain).

APPENDIX A: DERIVATION OF $J_r^{(3)}$

Here we want to prove Eq. (7). Equation (4) gives the definition

$$J_r^{(3)} = \sum_{\neq} \frac{x_{l_1}^r}{(x_{l_1} - x_{l_2})(x_{l_1} - x_{l_3})},$$

which, according to Ref. 9, can be transformed into

$$J_r^{(3)} = \frac{2}{3}\sum_{\neq} \frac{x_{l_1}^r - x_{l_2}^r}{(x_{l_1} - x_{l_2})(x_{l_1} - x_{l_3})}.$$

Here the symbol $\Sigma_{\neq}$ is a sum over l_1, l_2, and l_3 such that all of them are different among each other. Using the simple relations

$$\frac{x_{l_1}^r - x_{l_2}^r}{x_{l_1} - x_{l_2}} = \sum_{s=0}^{r-1} x_{l_1}^{r-1-s} x_{l_2}^s,$$
$$\sum_{l_2 \neq l_3} \frac{x_{l_2}^s}{x_{l_2} - x_{l_3}} = \frac{1}{2} \sum_{l_2 \neq l_3} \frac{x_{l_2}^s - x_{l_3}^s}{x_{l_2} - x_{l_3}}, \tag{A1}$$

one obtains after some manipulations the relation

$$J_r^{(3)} = \frac{1}{3} \sum_{s=1}^{r-1} \sum_{t=0}^{s-1} \sum_{\neq} x_{l_1}^{r-1-s} x_{l_2}^{s-1-t} x_{l_3}^t$$
$$= \frac{1}{3} \sum_{s=1}^{r-1} \sum_{t=0}^{s-1} B_3(r-1-s, s-1-t, t), \tag{A2}$$

where we have used the notation

$$B_3(t_1,t_2,t_3) = \sum_{\neq} x_{l_1}^{t_1} x_{l_2}^{t_2} x_{l_3}^{t_3} \tag{A3}$$

to write the last equality. Now we calculate the B_3 quantity by expanding the triple sum. Taking into account the definition $y_r = \Sigma_{l=1}^N x_l^r$ and making some straightforward algebraic operations, we obtain

$$B_3(t_1,t_2,t_3) = 2y_\sigma - \sum_{i=1}^{3} y_{t_i} y_{\sigma - t_i} + \sum_{i=1}^{3} y_{t_i},$$

where $\sigma = t_1 + t_2 + t_3$. The use of this in Eq. (A2) leads to

$$J_r^{(3)} = \frac{1}{3} \sum_{s=1}^{r-1} \sum_{t=0}^{s-1} [\, y_{r-1-s} y_{s-1-t} y_t$$
$$- 3 y_{r-1-s} y_{s-1} + 2 y_{r-2} \,].$$

Simple operations here produce the wanted value of $J_r^{(3)}$ given in Eq. (7).

APPENDIX B: DERIVATION OF $J_r^{(4)}$

Now let us prove Eq. (9). From Eq. (4) we have the definition

$$J_r^{(4)} = \sum_{\neq} \frac{x_{l_1}^r}{(x_{l_1} - x_{l_2})(x_{l_1} - x_{l_3})(x_{l_1} - x_{l_4})}.$$

It is shown in Ref. 9 that

$$J_r^{(4)} = \sum_{\neq} \frac{x_{l_1}^r - x_{l_2}^r}{(x_{l_1} - x_{l_2})(x_{l_1} - x_{l_3})(x_{l_1} - x_{l_4})}.$$

Here the symbol $\Sigma_{\neq}$ denotes a sum over l_1, l_2, l_3, and l_4 such that none of them are mutually equal. From here and by means of Eq. (28) and the following relation

$$\sum_{l_2 \neq l_3} \sum_{l_3=1}^{N} \frac{x_{l_2}^s}{x_{l_2} - x_{l_3}} = \frac{1}{2} \sum_{l_2 \neq l_3} \sum_{l_3=1}^{N} \frac{x_{l_2}^s - x_{l_3}^s}{x_{l_2} - x_{l_3}},$$

we can easily show that

$$J_r^{(4)} = \frac{1}{4} \sum_{s=2}^{r-1} \sum_{t=1}^{s-1} \sum_{u=0}^{t-1} \sum_{\neq} x_{l_1}^{r-1-s} x_{l_2}^{s-1-t} x_{l_3}^{t-1-u} x_{l_4}^u$$
$$= \frac{1}{4} \sum_{s=2}^{r-1} \sum_{t=1}^{s-1} \sum_{u=0}^{t-1} B_4(r-1-s, s-1-t, t-1-u, u). \tag{B1}$$

To write the last equality we have used the notation

$$B_4(t_1,t_2,t_3,t_4) = \sum_{\neq} x_{l_1}^{t_1} x_{l_2}^{t_2} x_{l_3}^{t_3} x_{l_4}^{t_4},$$

expanding completely this quadruple summation, regrouping terms, and bearing in mind the definition of y_r produces the following results:

$$B_4(t_1,t_2,t_3,t_4) = -6y_\sigma + 2 \sum_{i=1}^{4} y_{t_i} y_{\sigma - t_i}$$
$$+ \sum_{i=1}^{3} y_{t_i + t_{i+1}} y_{\sigma - t_i - t_{i+1}}$$
$$- \sum_{j>i}^{4} y_{t_i} y_{t_j} y_{\sigma - t_i - t_j} + \prod_{i=1}^{4} y_{t_i},$$

with $\sigma = t_1 + t_2 + t_3 + t_4$. The use of this result in Eq. (31) allows us to write

$$J_r^{(4)} = \frac{1}{4} \sum_{s=2}^{r-1} \sum_{t=1}^{s-1} \sum_{u=0}^{t-1} [\, y_{r-1-s} y_{s-1-t} y_{t-1-u} y_u$$
$$- 6 y_{r-1-s} y_{r-1-t} y_{t-1} + 8 y_{r-1-t} y_{t-2}$$
$$+ 3 y_{r-2-s} y_{s-1} - 6 y_{r-3} \,].$$

A long series of straightforward but tedious manipulations transforms this equation into the wanted expression (9) for $J_r^{(4)}$.

[1]C. Lanczos, J. Res. Natl. Bur. Standards **45**, 255 (1950); C. C. Paige, J. Inst. Math. Appl. **10**, 373 (1972).
[2]R. R. Whitehead, A. Watt, B. J. Cole, and I. Morrison, Adv. Nucl. Phys. **9**, 123 (1977); D. W. Bullet, R. Haydock, V. Heine, and M. J. Kelly, *Solid State Physics* (Academic, New York, 1980), Vol. 35.
[3]H. S. Wilf, *Mathematics for the Physical Sciences* (Wiley, London, 1962).
[4]F. J. Dyson, Phys. Rev. **92**, 1331 (1953).
[5]P. Dean, Proc. Cambridge Philos. Soc. **52**, 752 (1956).
[6]J. S. Dehesa, J. Comput. Appl. Math. **3**, 167 (1977); Linear Algebra Its Appl. **33**, 41 (1980).
[7]P. Nevai and J. S. Dehesa, SIAM J. Math. Anal. **10**, 1184 (1979); J. S. Dehesa, J. Phys. A: Math. Gen. **11**, L223 (1978).
[8]J. S. Dehesa, J. Comp. Appl. Math. **7**, 249 (1981).
[9]K. M. Case, J. Math. Phys. **21**, 702 (1980).
[10]H. Cramer, *Mathematical Methods of Statistics* (Princeton U. P., Princeton, 1946).
[11]W. Weibull, J. Appl. Mech. **18**, 293 (1951).
[12]W. P. Elderton, *Frequency Curves and Correlation* (Harren, Washington, DC, 1953).
[13]H. L. Krall, *On orthogonal polynomials satisfying fourth-order differential equations, Pennsylvania State College Studies, No. 6* (Pennsylvania State College, State College, PA, 1940).
[14]A. M. Krall, Proc. R. Soc. Edinburgh Sect. A **87**, 271 (1981).
[15]A. M. Krall and R. D. Morton, SIAM J. Math. Anal. **4**, 604 (1978).
[16]A. M. Krall, SIAM Rev. **22**, 436 (1980).
[17]J. S. Dehesa (to be published).

Nonlinear multigroup neutron-flux systems: Blowup, decay, and steady states

Gen-Shun Chen
Department of Chemical and Nuclear Engineering, University of Cincinnati, Cincinnati, Ohio 45221
Anthony W. Leung
Department of Mathematical Sciences, University of Cincinnati, Cincinnati, Ohio 45221

(Received 17 July 1984; accepted for publication 11 January 1985)

This article investigates a nonlinear system of partial differential equations describing multigroup neutron-flux reaction-diffusion inside a nuclear fission reactor. The neutrons are divided into n energy groups, with fission and scattering rates dependent on temperature, which gives rise to the $(n+1)$th equation. Cases concerning directly coupled and down scattering, when group transfers only occur from higher to lower energy groups, are considered. Quantitative conditions on the various fission and scattering rates are found for eventual "blowup" and "decay" of neutron concentrations. Finally, a system with fission and scattering rates dependent on neutron density is also investigated. Sufficient conditions for positive nontrivial steady state are found.

I. INTRODUCTION: NONLINEAR TEMPERATURE-DEPENDENT MODELS FOR NEUTRON FISSION

This paper studies the steady states of multigroup diffusion equations, describing neutron-flux reaction-diffusion inside a nuclear fission reactor. The reactor core is represented by a bounded domain $\mathscr{D}$ in R^d, $d \geqslant 2$. The functions $u_i(x)$, $i = 1,\dots,n$, $x = (x_1,\dots,x_d) \in \mathscr{D}$ are the neutron flux of the ith energy group. The $T(x)$ is the core temperature above coolant temperature. The following system of nonlinear temperature feedback multigroup elliptic diffusion equations will be considered in Sec. II:

$$\begin{aligned} &\Delta u_i + \sum_{j=1}^{n} H_{ij}(x,T)u_j = 0, \quad i = 1,\dots,n, \\ &\Delta T - c(x)T + \sum_{j=1}^{n} G_j(x,T)u_j = 0, \quad \text{in } \mathscr{D}. \end{aligned} \tag{1.1}$$

Here $\Delta \equiv \Sigma_{i=1}^{d} (\partial^2/\partial x_i^2)$, $c(x) > 0$ in $\mathscr{D}$ closure represents the cooling function. The functions determining interaction rates, $H_{ij}(x,T)$ and $G_j(x,T)$, are assumed to be functions of space and temperature. In more conventional notations of nuclear engineering,

$$H_{11} \equiv \sigma_1^{-1}[v_1\Sigma_{f_1} - \Sigma_R], \quad H_{1j} \equiv \sigma_1^{-1} v_j \Sigma_{f_j}, \quad \text{for } j \geqslant 2,$$

$$H_{ij} \equiv \sigma_i^{-1}\Sigma_{S_{ji}}, \quad \text{for } i > 1, \quad j = 1,\dots,i-1,i+1,\dots,n,$$

$$H_{ii} \equiv -\sigma_i^{-1}\Sigma_{a_i}, \quad \text{for } i > 1, \quad G_i = \theta_i \Sigma_{f_i}.$$

Here, Σ_{f_i} is the fission "macroscopic cross section" of group i. Also, Σ_R is the removal cross section, $\Sigma_{S_{ji}}$ is the group-transfer cross section from group j to i, and Σ_{a_i} is the absorption cross section of group i. The σ_i, v_i, and θ_i are diffusion, neutron release, and energy release parameters, respectively. Detailed descriptions of such symbols can be found in Ref. 1, p. 288.

Linear models, neglecting T and the last equation (1.1), had been extensively studied analytically and numerically (see, e.g., Refs. 1–3). The limitations of such linear models had been mentioned in Ref. 4. The advisability of a temperature-dependent feedback model had been proposed and studied in Refs. 5–7. Power levels of reactors are adjusted by moving the control rods, causing changes in temperature and neutron flux in the core. However, the atomic concentrations of materials in the core depend sensitively on temperature. As temperature changes, they may contract, expand, or change phase, eventually causing a change in the macroscopic cross section.

In Ref. 7, temperature feedback models are only investigated for two-group neutron flux where the group-transfer scattering effect is quite simple. In practice, the multigroup equations are commonly applied in cases of four or more groups. The scattering effect is more involved, and we will consider the situations of directly coupled and down scattering in n groups. Moreover, the coefficients or cross sections H_{ij} are now dependent on space x, while in Ref. 7 they are independent of x and can only be applied to homogeneous reactors.

In the case of down scattering, one assumes group transfer only from higher to lower energy groups (i.e., from group j to group i, $i > j$). On the other hand, fissions from each group produce neutrons in the first group. Therefore, we assume

$$H_{1j} > 0, \quad j = 2,\dots,n. \tag{1.2}$$

$$\text{For } i > 1: \quad H_{ij} > 0, \quad \text{if } j < i, \quad H_{ii} \overset{\text{(def)}}{=} -A_i(x,T) < 0, \quad \text{and } H_{ij} \equiv 0, \quad \text{if } j > i. \tag{1.3}$$

All inequalities are true for $x \in \overline{\mathscr{D}}$, $T \geqslant 0$. In the first row, $i = 1$, only H_{11} is not assumed positive because of the removal term $-\Sigma_R$ which can be adjusted by control rods.

We now clarify our notations, conventions, and assumptions in this article. Here, $\mathscr{D}$ is a bounded domain in R^d, $d \geqslant 2$, whose boundary $\delta\mathscr{D}$ is C^2 smooth [i.e., can be locally represented as $x_i = \phi(x)$ for some i, ϕ with continuous second derivatives and independent of x_i]; $\overline{\mathscr{D}}$ denotes $\mathscr{D}$ closure. The functions H_{ij}, G_i, $i,j = 1,\dots,n$ are continuous functions of $x \in \overline{\mathscr{D}}$, $T \geqslant 0$; $c(x)$ is continuous and positive in $\overline{\mathscr{D}}$.

 0022-2488/85/071553-07$02.50

For convenience, let

$$\tilde{h}_{ij} = \inf\{H_{ij}(x,T)|x\in\overline{\mathscr{D}},\ T\geqslant 0\},$$

$$\bar{h}_{ij} = \sup\{H_{ij}(x,T)|x\in\overline{\mathscr{D}},\ T\geqslant 0\},$$

for $i, j = 1,...,n$. For $i\neq 1$, $\tilde{a}_i = \inf\{A_i(x,T)|x\in\overline{\mathscr{D}},T\geqslant 0\}$, $\bar{a}_i = \sup\{A_i(x,T)|x\in\overline{\mathscr{D}},T\geqslant 0\}$, and similarly we define $\tilde{g}_i, \bar{g}_i$ to be the corresponding inf and sup of G_i, $i = 1,...,n$. We assume that

$$-\infty < \tilde{h}_{11}\leqslant\bar{h}_{11} < \infty,\quad 0<\tilde{h}_{ij}\leqslant\bar{h}_{ij} < \infty,$$
$$\text{for } i = 1, j = 2,...,n, \quad \text{and for } i>1, j<i, \tag{1.4}$$
$$0<\tilde{a}_i\leqslant\bar{a}_i < \infty,\quad i = 2,...,n,$$
$$0<\tilde{g}_i\leqslant\bar{g}_i < \infty,\quad i = 1,...,n.$$

[Note that (1.3) implies that for $i>1$, $\tilde{h}_{ij}\bar{h}_{ij} = 0$ if $j>i$.] These are very reasonable and general assumptions for the reactor model.

Let $\lambda_1 > 0$ denote the first eigenvalue of the eigenvalue problem: $\Delta w + \lambda w = 0$ in $\mathscr{D}$, $w = 0$ in $\delta\mathscr{D}$, where $\omega(x)$ is the corresponding normalized eigenfunction with $\max\{\omega(x)|x\in\overline{\mathscr{D}}\} = 1$. For positive integers r, $C^r(\mathscr{D})$ and $C^r(\overline{\mathscr{D}})$ denote r times continuously differentiable functions in $\mathscr{D}$ and $\overline{\mathscr{D}}$, respectively.

In Sec. II, we consider Eqs. (1.1) in $\mathscr{D}$ with non-negative or zero Dirichlet boundary conditions on $\delta\mathscr{D}$. Theorem 2.1 finds a very simple criterion when nontrivial non-negative solutions cannot exist (for the corresponding time-dependent parabolic problem, solutions with non-negative initial data would "blow up" as $t\to+\infty$). Theorem 2.2 considers the case of directly coupled scattering (a special case of down scattering); it finds some sufficient conditions for "blow up" when the criterion in Theorem 2.1 is violated. Theorem 2.3 extends Theorem 2.2 to the general down scattering situation. Theorem 2.4 finds "decay" conditions for the nonexistence of nontrivial non-negative solutions; such conditions for the corresponding time-dependent parabolic problem would imply solutions tending to zero as $t\to+\infty$. Corollary 2.5 gives a simple diagonally dominant criterion for an application of Theorem 2.4.

In Sec. III, a simpler version of temperature feedback is considered. We assume that fissions and other cross sections are promptly affected by the neutron flux, eliminating the last equation in (1.1). Theorems 3.1 and 3.2 find sufficient conditions for the existence of nontrivial non-negative equilibria for the cases of directly coupled and down scattering, respectively.

II. BLOWUP AND DECAY CRITERIA

In this section, hypotheses (1.2)–(1.4) are always assumed in every theorem, unless otherwise stated.

When the fission cross section of the first group is large compared with the removal cross section, one expects the reactor to blow up. Theorem 2.1 gives a simple quantitative criterion for this to happen, and nontrivial non-negative steady state cannot exist.

Theorem 2.1: Suppose that

$$\tilde{h}_{11} \equiv \inf\{H_{11}(x,s)|x\in\overline{\mathscr{D}},s\geqslant 0\}\geqslant\lambda_1. \tag{2.1}$$

Then Eq. (1.1) with boundary conditions

$$u_i(x) = u_i^0(x)\geqslant 0,\quad i = 1,...,n,\quad T(x) = T^0(x)\geqslant 0,$$

for $x\in\delta\mathscr{D}$ [here $u_1^0(x),...,u_n^0(x)$, $T^0(x)$ are given functions on $\delta\mathscr{D}$], has no solution $(\hat{u}_1(x),...,\hat{u}_n(x), \hat{T}(x))$ with the following properties:

(i) Each component is in $C^2(\mathscr{D})\cap C^1(\overline{\mathscr{D}})$.

(ii) $\hat{u}_i(x)\geqslant 0$, $i = 1,...,n$, $\hat{T}(x)\geqslant 0$ in $\overline{\mathscr{D}}$.

(iii) $\hat{u}_1(x)\not\equiv 0$ in $\overline{\mathscr{D}}$.

Proof: Assume that $(\hat{u}_1(x),...,\hat{u}_n(x),\hat{T}(x))$ exists as described with properties (i)–(iii). We will construct a family of lower bounds for the function $\hat{u}_1(x)$, parametrized by $\delta>0$. As $\delta\to\infty$, the lower bound will tend to ∞. For each $\delta>0$, define $u_1^\delta(x) = \delta\omega(x)$ for $x\in\overline{\mathscr{D}}$. For all $u_i(x)\geqslant 0$, $i = 2,...,n$, $T(x)\geqslant 0$, we have

$$\Delta u_1^\delta(x) + H_{11}(x,T(x))u_1^\delta(x) + \sum_{j=2}^n H_{1j}(x,T(x))u_j(x)$$
$$\geqslant\delta\omega(x)[-\lambda_1 + \tilde{h}_{11}]\geqslant 0, \tag{2.2}$$

in $\mathscr{D}$. From properties (i)–(iii), we now deduce that $\hat{u}_1(x) > u_1^\delta(x)$ for $x\in\mathscr{D}$ and $\delta>0$ sufficiently small. Let $C > |\tilde{h}_{11}|$; we have

$$\Delta\hat{u}_1(x) - C\hat{u}_1(x) = -[H_{11}(x,\hat{T}(x)) + C]\hat{u}_1(x)$$
$$-\sum_{j=2}^n H_{1j}(x,\hat{T}(x))\hat{u}_j(x)\leqslant 0,$$

in $\mathscr{D}$, $\hat{u}_1\geqslant 0$ in $\overline{\mathscr{D}}$. Maximum principle implies that $\hat{u}_1(x)>0$ in $\mathscr{D}$. Moreover, applying the maximum principle at the boundary, we find that outward normal derivatives $\partial\hat{u}_1/\partial\eta$ must be negative at those boundary points where $\hat{u}_1 = 0$. Consequently the set $\mathscr{S}\equiv\{s>0|\hat{u}_1(x)>u_1^\delta(x)$ for all $0\leqslant\delta\leqslant s$, $x\in\mathscr{D}\}$ is nonempty.

Suppose $\mathscr{S}$ has an upper bound; let its lub be $\bar{\delta}$. There must be a point in $\mathscr{D}$ where $\hat{u}_1 = u_1^{\bar{\delta}}$. Otherwise, we consider for $C>|\tilde{h}_{11}|$ that

$$\Delta(\hat{u}_1 - u_1^{\bar{\delta}}) - C(\hat{u}_1 - u_1^{\bar{\delta}})$$
$$= \left\{\Delta\hat{u}_1 + H_{11}(x,\hat{T})\hat{u}_1 + \sum_{j=2}^n H_{1j}(x,\hat{T})\hat{u}_j\right\}$$
$$-\left\{\Delta u_1^{\bar{\delta}} + H_{11}(x,\hat{T})u_1^{\bar{\delta}} + \sum_{j=2}^n H_{1j}(x,\hat{T})\hat{u}_j\right\}$$
$$-\{H_{11}(x,\hat{T}) + C\}(\hat{u}_1 - u_1^{\bar{\delta}})\leqslant 0, \tag{2.3}$$

in $\mathscr{D}$. [The last inequality is true because $\hat{T}(x)\geqslant 0$, $\hat{u}_j(x)\geqslant 0$, $j = 2,...,n$, and inequality (2.2) can be applied with $\delta = \bar{\delta}$.] Together with $\hat{u}_1 - u_1^{\bar{\delta}}\geqslant 0$ in $\overline{\mathscr{D}}$, this implies that $\partial\hat{u}_1/\partial\eta < \partial u_1^{\bar{\delta}}/\partial\eta$ at those points at the boundary where $\hat{u}_1 = u_1^{\bar{\delta}}$. Consequently, for sufficiently small $\epsilon>0$, $u_1^{\bar{\delta}+\epsilon} < \hat{u}_1$ for all $x\in\mathscr{D}$. This violates the definition of $\bar{\delta}$, and we conclude that there must be a point $\bar{x}\in\mathscr{D}$ where $u_1^{\bar{\delta}}(\bar{x}) = \hat{u}_1(\bar{x})$.

Now, inequality (2.3) and maximum principle again imply that $u_1^{\bar{\delta}}(x)\equiv\hat{u}_1(x)$ in $\overline{\mathscr{D}}$. Therefore, we have for $x\in\mathscr{D}$

$$0 = \Delta\hat{u}_1 + H_{11}(x,\hat{T})\hat{u} + \sum_{j=2}^n H_{1j}(x,\hat{T})\hat{u}_j$$
$$= \Delta u_1^{\bar{\delta}} + H_{11}(x,\hat{T})u_1^{\bar{\delta}} + \sum_{j=2}^n H_{1j}(x,\hat{T})\hat{u}_j$$
$$\geqslant[-\lambda_1 + \tilde{h}_{11}]\bar{\delta}\omega + \sum_{j=2}^n H_{1j}(x,\hat{T})\hat{u}_j. \tag{2.4}$$

Hypotheses (2.1) and (ii) and inequality (2.4) imply that $\hat{u}_j(x) \equiv 0$ in $\mathscr{D}$ for $j = 2,...,n$. Referring to the second equation in (1.1) (i.e., $i = 2$), one obtains $H_{21}\hat{u}_1 \equiv 0$, and by (1.3) one is led to the conclusion that $\hat{u}_1(x) \equiv 0$ in $\overline{\mathscr{D}}$. This contradicts hypothesis (iii).

The last two paragraphs show that the set $\mathscr{S}$ is unbounded. However, as $\delta \to +\infty$, $u_1^\delta(x) \to +\infty$ for $x \in \mathscr{D}$, and this contradicts the existence of $(\hat{u}_1(x),...,\hat{u}_n(x), \hat{T}(x))$ satisfying (i).

Remark: In Theorem 2.1, if it is further assumed that $u_i^0(x) \equiv 0$ on $\delta\mathscr{D}$, $i = 2,...,n$, then we can apply maximum principle to the ith equation successively, $i = 2,...,n$ in (1.1) to conclude that the only solution with properties (i) and (ii) is $(0,...,0,\hat{T}(x))$, where $\Delta\hat{T}(x) - c(x)\hat{T}(x) = 0$ in $\mathscr{D}$, $\hat{T}(x) = T^0(x) \geqslant 0$ on $\delta\mathscr{D}$.

When (2.1) is violated, which arises when the fission cross section of group 1 is small compared with removal cross section, it is still possible for blowup situations to occur. In order to make the conditions more readily understandable, we temporarily consider the simpler situation of "directly coupled scattering." In nuclear theory terms, this means that neutrons in a given energy group i only scatter into the next lower energy group $i + 1$. More precisely, hypotheses (1.3) and (1.4) are modified to the following two sets of conditions:

$$\text{For } i > 1: \quad H_{i,i-1} > 0, \quad H_{ii} = -A_i(x,T) < 0, \qquad H_{ij} \equiv 0, \ \text{if } j > i, \ \text{or } j < i - 1. \tag{2.5}$$

All inequalities are true for $x \in \overline{\mathscr{D}}$, $T \geqslant 0$

$$-\infty < \tilde{h}_{11} \leqslant \bar{h}_{11} < \infty, \quad 0 < \tilde{h}_{ij} \leqslant \bar{h}_{ij} < \infty,$$
$$\text{for } i = 1, \quad j = 2,...,n, \quad \text{and for } i > 1, \quad j = i - 1, \tag{2.6}$$
$$0 < \tilde{a}_i \leqslant \bar{a}_i < \infty, \quad i = 2,...,n,$$
$$0 < \tilde{g}_i < \bar{g}_i < \infty, \quad i = 1,...,n.$$

[Note that (2.5) implies that for $i > 1$, $\tilde{h}_{ij} = \bar{h}_{ij} = 0$ if $j > i$ or $j < i - 1$.]

Theorem 2.2 gives sufficient conditions that neutron formations in the first m groups are fast enough to blow up. Consequently, no finite steady state can exist.

Theorem 2.2 (directly coupled scattering): Suppose that

$$\tilde{h}_{11} \equiv \inf\{H_{11}(x,s) | x \in \overline{\mathscr{D}}, s \geqslant 0\} < \lambda_1,$$

and there exist positive constants $\delta_2,...,\delta_m$, $2 \leqslant m \leqslant n$, with $\Sigma_{i=1}^m \delta_i = 1$, so that

$$\tilde{h}_{21}\tilde{h}_{12} > (\lambda_1 + \bar{a}_2)(\lambda_1 - \tilde{h}_{11})\delta_2, \tag{2.7}$$

$$\delta_{k-1}\tilde{h}_{k,k-1}\tilde{h}_{1k} > (\lambda_1 + \bar{a}_k)h_{1,k-1} \cdot \delta_k, \quad k = 3,...,m. \tag{2.8}$$

Then Eq. (1.1) [under hypothesis (1.2), with (2.5) and (2.6) replacing (1.3) and (1.4)] with prescribed boundary conditions

$$u_i(x) = u_i^0(x) \geqslant 0, \quad i = 1,...,n, \quad T(x) = T^0(x) \geqslant 0,$$

has no solution $(\hat{u}_1(x),...,\hat{u}_n(x),\hat{T}(x))$ with properties (i)–(iii) as described in Theorem 2.1.

Remark: Suppose $m = 2$; then (2.7) is supposed to be valid for $\delta_2 = 1$ and inequalities (2.8) will all be absent. Equation (2.8) involves $(m - 2)$ inequalities, which are more readily satisfied by choosing δ_k smaller for increasing k.

Proof: Assume that $(\hat{u}_1(x),...,\hat{u}_n(x),\hat{T}(x))$ exists with properties (i)–(iii). We will construct a family of lower bounds for the solution as in Theorem 2.1, and will eventually lead to a contradiction. Let $\tilde{\tilde{h}}_{11} < \tilde{h}_{11}$ so that (2.7) is still valid with $\tilde{h}_{11}$ replaced by $\tilde{\tilde{h}}_{11}$; for each $\delta > 0$, define $u_1^\delta(x) = \delta\omega(x)$, $u_i^\delta(x) = \delta_i \tilde{h}_{1i}^{-1}(\lambda_1 - \tilde{\tilde{h}}_{11})\delta\omega$, $i = 2,...,m$, $u_j^\delta(x) \equiv 0$, $j = m + 1,...,n$, and $T^\delta(x) = 0$, for $x \in \overline{\mathscr{D}}$. For all $u_k(x) \geqslant u_k^\delta(x)$, $k = 2,...,n$, $T(x) \geqslant T^\delta(x)$, we have

$$\begin{aligned} \Delta u_1^\delta(x) &+ H_{11}(x,T(x))u_1^\delta(x) + \sum_{j=2}^n H_{1j}(x,T(x))u_j \\ &\geqslant [-\lambda_1 + \tilde{h}_{11}]\delta\omega(x) \\ &\quad + \sum_{j=2}^m \tilde{h}_{1j}\,\delta_j\,\tilde{h}_{1j}^{-1}(\lambda_1 - \tilde{\tilde{h}}_{11})\delta\omega > 0, \end{aligned} \tag{2.9}$$

in $\mathscr{D}$. For all $u_1(x) \geqslant u_1^\delta(x)$, $T(x) \geqslant T^\delta(x)$, we have

$$\begin{aligned} \Delta u_2^\delta(x) &+ H_{21}(x,T(x))u_1(x) - A_2(x,T(x))u_2^\delta(x) \\ &\geqslant \delta\omega(x)[\tilde{h}_{12}^{-1}\delta_2(-\lambda_1)(\lambda_1 - \tilde{\tilde{h}}_1) \\ &\quad + \tilde{h}_{21} - \bar{a}_2\tilde{h}_{12}^{-1}\delta_2(\lambda_1 - \tilde{\tilde{h}}_{11})] > 0, \end{aligned} \tag{2.10}$$

in $\mathscr{D}$. [The last inequality is valid due to (2.7) and the choice of $\tilde{\tilde{h}}_{11}$.] For each $i = 3,...,m$, all $u_k(x) \geqslant u_k^\delta(x)$, $k \neq i$, $T(x) \geqslant T^\delta(x)$, we have

$$\begin{aligned} \Delta u_i^\delta(x) &+ H_{i,i-1}(x,T(x))u_{i-1}(x) - A_i(x,T(x))u_i^\delta(x) \\ &\geqslant \delta\omega(x)(\lambda_1 - \tilde{\tilde{h}}_{11})[(-\lambda_1)\delta_i\tilde{h}_{1i}^{-1} + \tilde{h}_{i,i-1}\delta_{i-1}\tilde{h}_{1,i-1}^{-1} \\ &\quad - \bar{a}_i\,\delta_i\,\tilde{h}_{1i}^{-1}] > 0, \end{aligned} \tag{2.11}$$

in $\mathscr{D}$. The last inequality is due to (2.8).

We now show that properties (i)–(iii) imply that $\hat{u}_i(x) > 0$ for $x \in \mathscr{D}$, $i = 1,...,m$. The case for $i = 1$ is the same as in Theorem 2.1. For $2 \leqslant i \leqslant m$, let $P > \max\{\bar{a}_i | i = 2,...m\}$; we have

$$\Delta\hat{u}_i - P\hat{u}_i = -H_{i,i-1}(x,\hat{T})\hat{u}_{i-1} - [P - A_i(x,\hat{T})]$$
$$\hat{u}_i \leqslant 0, \ \text{in } \mathscr{D}, \quad \hat{u}_i \geqslant 0, \ \text{in } \overline{\mathscr{D}}.$$

Maximum principle implies that $\hat{u}_i > 0$ in $\mathscr{D}$ or $\hat{u}_i \equiv 0$ in $\overline{\mathscr{D}}$. However, (iii) implies successively that the trivial function is not a solution of the ith equation in (1.1), $i = 2,...,m$. Hence $\hat{u}_i(x) > 0$ in $\mathscr{D}$. Moreover, maximum principle at the boundary indicates that outward normal derivatives $\partial\hat{u}_i/\partial\eta$, $i = 1,...,m$, are negative at those boundary points where the corresponding function is 0.

From the above paragraph, we see that the set $\mathscr{S} \equiv \{s > 0 | \hat{u}_i(x) > \hat{u}_i^\delta(x), i = 1,...,m, \text{for all } 0 \leqslant \delta < s, x \in \mathscr{D}\}$ is nonempty. Suppose $\mathscr{S}$ has an upper bound; let its lub be $\bar{\delta}$. If there is a point at the boundary where $u_i^{\bar{\delta}} = \hat{u}_i$, some $i = 1,...,m$, we deduce a contradiction to the definition of $\bar{\delta}$ by using maximum principle at the boundary, with the inequalities as in (2.3) and

$$\begin{aligned} \Delta(\hat{u}_i - u_i^{\bar{\delta}}) &- A_i(x,\hat{T}(x))(\hat{u}_i - u_i^{\bar{\delta}}) \\ &= \{\Delta\hat{u}_i + H_{i,i-1}(x,\hat{T})\hat{u}_{i-1} - A_i(x,\hat{T})\hat{u}_i\} \\ &\quad - \{\Delta u_i^{\bar{\delta}} + H_{i,i-1}(x,\hat{T})\hat{u}_{i-1} - A_i(x,\hat{T})u_i^{\bar{\delta}}\} < 0, \end{aligned} \tag{2.12}$$

in $\mathscr{D}$, $i = 2,...,m$. On the other hand, suppose that there is a point $\bar{x} \in \mathscr{D}$ where $u_i^{\bar{\delta}}(\bar{x}) = \hat{u}_i(\bar{x})$, some $i = 1,...,m$. Inequalities (2.3) and (2.12) and maximum principle imply that $u_i^{\bar{\delta}}(x) \equiv \hat{u}_i(x)$ in $\overline{\mathscr{D}}$. If $i = 1$, we consider

$$0 = \Delta \hat{u}_1 + H_{11}(x,\widehat{T})\hat{u}_1 + \sum_{j=2}^{n} H_{1j}(x,\widehat{T})\hat{u}_j$$
$$= \Delta u_1^{\bar{\delta}} + H_{11}(x,\widehat{T})u_1^{\bar{\delta}} + \sum_{j=2}^{n} H_{1j}(x,\widehat{T})\hat{u}_j > 0,$$

which is a contradiction. The last inequality is true by letting $\delta = \bar{\delta}$ in (2.9). If $i = 2,...,m$, we consider $0 = \Delta \hat{u}_i + H_{i,i-1}(x,\widehat{T})\hat{u}_{i-1} - A_i(x,\widehat{T})\hat{u}_i = \Delta u_i^{\bar{\delta}} + H_{i,i-1}(x,\widehat{T})\hat{u}_{i-1} - A_i(x,\widehat{T})u_i^{\bar{\delta}} > 0$ by (2.10) and (2.11), and again arrive at a contradiction.

The set $\mathscr{S}$ is consequently unbounded. However, $u_i^{\delta}(x) \to +\infty$ as $\delta \to +\infty$, $i = 1,...,m$. This proves the nonexistence of $(\hat{u}_1(x),...,\hat{u}_n(x),\widehat{T}(x))$.

The following theorem returns to the general down scattering condition. A more involved blowup condition is found. Neutrons in groups $k_1,k_2,...,k_p$ are formed too quickly so that steady state cannot occur.

Theorem 2.3: Suppose that

$$\tilde{h}_{11} \equiv \inf\{H_{11}(x,s) | x \in \overline{\mathscr{D}}, s \geqslant 0\} < \lambda_1,$$

and there exist integers $1 < k_1 < k_2 < \cdots < k_p \leqslant n$, with corresponding positive constants $\delta_{k_1},...,\delta_{k_p}$ with $\Sigma_{i=1}^{p} \delta_{k_i} = 1$, so that

$$\tilde{h}_{k_1 1}\, \tilde{h}_{1k_1} > (\lambda_1 + \bar{a}_{k_1})(\lambda_1 - \tilde{h}_{11})\delta_{k_1} \tilde{h}_{1k_i}$$
$$\times \left[\tilde{h}_{k_i 1} + \sum_{s=1}^{i-1} \delta_{k_s} \tilde{h}_{k_i k_s}\, \tilde{h}_{1k_s}^{-1}(\lambda_1 - \tilde{h}_{11}) \right]$$
$$> (\lambda_1 + \bar{a}_{k_i})(\lambda_1 - \tilde{h}_{11})\delta_{k_i}, \quad \text{for } i = 1,...,p.$$

Then Eq. (1.1) with prescribed boundary conditions:

$$u_i(x) = u_i^0(x) \geqslant 0, \quad i = 1,...,n, \quad T(x) = T^0(x) \geqslant 0$$

has no solution $(\hat{u}_1(x),...,\hat{u}_n(x),\widehat{T}(x))$ with properties (i)–(iii) as described in Theorem 2.1.

The proof is exactly analogous to that of Theorem 2.2, using the family of lower solutions $u_1^{\delta}(x) = \delta\omega(x)$, $u_{k_i}^{\delta} = \delta_{k_i} \tilde{h}_{1k_i}^{-1}(\lambda_1 - \tilde{\tilde{h}}_{11})\delta\omega(x)$ for $i = 1,...,p$, $u_j^{\delta}(x) \equiv 0$ for $j \neq k_i$, all i. The details will be omitted.

The remaining parts of this section discuss conditions when the neutron density of each group will decay to zero for the time-dependent parabolic model. This means that the only non-negative steady state is the trivial one.

Theorem 2.4: Suppose that

$$\bar{h}_{11} = \sup\{H_{11}(x,s) | x \in \overline{\mathscr{D}} s \geqslant 0\} < \lambda_1, \tag{2.13}$$

and there exist positive constants $c_i > 0$, $i = 1,...,n$ such that

$$\sum_{j=2}^{n} c_j \bar{h}_{1j} + c_1(-\lambda_1 + \bar{h}_{11}) < 0 \tag{2.14}$$

and

$$\sum_{\substack{j=1 \\ j \neq i}}^{n} c_j \bar{h}_{ij} + c_i(-\lambda_1 - \tilde{a}_i) < 0, \quad \text{for } i = 2,...,n. \tag{2.15}$$

Then Eq. (1.1) with boundary conditions

$$u_i(x) = 0, \quad i = 1,...,n, \quad T(x) = 0, \quad \text{for } x \in \delta\mathscr{D}$$

has the solution $(0,...,0)$ as the only solution with the properties that each component is in $C^2(\mathscr{D}) \cap C^1(\overline{\mathscr{D}})$ and non-negative in $\overline{\mathscr{D}}$.

Proof: We will consider a parabolic system related to (1.1)

$$\frac{\partial \tilde{u}_i}{\partial t} = \Delta \tilde{u}_i + \sum_{j=1}^{n} H_{ij}(x,\widetilde{T})\tilde{u}_j, \quad i = 1,...,n,$$
$$\frac{\partial \widetilde{T}}{\partial t} = \Delta \widetilde{T} - c(x)\widetilde{T} + \sum_{j=1}^{n} G_j(x,\widetilde{T})\tilde{u}_j, \tag{2.16}$$

for $(x,t) \in \mathscr{D} \times (0,\infty)$, with boundary conditions

$$\tilde{u}_i(x,t) = 0, \quad i = 1,...,n, \quad \widetilde{T}(x,t) = 0, \tag{2.17}$$

for $(x,t) \in \delta\mathscr{D} \times (0,\infty)$. Here $\tilde{u}_i$, $i = 1,...,n$, $\widetilde{T}$ are functions in $\overline{\mathscr{D}} \times [0,\infty)$. We will prove that all solutions of (2.16) and (2.17) with components in $C^2(\mathscr{D} \times (0,\infty)) \cap C^1(\overline{\mathscr{D}} \times [0,\infty))$ and initial conditions which are non-negative for all $x \in \overline{\mathscr{D}}$, $t = 0$, will tend to zero at $t \to +\infty$. Consequently, the equilibrium solution as stated in the theorem can only be the trivial one.

Define $v_1 \equiv v_2 \equiv \cdots \equiv v_{n+1} \equiv 0$. Let $k > 0$ be a constant such that $kc_i\omega(x) \geqslant \tilde{u}_i(x,0)$, $i = 1,...,n$ for each $x \in \overline{\mathscr{D}}$ (c_i are those stated in the theorem). Let $d = \min\{c(x) | x \in \overline{\mathscr{D}}\}$ and σ be a small enough constant with $0 < \sigma < d$ so that inequalities (2.14) and (2.15) $i = 2,...,n$ are all valid with $(-\lambda_1 + \bar{h}_{11})$ and $(-\lambda_1 - \tilde{a}_i)$, respectively, replaced by $(-\lambda_1 + \bar{h}_{11} + \sigma)$ and $(-\lambda_1 - \tilde{a}_i + \sigma)$. Choose $c_{n+1} > 0$ so that

$$c_{n+1} > \max\left\{ \max_{x \in \mathscr{D}} \widetilde{T}(x,0), (d-\sigma)^{-1} \sum_{i=1}^{n} \bar{g}_i\, kc_i \right\}.$$

Finally, define $w_i = kc_i\, \omega(x)e^{-\sigma t}$, $i = 1,...,n$ and $w_{n+1} = c_{n+1}e^{-\sigma t}$. Consider the set

$$J \equiv \{(x,t,z_1,...,z_{n+1}) | (x,t) \in \mathscr{D} \times (0,\infty),$$
$$v_i(x,t) \leqslant z_i \leqslant w_i(x,t), \text{ each } i = 1,...,n+1\}.$$

Clearly, we have for each $i = 1,...,n$

$$\Delta v_i + \sum_{\substack{j=1 \\ j \neq i}}^{n} H_{ij}(x,z_{n+1})z_j + H_{ii}(x,z_{n+1})v_i - \frac{\partial v_i}{\partial t}$$
$$= \sum_{\substack{j=1 \\ j \neq i}}^{n} H_{ij}(x,z_{n+1})z_j \geqslant 0, \tag{2.18}$$

$$\Delta v_{n+1} - c(x)v_{n+1} + \sum_{j=1}^{n} G_j(x,z_{n+1})z_j - \frac{\partial v_{n+1}}{\partial t}$$
$$= \sum_{j=1}^{n} G_j(x,z_{n+1})z_j \geqslant 0, \tag{2.19}$$

for all $(x, t, z_1,..., z_{n+1}) \in J$. On the other hand, for all $(x, t, z_1,..., z_{n+1}) \in J$

$$\Delta w_1 + H_{11}(x,z_{n+1})w_1 + \sum_{j=2}^{n} H_{1j}(x,z_{n+1})z_j - \frac{\partial w_1}{\partial t}$$
$$\leqslant k\omega(x)e^{-\sigma t}\left\{(-\lambda_1 + \bar{h}_{11} + \sigma)c_1 + \sum_{j=2}^{n} c_j \bar{h}_{ij}\right\} < 0, \tag{2.20}$$

$$\Delta w_i + \sum_{\substack{j=1 \\ j \neq i}}^{n} H_{ij}(x,z_{n+1})z_j - A_i(x,z_{n+1})w_i - \frac{\partial w_i}{\partial t}$$

$$\leqslant k\omega(x)e^{-\sigma t}\sum_{\substack{j=1\\ j\neq i}}^{n} c_j\,\bar{h}_{ij} + (-\lambda_1 - \tilde{a}_i + \sigma)c_i\} < 0,$$

$$i = 2,\ldots,n, \tag{2.21}$$

$$\Delta w_{n+1} - c(x)w_{n+1} + \sum_{j=1}^{n} G_j(x,z_{n+1})z_j - \frac{\partial w_{n+1}}{\partial t}$$
$$\leqslant e^{-\sigma t}\left\{(-d+\sigma)c_{n+1} + \sum_{j=1}^{n} \bar{g}_j\, kc_j \max_{x\in\bar{\mathscr{D}}} \omega(x)\right\} < 0, \tag{2.22}$$

because of the choice of $c_j, j = 1,\ldots,n+1$, and σ. Moreover, we have $x\in\bar{\mathscr{D}}$

$$v_i(x,0)\leqslant\tilde{u}_i(x,0)\leqslant w_i(x,0), \quad i = 1,\ldots,n$$

and (2.23)

$$v_{n+1}(x,0)\leqslant\tilde{T}(x,0)\leqslant w_{n+1}(x,0),$$

and for $(x,t)\in\delta\mathscr{D}\times[0,\infty)$,

$$v_i(x,t)\leqslant\tilde{u}_i(x,t)\leqslant w_i(x,t), \quad i = 1,\ldots,n$$

and (2.24)

$$v_{n+1}(x,t)\leqslant\tilde{T}(x,t)\leqslant w_{n+1}(x,t).$$

Therefore, if such a solution $(\tilde{u}_1(x,t),\ldots,\tilde{u}_n(x,t),\tilde{T}(x,t))$ exists in $\bar{\mathscr{D}}\times[0,\infty)$, it will satisfy (2.24) for all $(x,t)\in\bar{\mathscr{D}}\times[0,\infty)$, by inequalities (2.18)–(2.24) above. (See, e.g., Lemma 2.1 in Ref. 8 or Ref. 9 for a variant of the comparison principles used here.)

Let $(\hat{u}_1(x),\ldots,\hat{u}_n(x),\hat{T}(x))$ be a solution of the boundary value problem described in the statement of the theorem, with properties as stated. It will be a solution of (2.16) and (2.17) with the appropriate smoothness and non-negative condition at $t=0$. Letting $(\tilde{u}_1(x,t),\ldots,\tilde{u}_n(x,t),\tilde{T}(x,t)) = (\hat{u}_1(x),\ldots,\hat{u}_n(x),\hat{T}(x))$ for $(x,t)\in\bar{\mathscr{D}}\times[0,\infty)$, inequality (2.24) for $(x,t)\in\bar{\mathscr{D}}\times[0,\infty)$ implies that

$$0\leqslant\hat{u}_i(x)\leqslant kc_i\,\omega(x)e^{-\sigma t}, \quad i = 1,\ldots,n, \quad 0\leqslant\hat{T}(x)\leqslant c_{n+1}e^{-\sigma t},$$

for $(x,t)\in\bar{\mathscr{D}}\times[0,\infty)$. Consequently $(\hat{u}_1,\ldots,\hat{u}_n,\hat{T}) \equiv (0,\ldots,0)$.

We now observe a few very direct consequences of Theorem 2.4. Note that

$$\bar{h}_{ii} \overset{\text{(def)}}{\equiv} \sup\{H_{ii}(x,T)|x\in\bar{\mathscr{D}},\ T\geqslant 0\} = -\tilde{a}_i, \ i = 2,\ldots,n.$$

Define Φ and Ω to be $n\times n$ constant matrices

$$\Phi = [\bar{h}_{ij}], \quad \text{whose } (i,j)\text{th entry is } \bar{h}_{ij}, \ 1\leqslant i,j\leqslant n, \tag{2.25}$$

$$\Omega = \Phi - \lambda_1 I, \quad \text{where } I \text{ is the identity matrix.}$$

If (2.13) is satisfied, the matrix Ω has all its diagonal entries negative and all other entries positive. Suppose further that Ω is strictly "diagonally dominant," i.e.,

$$|\bar{h}_{11} - \lambda_1| > \sum_{\substack{j=1\\ j\neq i}}^{n} \bar{h}_{ij},$$

then (2.14) and (2.15) are satisfied by choosing $c_1 = c_2 = \cdots = c_n = 1$. Consequently the next corollary is true.

Corollary 2.5: Suppose that the $n\times n$ constant matrix Ω in (2.25) is strictly "diagonally dominant"; then the boundary value problem in Theorem 2.4 has the solution $(0,\ldots,0)$ as the only solution with the properties that each component is in $C^2(\mathscr{D})\cap C^1(\bar{\mathscr{D}})$ and non-negative in $\bar{\mathscr{D}}$.

Remark: Suppose that (2.13) is true and $n = 2$ in (1.1). If

$$\bar{h}_{21}\,\bar{h}_{12} < (\tilde{a}_2 + \lambda_1)(\lambda_1 - \bar{h}_{11})$$

is satisfied, choose c_1,c_2 to be positive constants so that $(\tilde{a}_2 + \lambda_1)\bar{h}_{21}^{-1} > c_1c_2^{-1} > \bar{h}_{12}(\lambda_1 - \bar{h}_{11})^{-1}$; then (2.14) and (2.15) are satisfied and Theorem 2.4 can be applied.

III. POSITIVE EQUILIBRIA FOR PROMPT FEEDBACK MULTIGROUP EQUATIONS

Conditions for the existence of positive steady states were not found for the general model (1.1) in the previous section, under zero Dirichlet boundary condition. To make the analysis more tractable, we consider a slightly simpler model. We now assume that the reaction coefficients (i.e., cross sections) are functions of the neutron fluxes u_i directly. That is, the feedback is prompt, and does not have to be regulated through the change in T indirectly through the last equation in (1.1). More precisely, we have

$$\Delta u_i + \sum_{j=1}^{n} H_{ij}(u_1,\ldots,u_n)u_j = 0, \text{ in } \mathscr{D}, \quad i = 1,\ldots,n, \tag{3.1}$$

$$u_i(x) = 0, \quad x\in\delta\mathscr{D}, \quad i = 1,\ldots,n. \tag{3.2}$$

As in (2.5), we first restrict ourselves to the directly coupled scattering case

$$\begin{aligned} &H_{1j}(u_1,\ldots,u_n) > 0, \quad j = 2,\ldots,n. \\ &\text{For } i>1: \quad H_{i,i-1}(u_1,\ldots,u_n) > 0, \\ &\qquad\qquad H_{ii} \overset{\text{(def)}}{=} -A_i(u_1,\ldots,u_n) < 0, \\ &\qquad\qquad H_{i,j} \equiv 0 \text{ if } j>i \text{ or } j<i-1. \end{aligned} \tag{3.3}$$

All formulas are valid for $u_k\geqslant 0$, each $k = 1,\ldots,n$. Define

$$h'_{ij} = \inf\{H_{ij}(u_1,\ldots,u_n)|u_k\geqslant 0, k = 1,\ldots,n\},$$
$$h''_{ij} = \sup\{H_{ij}(u_1,\ldots,u_n)|u_k\geqslant 0, k = 1,\ldots,n\},$$

$i,j = 1,\ldots,n$. The functions $u_j\,H_{ij}(u_1,\ldots,u_n)$, $i,j = 1,\ldots,n$ are assumed to belong to the class C^α in the set $\{(u_1,\ldots,u_n)|u_k\geqslant 0, k = 1,\ldots,n\}$, i.e., they are locally Hölder continuous in $(u_1,\ldots,u_n)$ with Hölder exponent α, $0<\alpha<1$. Let $C^{2+\alpha}(\bar{\mathscr{D}})$ denote the Banach space of real-valued functions in $\bar{\mathscr{D}}$, with first and second derivatives also continuous in $\bar{\mathscr{D}}$, with finite value for the usual norm $|u|^{(2+\alpha)}_{\mathscr{D}}$. We assume the boundary $\delta\mathscr{D}$ belongs to class $C^{2+\alpha}$ (see, e.g., Ref. 10 for details of these symbols). The following three conditions will be assumed.

(P1) $-\infty < h'_{11}\leqslant h''_{11} < \infty$, $0 < h'_{1j}\leqslant h''_{1j} < \infty$, for $j = 2,\ldots,n$;

$0 < h'_{i,i-1} < h''_{i,i-1} < \infty$, for $i = 2,\ldots,n$;

$a'_i \overset{\text{(def)}}{=} -h''_{ii} > 0$, $a''_{ii} \overset{\text{(def)}}{=} -h'_{ii} < \infty$ for $i = 2,\ldots,n$

(hence $0 < a'_i\leqslant A_i\leqslant a''_i < \infty$).

(P2) $H_{11}(0,\ldots,0) > \lambda_1$. In the set $M \overset{\text{(def)}}{=} \{(u_1,\ldots,u_n)|u_k$

$\geqslant 0$, $k=1,...,n\}$, $H_{11}(u_1,...,u_n)$ is continuously differentiable with respect to $u_2,...,u_n$, and $|\partial H_{11}/\partial u_j| < K$ for all $(u_1,...,u_n) \in M$, $j=2,...,n$, where K is some positive constant.

(P3) There exist positive constants p and U^* such that

$$H_{11}(u_1,...,u_n) \leqslant -p, \text{ for all } (u_1,...,u_n) \in M \text{ with } u_1 \geqslant U^*.$$

We have the following existence theorem for positive steady state, in the directly coupled scattering case.

Theorem 3.1 (directly coupled scattering): Suppose there exist positive constants $\beta_i > 0$, $i=2,...,n$ with $\Sigma_{i=2}^n \beta_i \leqslant 1$ such that

$$h''_{21} h''_{12} < \beta_2 a'_2 p,$$
$$\beta_{i-1} h''_{i,i-1} h''_{1i} < \beta_i a'_i h''_{1,i-1}, \quad i=3,...,n. \tag{3.4}$$

Then the boundary value problem (3.1)–(3.3) under the conditions (P1)–(P3) has a solution $(\hat{u}_1(x),...,\hat{u}_n(x))$ with components in $C^{2+\alpha}(\overline{\mathscr{D}})$ and $\hat{u}_i(x) > 0$ in $\mathscr{D}$, $i=1,...,n$. [Here, the numbers p and U^* satisfy (P3) together with (3.4).]

Proof: We will construct upper and lower solutions for (3.1)–(3.3) and apply a theorem in Ref. 11 to conclude the existence of a positive solution. By (P2), there is a small constant $k>0$ so that $H_{11}(u,0,...,0) > \lambda_1$ for $0 \leqslant u < k$. Choose $0 < \epsilon < \min\{k, K^{-1}h'_{12},..., K^{-1}h'_{1n}\}$, $0 < \delta_2 < \epsilon h'_{21}[\lambda_1 + a''_2]^{-1}$, and $0 < \delta_i < \delta_{i-1} h'_{i,i-1}[\lambda_1 + a''_i]^{-1}$, for $i=3,...,n$. Define lower solutions as

$$v_1(x) = \epsilon\omega(x), \quad v_i(x) = \delta_i \omega(x), \quad i=2,...,n,$$

for $x \in \overline{\mathscr{D}}$. Define upper solutions as

$$w_1(x) = U^*, \quad w_i(x) = \beta_i p U^*/h''_{1i}, \quad i=2,...,n,$$

for $x \in \overline{\mathscr{D}}$. We now check the appropriate inequalities for the v_i, w_i, $i=1,...,n$. We have

$$\Delta v_1 + H_{11}(v_1,u_2,...,u_n)v_1 + \sum_{j=2}^n H_{1j}(v_1,u_2,...,u_n)u_j$$
$$> \epsilon\omega(x)[-\lambda_1 + H_{11}(\epsilon\omega(x),u_2,...,u_n)] + \sum_{j=2}^n h'_{1j} u_j,$$

for $u_j \geqslant 0$, $j=2,...,n$, $x \in \mathscr{D}$. However, $[-\lambda_1 + H_{11}(\epsilon\omega,0,...,0)] > 0$ in $\mathscr{D}$, and $F(s,x,u_2,...,u_n) \overset{\text{(def)}}{\equiv} \epsilon\omega(x)[-\lambda_1 + H_{11}(\epsilon\omega(x),su_2,...,su_n)] + \Sigma_{j=2}^n h'_{1j} su_j$ is an increasing function of $s \geqslant 0$, for fixed $u_j \geqslant 0$, $j=2,...,n$, each $x \in \mathscr{D}$ (by the choice of ϵ). Consequently, we have

$$\Delta v_1(x) + H_{11}(v_1(x),u_2,...,u_n)v_1 + \sum_{j=2}^n H_{1j}(v_1(x),u_2,...,u_n)u_j > 0, \tag{3.5}$$

for all $u_j \geqslant 0$, $j=2,...,n$, $x \in \mathscr{D}$. For $v_2(x)$, we have

$$\Delta v_2(x) + H_{21}(u_1,v_2(x),u_3,...,u_n)u_1 - A_2(u_1,v_2(x),u_3,...,u_n)v_2(x)$$
$$\geqslant \delta_2\omega(x)[-\lambda_1 - a''_2] + h'_{21}u_1$$
$$> -\epsilon h'_{21}\omega(x) + h'_{21}\epsilon\omega(x) = 0, \tag{3.6}$$

for $v_j(x) \leqslant u_j \leqslant w_j(x)$, $j \neq 2$, $x \in \mathscr{D}$. For $v_i(x)$, $i=3,...,n$, we have the inequalities

$$\Delta v_i(x) + H_{i,i-1}(u_1,...,u_{i-1},v_i(x),u_{i+1},...,u_n)u_{i-1} - A_i(u_1,...,v_i(x),...,u_n)v_i(x)$$
$$\geqslant \delta_i\omega(x)[-\lambda_1 - a''_i] + h'_{i,i-1} u_{i-1}$$
$$> -\delta_{i-1} h'_{i,i-1}\omega(x) + h'_{i,i-1}\delta_{i-1}\omega(x) = 0, \tag{3.7}$$

for $v_j(x) \leqslant u_j \leqslant w_j(x)$, $j \neq i$, $x \in \mathscr{D}$ (by the choice of δ_i). For the upper solutions, we have

$$\Delta w_1(x) + H_{11}(w_1(x),u_2,...,u_n)w_1(x) + \sum_{j=2}^n H_{1j}(w_1(x),u_2,...,u_n)u_j$$
$$\leqslant -pU^* + \sum_{j=2}^n h''_{1j}\beta_j pU^*(h''_{1j})^{-1} \leqslant 0, \tag{3.8}$$

for $v_j(x) \leqslant u_j \leqslant w_j(x)$, $j=2,...,n$, $x \in \mathscr{D}$ (since $\Sigma_{j=2}^n \beta_j \leqslant 1$), and

$$\Delta w_i(x) + H_{i,i-1}(u_1,...,u_{i-1},w_i(x),u_{i+1},...,u_n)u_{i-1} - A_i(u_1,...,w_i(x),...,u_n)w_i(x)$$
$$\leqslant h''_{i,i-1} w_{i-1} - a'_i \beta_i U^* p(h''_{1i})^{-1} < 0, \tag{3.9}$$

for each $i=2,...,n$, $v_j(x) \leqslant u_j \leqslant w_j(x)$, $j \neq i$, $x \in \mathscr{D}$ [by the properties of β_i, $i=2,...,n$ in (3.4)]. By Ref. 11, (3.5)–(3.9) imply that there exists a solution $(\hat{u}_1(x),...,\hat{u}_n(x))$ as described in the statement of the theorem with $v_i(x) \leqslant \hat{u}_i(x) \leqslant w_i(x)$, $i=1,...,n$, $x \in \overline{\mathscr{D}}$. Consequently $\hat{u}_i(x) > 0$ in $\mathscr{D}$, $i=1,...,n$.

Remark: A simple situation with which conditions in (3.4) are all satisfied is

$$h''_{21} h''_{12} < a'_2 p/(n-1),$$
$$h''_{i,i-1} h''_{1i} < a'_i h''_{1,i-1}, \quad i=3,...,n.$$

This corresponds to having $\beta_i = (n-1)^{-1}$, $i=2,...,n$ in (3.4). Theorem 3.1 is then applicable.

We now return to the more general case of down scattering. Properties (3.3) will be replaced by

$$\begin{aligned} &H_{1j}(u_1,...,u_n) > 0, \quad j=2,...,n; \\ &\text{for } i>1: \quad H_{ij}(u_1,...,u_n) > 0, \text{ if } j<i, \\ &\qquad H_{ii} \overset{\text{(def)}}{=} -A_i(u_1,...,u_n) < 0, \\ &\qquad H_{ij} \equiv 0, \quad \text{if } j>i. \end{aligned} \tag{3.3'}$$

All formulas are valid for $u_k \geqslant 0$, each $k=1,...,n$.

Hypothesis (P1) will be modified accordingly to (P1a)

$$-\infty < h'_{11} < h''_{11} < \infty, \quad 0 < h'_{1j} < h''_{1j} < \infty, \quad \text{for } j=2,...,n;$$
$$0 < h'_{ij} < h''_{ij} < \infty, \quad \text{for } j<i, \ i=2,...,n;$$
$$a'_i \overset{\text{(def)}}{=} -h''_{ii} > 0, \quad a''_i \overset{\text{(def)}}{=} -h'_{ii} < \infty, \quad \text{for } i=2,...,n.$$

The following is a more involved existence theorem for positive steady state.

Theorem 3.2: Suppose there exist positive constants $\beta_i > 0$, $i=2,...,n$ with $\Sigma_{i=2}^n \beta_i \leqslant 1$ such that

$$h''_{21} h''_{12} < \beta_2 a'_2 p,$$
$$h''_{1i}\left[h''_{i1} + \sum_{j=2}^{i-1} \beta_j h''_{ij} p(h''_{1j})^{-1}\right] < \beta_i a'_i p, \quad i=3,...,n.$$

Then the boundary value problem (3.1), (3.2), and (3.3′) under the conditions (P1a), (P2), and (P3) has a solution $(\hat{u}_1(x),...,\hat{u}_n(x))$ with components in $C^{2+\alpha}(\overline{\mathscr{D}})$ and $\hat{u}_i(x) > 0$ in $\mathscr{D}$, $i = 1,...,n$.

The proof is exactly analogous to that of Theorem 3.1. Details are therefore omitted.

[1]J. J. Duderstadt and L. J. Hamilton, *Nuclear Reactor Analysis* (Wiley, New York, 1976).

[2]G. J. Habetler and M. A. Martino, "Existence theorems and spectral theory for the multigroup diffusion model, Nuclear Reactor Theory," in *Proceedings of Symposia in Applied Mathematics, Vol. XI* (Am. Math. Soc., Providence, RI, 1961), pp. 127–139.

[3]R. S. Varga, "Numerical methods for solving multi-dimensional multigroup diffusion equations, Nuclear Reactor Theory," in *Proceedings of Symposia in Applied Mathematics, Vol. XI* (Am. Math. Soc., Providence, RI, 1961), pp. 146–189.

[4]I. Stakgold and L. E. Payne, "Nonlinear problems in nuclear reactor analysis," in *Proceedings of the Conference on Nonlinear Problems in Physical Sciences and Biology, Springer Lecture Notes in Mathematics, No. 322* (Springer, New York, 1972), pp. 298–307.

[5]W. E. Kastenberg, Nucl. Sci. Eng. **37**, 19 (1969).

[6]W. E. Kastenberg, "Stability analysis of nonlinear space dependent reactor kinectics," in *Advances in Nuclear Science and Technology* (Academic, New York, 1969), Vol. 5, pp. 51–91.

[7]A. Leung and G. S. Chen, "Positive solutions for temperature-dependent two-group neutron flux equations: Equilibria and stabilities," SIAM J. Math. Anal. **15**, 490 (1984).

[8]A. Leung, J. Math. Anal. Appl. **73**, 204 (1980).

[9]W. Walter, *Differential and Integral Inequalities* (Springer, Berlin, 1970).

[10]A. Friedman, *Partial Differential Equations* (Holt, Rhinehart, and Winston, New York, 1969).

[11]Y. Tsai, Bull. Inst. Math. Acad. Sinica **5**, 157 (1977).

Differential-Stäckel matrices

E. G. Kalnins
Mathematics Department, University of Waikato, Hamilton, New Zealand
Willard Miller, Jr.
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

(Received 28 November 1984; accepted for publication 8 February 1985)

We show that additive separation of variables for linear homogeneous equations of all orders is characterized by differential-Stäckel matrices, generalizations of the classical Stäckel matrices used for multiplicative separation of (second-order) Schrödinger equations and additive separation of Hamilton–Jacobi equations. We work out the principal properties of these matrices and demonstrate that even for second-order Laplace equations additive separation may occur when multiplicative separation does not.

I. INTRODUCTION

Our motivation for this study of additive separation of variables for linear differential equations was the following example in Ref. 1:

$$(x_1+x_2)(\partial_{11}u+\partial_{22}u)-2(\partial_1 u+\partial_2 u)=E.$$

This equation admits a five-parameter separable solution in the coordinates x_1, x_2:

$$u=(\alpha x_1^3+\beta x_1^2+\gamma x_1-\tfrac{1}{2}Ex_1)+(-\alpha x_2^3+\beta x_2^2-\gamma x_2+\delta).$$

The mechanism of separation was puzzling to us until we realized that the appropriate separation equations are

$$\begin{aligned}
&\partial_1 u + E/2 - \gamma - 2\beta x_1 - 3\alpha x_1^2 = 0,\\
&\partial_{11}u - 2\beta - 6\alpha x_1 = 0,\\
&\partial_2 u + \gamma - 2\beta x_2 + 3\alpha x_2^2 = 0,\\
&\partial_{22}u - 2\beta + 6\alpha x_2 = 0.
\end{aligned}$$

The associated "Stäckel matrix" responsible for the separation is

$$\begin{bmatrix} \frac{1}{2} & -1 & -2x_1 & -3x_1^2\\ 0 & 0 & -2 & -6x_1\\ 0 & 1 & -2x_2 & 3x_2^2\\ 0 & 0 & -2 & 6x_2 \end{bmatrix}.$$

This is not a true Stäckel matrix since more than one row depends on a given variable x_I (Refs. 2 and 3). Moreover, the second and fourth rows are the derivatives of the first and third rows, respectively. It is a nontrivial example of a differential-Stäckel matrix.

In this paper we show that the above example is not isolated. All additive separation of nth-order linear differential equations $L=E$ or $L=0$ is associated with differential-Stäckel matrices. In Sec. II we derive, in the form of a coupled system of nonlinear partial differential equations, necessary and sufficient conditions that a linear differential equation admits additive separation in a given coordinate system. In Sec. III we develop the principal properties of the matrices inverse to differential-Stäckel matrices, and in Sec. IV we find all solutions of the separability conditions and show that they correspond to differential-Stäckel matrices. Our method is an extension of Eisenhart's study of true Stäckel matrices.[3] In Sec. V we comment on the relation between multiplicative separation and additive separation for Laplace equations on Riemannian manifolds, and we give an example to show that additive separation may occur for a Laplace equation in a given coordinate system even when multiplicative separation is absent.

All functions appearing in this paper are assumed to be locally real analytic.

II. ADDITIVE SEPARABILITY FOR LINEAR DIFFERENTIAL EQUATIONS

In Ref. 1 the authors introduced a general definition of additive separation of variables for a partial differential equation

$$H(x_I,u,u_I,u_{IJ},\ldots)=E, \tag{2.1}$$

in the coordinates $x_1,\ldots,x_N$. Here u is the dependent variable, $u_I=\partial_{x_I}u$, $u_{IJ}=\partial_{x_I}\partial_{x_J}u$, etc., and E is a parameter. A separable solution of (2.1) is a solution of the form $u=\Sigma_{J=1}^N S^{(J)}(x_J,E)$. We briefly review this definition (a generalization of that of Levi-Civita[4] and its simple consequences. (See Ref. 5 for a discussion of other definitions of separability.)

For convenience we suppose H is a polynomial in the derivatives $u_I,u_{IJ},\ldots$. Furthermore, there is no loss of generality in setting all mixed partial derivatives identically equal to zero (since $u_{IJ}=0$ for $I\neq J$ if u is a separable solution) and writing (2.1) in the form

$$H(x_I,u,u_I,u_{II},\ldots)=E. \tag{2.2}$$

We introduce the new notation $u_{I,1}=u_I$, $u_{I,i+1}=\partial_{x_I}u_{I,i}$, $i=1,2,\ldots$, and define n_I to be the largest number l such that $\partial u_{I,l}H=H_{u_{I,l}}\not\equiv 0$. To avoid discussion of degenerate cases we require $n_I>0$ for $I=1,\ldots,N$.

Let the truncated differentiation operator $\widehat{D}_I$ be defined by

$$\widehat{D}_I=\partial_{x_I}+u_{I,1}\,\partial_u+\cdots+u_{I,n_I}\,\partial_{u_{i,n_I-1}}.$$

In Ref. 1 we showed that every separable solution u of (2.2) satisfies the integrability conditions

$$\begin{aligned}
&H_{u_{i,n_I}}H_{u_{J,n_J}}(\widehat{D}_I\widehat{D}_J H)+H_{u_{I,n_I}u_{J,n_J}}(\widehat{D}_I H)(\widehat{D}_J H)\\
&\quad -H_{u_{J,n_J}}(\widehat{D}_I H)(\widehat{D}_J H_{u_{I,n_I}})-H_{u_{I,n_I}}(\widehat{D}_J H)(\widehat{D}_I H_{u_{J,n_J}})\\
&=0,\quad 1\leqslant I<J\leqslant N.
\end{aligned} \tag{2.3}$$

If (2.3) is an identity in the dependent variables $u, u_{K,k}$ then we say that $\{x_I\}$ is a regular separable coordinate system. In this case the separable solutions involve $\Sigma_{J=1}^N n_J + 1$ independent parameters; that is, at a fixed point $\mathbf{x}^0$ the separable solutions are uniquely determined by prescribing $u(\mathbf{x}^0)$ and the $\Sigma_{J=1}^N n_J$ derivatives $u_{I,i}(\mathbf{x}^0)$, $1 \leqslant I \leqslant N$, $1 \leqslant i \leqslant n_I$. If the integrability conditions (2.3) do not hold identically then the separation is nonregular; separable solutions may exist but they will involve (strictly) fewer parameters than the regular case. In the following when we speak of variable separation we mean regular separation. [Note that multiplicative separation can easily be treated by the preceding definition since $v = \Pi_{J=1}^N T^{(J)}(x_J)$ is multiplicatively separable if and only if $u = \ln v$ is additively separable.]

For Laplacelike equations

$$H(x_I, u, u_I, u_{II}, \ldots) = 0, \tag{2.4}$$

there is a minor modification of the integrability conditions. Denoting by F_{IJ} the left-hand side of Eqs. (2.3) we can state the integrability conditions for (2.4) in the form

$$F_{IJ} = P_{I,J} H, \quad 1 \leqslant I < J < N, \tag{2.5}$$

where $P_{I,J}(x_K, u, u_{K,k})$ are polynomials in $u_{K,k}$. If (2.5) is satisfied identically in the dependent variables $u, u_{J,j}$ we say that $\{x_K\}$ is a regular separable coordinate system. In this case the separable solutions depend on $\Sigma_{J=1}^N n_J$ independent parameters. For nonregular separation the separable solutions depend on fewer parameters.

Now we will apply these criteria to determine additive separability conditions for the linear equations

$$L = E \tag{2.6}$$

and

$$L = 0, \tag{2.7}$$

where

$$L = \sum_{J=1}^{N} \sum_{j=1}^{n_J} H_{(J,j)}(x) u_{J,j}, \quad H_{(J,n_J)} \neq 0. \tag{2.8}$$

Introducing the abbreviation $H_J \equiv H_{(J,n_J)}$ we can write the integrability conditions (2.3) for $L = E$ in the form

$$\hat{D}_I \hat{D}_J L - \hat{D}_I L\, \partial_J \ln H_I - \hat{D}_J L\, \partial_I \ln H_J = 0, \quad I \neq J, \tag{2.9}$$

where $\partial_I = \partial_{x_I}$. Equating to zero the coefficients of the derivatives $u_{J,j}$ on the left-hand side of (2.9) we obtain the following necessary and sufficient conditions for regular separation:

$$\partial_{IJ} H_{(P,p)} - \partial_I H_{(P,p)}\, \partial_J \ln H_I - \partial_J H_{(P,p)}\, \partial_I \ln H_J = 0,$$
$$P \neq I, J, \quad p = 1, \ldots, n_P, \tag{2.10a}$$

$$\partial_{IJ} H_{(J,j)} - \partial_I H_{(J,j)}\, \partial_J \ln H_I - \partial_J H_{(J,j)}\, \partial_I \ln H_J$$
$$= H_{(J,j-1)}\, \partial_I \ln H_J - \partial_I H_{(J,j-1)}, \tag{2.10b}$$
$$j = 1, \ldots, n_J.$$

Here $I \neq J$, $H_{(J,0)} \equiv 0$. In terms of the linear operators

$$A_{IJ} = \partial_{IJ} - \partial_J \ln H_I \partial_I - \partial_I \ln H_J\, \partial_J,$$
$$B_{IJ} = -\partial_I + \partial_I \ln H_J, \quad I \neq J, \tag{2.11}$$

these conditions can be written as

$$A_{IJ} H_{(P,p)} = 0, \quad P \neq I, J,$$
$$A_{IJ} H_{(J,j)} = B_{IJ} H_{(J,j-1)}, H_{(J,0)} = 0, \quad I \neq J. \tag{2.12}$$

[The possibility that L, Eq. (2.8), has an additive term of the form $H(x)u$ can be treated as a special case of our considerations. Formally Eqs. (2.10) are still the separation conditions for $L = E$ where now $H_{(J,0)} = H$, $H_{(J,-1)} = 0$, $J = 1, \ldots, N$, and the index j takes the values $0, 1, \ldots, n_J$.]

Similarly, the integrability conditions (2.5) for the homogeneous equation $L = 0$ take the form

$$A_{IJ} H_{(P,p)} = Q_{IJ}(x) H_{(P,p)}, \quad P \neq I, J,$$
$$A_{IJ} H_{(J,j)} = B_{IJ} H_{(J,j-1)} + Q_{IJ}(x) H_{(J,j)}, \tag{2.13}$$
$$H_{(J,0)} = 0, \quad I \neq J,$$

where Q_{IJ} is a function of the independent variables x_K alone.

The two sets of integrability conditions are closely related.

Lemma 1: If the functions $\{H_{(I,i)}\}$ satisfy (2.12) and $R(x)$ is nonzero then the functions $\{H'_{(I,i)} = R H_{(I,i)}\}$ satisfy (2.13). Indeed

$$-Q_{IJ} = 2\, \partial_I \ln R \partial_J \ln R$$
$$+ \partial_I \ln R\, \partial_J \ln H_I + \partial_J \ln R \partial_I \ln H_J.$$

Suppose the functions $\{H_{(I,i)}\}$, $I = 1, \ldots, N$, $i = 1, \ldots, n_I$ are not all zero and set $H_{I,0} = 0$. Then there must exist some $H_{(K,k)} \neq 0$ such that $H_{(K,k-1)} = 0$. Let $H'_{(I,i)} = H_{(I,i)}/H_{(K,k)}$.

Lemma 2: The functions $\{H_{(I,i)}\}$ satisfy conditions (2.13) if and only if the functions $\{H'_{(I,i)}\}$ satisfy (2.12), i.e.,

$$\partial_{IJ} H'_{(P,p)} - \partial_J \ln H'_I \partial_I H'_{(P,p)} - \partial_I \ln H'_J \partial_J H'_{(P,p)} = 0,$$
$$P \neq I, J, \tag{2.14}$$

$$\partial_{IJ} H'_{(J,j)} - \partial_J \ln H'_I \partial_I H'_{(J,j)} - \partial_I \ln H'_J\, \partial_J H'_{(J,j)}$$
$$= H'_{(J,j-1)}\, \partial_I \ln H'_J - \partial_I H'_{(J,j-1)},$$

for $I \neq J$.

It follows from these lemmas that all sets of functions $\{H_{(I,i)}\}$ satisfying conditions (2.13) are of the form $H_{(I,i)} = R H'_{(I,i)}$, where the $\{H'_{(I,i)}\}$ satisfy conditions (2.12). In the next section we will show how to find all solutions of Eqs. (2.12).

III. D-STÄCKEL MATRICES

Consider a coordinate set $x_1, \ldots, x_N$ and let $n_1, \ldots, n_N$ be positive integers with $n = \Sigma_{I=1}^n n_I$. Let $S = (S_{(I,i),l}(x_I))$ be an $n \times n$ matrix with the properties following.

(1) $$S_{(I,i),l}(x_I) = \frac{d^{i-1}}{dx_I^{i-1}} S_{(I,1),l}(x_I),$$
$$i = 1, 2, \ldots, n_I. \tag{3.1}$$

[Here, the rows of S are designated by the index (I,i), where $I = 1, \ldots, N$, $i = 1, \ldots, n_I$. The columns of S are designated by the index $l = 1, 2, \ldots, n$. Thus row (I,i) depends only on the variable x_I and is the $i-1$ derivative of row $(x,1)$.]

(2) $\det S \neq 0$.

(3) $T^{1,(J,j)} \neq 0$, $J = 1, \ldots, N$, $h = 1, \ldots, n_J$, where $T = S^{-1}$, i.e.,

$$\sum_{l=1}^{n} S_{(I,i),l}(x_I)T^{l,(J,j)} = \delta_{(I,i)}^{(J,j)}. \qquad (3.2)$$

We say that a matrix S satisfying properties (1)–(3) is a differential-Stäckel matrix (D-Stäckel matrix). If $n_1 = \cdots = n_N = 1$ then S is simply the usual Stäckel matrix.[2,3] In order to obtain results about D-Stäckel matrices that are useful for separation of variables we need to characterize the inverse matrix T. For this we generalize Eisenhart's study of Stäckel matrices.[3,6]

Differentiating (3.2) with respect to x_I, we obtain

$$\sum_I S_{(I,i+1),l}(x_I)T^{l,(J,j)} + \sum_I S_{(I,i),l}\, \partial_{x_I} T^{l,(J,j)} = 0, \qquad (3.3)$$

where we adopt the convention $S_{(I,n_I+1),l} = 0$. Since S is nonsingular it follows that

$$\partial_{x_I} T^{l,(J,j)} = f_I^{(J,j)}\, T^{l,(I,n_I)} - T^{l,(J,j-1)}\delta_I^J, \qquad (3.4)$$

where $f_I^{(J,j)}$ is a function and we adopt the convention $T^{l,(J,0)} = 0$.

Now set

$$T^{l,(J,j)} = \rho_{(J,j)}^l H_{(J,j)}, \quad \rho_{(J,j)}^1 = 1. \qquad (3.5)$$

In particular, $T^{1,(J,j)} = H_{(J,j)}$. We will characterize T in terms of the "roots" $\rho_{(J,j)}^l$ and the $H_{(J,j)}$. Substituting (3.5) in (3.4) we obtain

$$\partial_I \rho_{(J,j)}^l H_{(J,j)} + \rho_{(J,j)}^l\, \partial_I H_{(J,j)} = f_I^{(J,j)} \rho_{(I,n_I)}^l H_{(I,n_I)} - \rho_{(J,j-1)}^l H_{(J,j-1)}\delta_I^J, \qquad (3.6)$$

where $\rho_{(J,0)}^l = H_{(J,0)} = 0$. For $l = 1$, Eq. (3.6) reduces to

$$\partial_I H_{(J,j)} = f_I^{(J,j)} H_{(I,n_I)} - H_{(J,j-1)}\delta_I^J, \qquad (3.7)$$

in view of (3.5). Solving this expression for $f_I^{(J,j)}H_{(I,n_I)}$ and substituting into (3.6) we obtain the desired characterization

$$\partial_I \rho_{(J,j)}^l = (\rho_{(I,n_I)}^l - \rho_{(J,j)}^l)\partial_I \ln H_{(J,j)} + (\rho_{(I,n_I)}^l - \rho_{(J,j-1)}^l)\frac{H_{(J,j-1)}}{H_{(J,j)}}\delta_I^J, \qquad (3.8)$$

$I, J = 1,\ldots,N$, $h = 1,\ldots,n_J$.

At this point we have shown that if S is a D-Stäckel matrix then the system of equations

$$\partial_I \rho_{(J,j)} = (\rho_{(I,n_I)} - \rho_{(J,j)})\partial_I \ln H_{(J,j)} + (\rho_{(I,n_I)} - \rho_{(J,j-1)})\frac{H_{(J,j-1)}}{H_{(J,j)}}\delta_I^J, \qquad (3.9)$$

$I,J = 1,\ldots,N$, $h = 1,\ldots,n_J$, where $H_{(J,j)} = T^{1,(J,j)}$ admits a full linearly independent set of n vector-valued solutions $\{\rho_{(J,j)}^l\}$, $l = 1,\ldots,n$.

Conversely, suppose we are given a set of n nonzero functions $\{H_{(J,j)}\}$ such that the system (3.9) admits a full linearly independent set of n vector-valued solutions $\{\rho_{(J,j)}^l\}$. Since $\rho_{(J,j)} \equiv 1$, all J, j, is a solution, without loss of generality we can include it in our basis set and assume $\rho_{(J,j)}^1 = 1$. It follows that the $n \times n$ matrix T defined by (3.5) is invertible. Let $S = T^{-1}$, i.e.,

$$\sum_{l=1}^{n} S_{(I,i),l} T^{l,(J,j)} = \delta_{(I,i)}^{(J,j)}. \qquad (3.10)$$

It follows from (3.9) and (3.5) that (3.4) holds with

$$f_I^{(J,j)} = H_{(I,n_I)}^{-1}(\partial_I H_{(J,j)} + H_{(J,j-1)}\delta_I^J).$$

Differentiating both sides of (3.10) with respect to x_K and using (3.4), we find

$$\sum_{l=1}^{n} \partial_K S_{(I,i),l} T^{l,(J,j)} = \delta_K^J \delta_{(I,i)}^{(J,j-1)} - f_K^{(J,j)}\delta_{(I,i)}^{(K,n_K)}. \qquad (3.11)$$

It follows that $\partial_K S_{(I,i),l} = 0$ if $K \neq I$ and $\partial_I S_{(I,i),l} = S_{(I,i+1),l}$ for $i = 1,\ldots,n_I - 1$. Thus S is a D-Stäckel matrix.

Theorem 1: Let $\{H_{(J,j)}\}$ ($J = 1,\ldots,N$, $j = 1,\ldots n_J$, $\Sigma_J n_J = n$) be a set of n nonzero functions of the N variables x_I. There exists an $n \times n$ D-Stäckel matrix $S = (S_{(I,i),l}(x_I))$ with inverse $T = (T^{l,(J,j)})$ such that $H_{(J,j)} = T^{1,(J,j)}$ if and only if $\{H_{J,j)}\}$ satisfies Eqs. (2.10a) and (2.10b).

Proof: It is straightforward to verify that (2.10a) and (2.10b) are simply the integrability conditions $\partial_K(\partial_I \rho_{(J,j)}) = \partial_I(\partial_K \rho_{(J,j)})$, $K \neq I$, for the system (3.9). Thus if $\{H_{(J,j)}\}$ satisfies the integrability conditions then (3.9) has n independent vector-valued solutions and we can construct a D-Stäckel matrix S such that $H_{(J,j)} = T^{1,(J,j)}$.

Conversely, if $H_{(J,j)} = T^{1,(J,j)}$ and $T^{-1} = S$ for some D-Stäckel matrix S then the system (3.9) admits n independent vector-valued solutions and the integrability conditions (2.10) must be satisfied.

We now have a partial answer to the problems posed in Sec. II. Consider the equation $L = E$ in N independent variables x_I, where

$$L = \sum_{J=1}^{N} \sum_{j=1}^{n_J} H_{(J,j)}(x) u_{J,j},$$

and suppose each of the $H_{(J,j)}$ is nonzero. This equation admits (regular) additively separable solutions provided the conditions (2.10) are satisfied. These conditions imply the existence of a D-Stäckel matrix S such that $H_{(J,j)} = T^{1,(J,j)}$. The separation equations are evident

$$u_{J,j} + \sum_{l=1}^{n} S_{(J,j),l}(x_J)\lambda_l = 0, \qquad (3.12)$$

$$1 \leqslant J \leqslant N, \quad 1 \leqslant j \leqslant n_J, \quad \lambda_1 = -E.$$

Here there are n separation parameters λ_l. The separable solutions u are obtained by integrating the N first-order ordinary differential equations

$$u_{J,1} + \sum_{l=1}^{n} S_{(J,1),l}(x_J)\lambda_l = 0. \qquad (3.13)$$

The remaining $n - N$ equations are redundant since they are obtained by differentiating the basic set (3.13). The number of parameters in the solution u is $n + \Sigma_J n_J + 1$, in agreement with the prediction in Sec. II. Multiplying the separation equation (3.12) for $u_{J,j}$ by $T^{1,(J,j)}$ and summing over the index (J,j) we once again obtain $L = E$ for $E = -\lambda_1$.

The treatment for the equation $L = 0$ is similar. Suppose

$$L = \sum_{J=1}^{N} \sum_{j=1}^{n_J} H'_{(J,j)}(x) u_{J,j},$$

where none of the $H'_{(J,j)}$ is zero and suppose these functions satisfy the separability conditions (2.14). Then there is a nonzero function $R(x)$ such that $H'_{(J,j)} = RH_{(J,j)}$, where the $H_{(J,j)}$ satisfy conditions (2.10) and, thus, determine a D-Stäckel matrix S. The separation equations are (3.12) with $\lambda_1 = 0$. There are $n - 1$ separation parameters and a separated solution contains $n = \Sigma_J n_J$ parameters. Multiplying the separation equation (3.12) for $u_{J,j}$ by $RT^{1,(J,j)}$ and summing over the index (J,j), we rederive $L = 0$, since $\lambda_1 = 0$.

IV. ANALYSIS OF THE SEPARATION EQUATIONS

We do not as yet have a complete solution of the integrability conditions characterizing regular separation for the linear equation $L = E$

$$\begin{aligned} &A_{IJ} H_{(P,p)} = 0, \quad P \neq I,J, \\ &A_{IJ} H_{(J,j)} = B_{IJ} H_{(J,j-1)}, \quad H_{(J,0)} = 0, \quad I \neq J, \end{aligned} \tag{4.1}$$

where

$$A_{IJ} = \partial_{IJ} - \partial_J \ln H_I \partial_I - \partial_I \ln H_J \partial_J,$$

$$B_{IJ} = -\partial_I + \partial_I \ln H_J, \quad H_J = H_{(J,n_J)} \neq 0, \tag{4.2}$$

$$1 \leqslant J \leqslant N, \quad 1 \leqslant j \leqslant n_J, \quad \sum_J n_J = n.$$

In order that Theorem 1 can be applied to obtain a D-Stäckel matrix we must have all $H_{(J,j)} \neq 0$. However, we are assuming only that $H_J \neq 0$. Furthermore, it is easy to construct examples of separable systems where at least one of the $H_{(J,j)}$ vanishes.

A more detailed analysis of the structure of Eqs. (4.1) will resolve the difficulty. Suppose we are given N nonzero functions H_J satisfying $A_{IJ} H_P = 0$ for $P \neq I,J, I \neq J$. Our task will be to construct a finite set of functions $H_{(J,j)}$ with $H_{(J,n_J)} = H_J$ such that Eqs. (4.1) are satisfied. (We do not require that the $H_{(J,j)}$ are all nonzero.) Initially we will not know the values of the integers n_J.

The construction process is based on the second equation of (4.1), which we can write in the form

$$\partial_I [H_{(K,k-1)}/H_K] = (\partial_K \ln H_I \, \partial_I H_{(K,k)} + \partial_I \ln H_K \partial_K H_{(K,k)} - \partial_{IK} H_{(K,k)})/H_K, \quad I \neq K. \tag{4.3}$$

If $H_{(K,k)}$ is known then we can construct $H_{(K,k-1)}$ from (4.3) by quadrature.

Lemma 5: Suppose the N nonzero functions H_P satisfy $A_{IJ} H_P = 0$ for $P \neq I, J, I \neq J$ and suppose the function $H_{(K,k)}$ (fixed K,k) satisfies $A_{IJ} H_{(K,k)} = 0, K \neq I,J, I \neq J$. Then the $N-1$ equations (4.3) are compatible and have the general solution

$$H_{(K,k-1)} = \widetilde{H}_{(K,k-1)} + f^{(k-1)}(x_K) H_K, \tag{4.4}$$

where $\widetilde{H}_{(K,k-1)}$ is a particular solution and $f^{(k-1)}$ is an arbitrary function of x_K. The solution satisfies

$$A_{IJ} H_{(K,k-1)} = 0, \quad K \neq I,J, \quad I \neq J. \tag{4.5}$$

Proof: The compatibility requirement $\partial_J(\partial_I(H_{(K,k-1)}/H_K)) = \partial_I(\partial_J(H_{(K,k-1)}/H_K))$, $I,J \neq K$ and (4.5) are straightforward consequences of (4.3) and the conditions $A_{IJ} H_P = 0$, $A_{IJ} H_{(K,k)} = 0$.

It follows from Lemma 5 that for each K we can always construct functions $H_{(K,k-1)}$ through a step-by-step procedure using the second of Eqs. (4.1), such that the first of Eqs. (4.1) is automatically satisfied. At each step the solution $H_{(K,k-1)}$ is arbitrary up to the additive term $f^{(k-1)}(x_K) H_K$ and we simply choose one of these solutions. Thus we generate an infinite sequence $\{H_{(K,k)} = H_K^{(l)}\}$, $l = 0,1,2,...$, where $n_{K-l} = k$ (but n_K is unknown)

$$A_{IK} H_K^{(l)} = B_{IK} H_K^{(l+1)}, \quad I \neq K, \quad H_K = H_K^{(0)}. \tag{4.6}$$

The following properties of the operators A_{IK}, B_{IK} will prove useful:

$$\begin{aligned} &B_{IK} F(x) = 0, \text{ for all } I \neq K \\ &\quad \Leftrightarrow F(x) = f(x_K) H_K, \end{aligned} \tag{4.7}$$

$$A_{IK}(f(x_K) H_K^{(l)}) = B_{IK}(f H_K^{(l+1)} - f' H_K^{(l)}), \tag{4.8}$$

where $f' = \partial_K f$.

Suppose there is a smallest finite positive integer m_K for which functions $f_{(i)}(x_K)$ exist such that

$$H_K^{(m_K)} = \sum_{i=0}^{m_K - 1} f_{(i)}(x_K) H_K^{(i)}. \tag{4.9}$$

Lemma 6: Each $H_K^{(m_K + s)}$, $s = 0,1,2,...$, is a linear condition of the finite set $\{H_K^{(l)}: l = 0,...,m_K - 1\}$ with coefficients that are functions of x_K.

Proof: The proof is by induction on s. The statement is clearly true for $s = 0$. We assume it holds for $s = t$

$$H_K^{(m_K + t)} = \sum_{i=0}^{m_K - 1} g_{(i)}(x_K) H_K^{(i)}.$$

Now

$$B_{IK} H_K^{(m_K + t - 1)} = A_{IK} H_K^{(m_K + t)} = A_{IK}\left(\sum_0^{m_K - 1} g_{(i)} H_K^{(i)}\right)$$

$$= B_{IK}\left(\sum_0^{m_K - 1} g_{(i)} H_K^{(i+1)} - \sum_0^{m_K - 1} g'_{(i)} H_K^{(i)}\right).$$

Hence, by (4.7) there is a function $g(x_K)$ such that

$$H_K^{(m_K + t + 1)} = \sum_{i=0}^{m_K - 1} h_{(i)}(x_K) H_K^{(i)},$$

$$h_{(i)}(x_K) = \begin{cases} g_{(i-1)} - g'_{(i)}, & 1 \leqslant i \leqslant m_K - 1, \\ g - g'_{(0)}, & i = 0. \end{cases}$$ Q.E.D.

Let $\{\mathscr{H}_K^{(l)}\}$, $\{\mathscr{k}_K^{(l)}\}$, $l = 0,1,2,...$, be two sequences constructed by the procedure (4.6).

Lemma 7: There is a sequence of functions $g_1(x_K), g_2(x_K),...$, and expressions $L_{i,j}(g_1, g_2, ..., g_{i-j-1})$ with $L_{i0} = 0$, $L_{i,i-1} = 0$, and $L_{i+1,j} = L_{i,j-1} + g'_{i-j} - L'_{i,j}$ such that

$$\mathscr{H}_K^{(i)} = \mathscr{k}_K^{(i)} + \sum_{j=0}^{i-1} (g_{i-j}(x_K) - L_{i,j}(x_K)) \mathscr{k}_K^{(j)},$$

$$i = 0,1,2,... . \tag{4.10}$$

Any such sequence $\{g_l(x_K)\}$ together with $\{\mathscr{k}_K^{(l)}\}$ determines a new sequence of solutions $\{\mathscr{H}_K^{(i)}\}$ of (4.6). The induction proof of this result is similar to that of the preceding lemma.

Now let $\{H_K^{(l)}\}$ be the solution sequence treated in Lemma 6 and consider the relation (4.9). Set $\hbar_K^{(i)} = H_K^{(i)}$ in (4.10) and choose $g_1,...,g_{m_K - 1}$ recursively such that

$$-f_{(j)} = g_{m_K - j} - L_{m_K, j}, \quad j = 0,1,...,m_K - 1.$$

Then $\mathscr{H}_K^{(m_K)} = 0$. We see that there is a solution sequence $\{\mathscr{H}_K^{(l)}\}$ with $\mathscr{H}_K^{(0)},...,\mathscr{H}_K^{(m_K - 1)}$ nonzero and all further terms zero. According to Lemma 7 all other solution sequences are linear combinations of these m_K nonzero terms.

Lemma 8: The integer m_K, if it exists, is unique.

In general, there is no finite integer m_K for which (4.9) holds. [An example is $N = 2$, $H_1 = 1$, $H_2 = \exp(x_1 x_2)$. Here $m_1 = 1$, but m_2 does not exist.] However, if the $H_{(J,j)}$ satisfy equations (4.1), i.e., if they correspond to a regular separable system for the equation $L = E$ then the integers m_J always exist and $1 \leqslant m_J \leqslant n_J$. Thus there is a set of $\Sigma_{J=1}^N m_J$ functions $\{\hbar_{(J,j)}\}$, $1 \leqslant j \leqslant m_J$ satisfying (4.1) such that $H_J = \hbar_{(J,m_J)}$ and each $\hbar_{(J,j)}$ is nonzero. Using Lemma 7 we can express the equation $L = E$ in terms of the new functions $\hbar_{(J,j)}$.

Lemma 9:

$$L = \sum_{K=1}^N \sum_{k=1}^{n_K} H_{(K,k)} u_{K,k} = \sum_{K=1}^N \sum_{k=1}^{m_K} \hbar_{(K,k)} U_{K,k},$$

where

$$U_{K,k}(u_{k,l}, x_K) = u_{K, n_K - m_K + k} + \sum_{s=1}^{n_K - m_K + k - 1} (g_{n_K - m_K - s + k}(x_K) - L_{n_K - s, m_K - k}(x_K)) u_{K,s}.$$

In particular,

$$\partial_K U_{K,k} = U_{K,k+1}, \quad 1 \leqslant k \leqslant m_K - 1.$$

It follows from this result and Theorem 1 that when $L = E$ is separable then there exists a set of $m = \Sigma_{J=1}^N m_J$ nonzero functions $\hbar_{(J,j)}$ and an associated $m \times m$ D-Stäckel matrix S such that $\hbar_{(J,j)} = T^{1,(J,j)}$, where $T = S^{-1}$ and the separation equations for $L = E$ take the form

$$U_{K,k} + \sum_{l=1}^m S_{(K,k),l}(x_K)\lambda_l = 0, \quad K = 1,...,N, \tag{4.11}$$
$$1 \leqslant K \leqslant N, \quad 1 \leqslant k \leqslant m_K \leqslant n_K, \quad \lambda_1 = -E.$$

There are m separation parameters λ_l. The separable solutions u are determined by solving the N ordinary differential equations

$$U_{K,1} + \sum_{l=1}^m S_{(K,1),l}(x_K)\lambda_l = 0. \tag{4.12}$$

The remaining $m - N$ equations (4.11) are redundant, since they are obtained by differentiating the basic set (4.12). The highest derivative term in $U_{K,1}$ is $u_{K, n_K - m_K + 1}$ so each equation (4.12) is of order $n_K - m_K + 1$. The number of parameters in the solution u is $m + \Sigma_K(n_K - m_K) + 1 = n + 1$. We now have the complete solution of the separation of Eqs. (2.10).

V. SEPARATION OF LAPLACE EQUATIONS

Suppose Δ_N is the Laplace–Beltrami operator on a local pseudo-Riemannian manifold V^N. In local orthogonal coordinates x_I we have

$$\Delta_N = \frac{1}{h} \sum_{I=1}^N \partial_I (h H_I \partial_I), \tag{5.1}$$

where

$$ds^2 = \sum_{I=1}^N H_I^{-1} dx_I^2, \quad h^2 = \prod_I H_I^{-1}.$$

It is of interest to determine the relationships between the well-developed theory of multiplicative R separation for the Laplace equation $\Delta_N u = 0$ (Refs. 7–9) and additive separation. Recall that multiplicative R separation in the orthogonal coordinates x_I leads to solutions for the Laplace equation of the form

$$u = e^{R(x)} \prod_{I=1}^N u^{(I)}(x_I), \tag{5.2}$$

where the fixed function R is independent of the separation parameters. Similarly we can introduce additive R separation

$$u = e^{R(x)} \left(\sum_{I=1}^N u^{(I)}(x_I) \right). \tag{5.3}$$

The following is a straightforward consequence of the principal results of this paper.

Theorem 2: If the Laplace equation $\Delta_N U = 0$ is multiplicatively R separable in the orthogonal coordinates x_I then it is additively R separable in these same coordinates if and only if $e^{-R} \Delta_N e^R = c H_I$, $I = 1,...,N$, where c is a constant.

In each of these cases a true Stäckel matrix determines the separation; no nontrivial D-Stäckel matrices appear. Note that true multiplicative separation ($R \equiv 1$) always leads to additive separation.

On the other hand, Laplace equations may admit additive separation in an orthogonal coordinate system for which no multiplicative R separation is possible. For example, consider the three-dimensional manifold with metric coefficients

$$H_1 = H_2 = (x_1 + x_2)^5, \quad H_3 = (x_1 + x_2)^4. \tag{5.4}$$

Then

$$\Delta_3 = (x_1 + x_2)^4 [(x_1 + x_2)(\partial_{11} + \partial_{22}) - 2(\partial_1 + \partial_2) + \partial_{33}], \tag{5.5}$$

and since (5.4) is not conformal to a Stäckel form metric, no multiplicative R separation is possible. However, since

$$(x_1 + x_4)^{-4} \Delta_3 u = (x_1 + x_2)(u_{1,2} + u_{2,2}) - 2(u_{1,1} + u_{2,1}) + u_{3,2} = 0, \tag{5.6}$$

we have

$$H_1' = H_2' = x_1 + x_2, \quad H_3' = 1,$$

$$H'_{(1,1)} = H'_{(2,1)} = -2,$$

which satisfies Eqs. (2.10) [or (2.14)] with $n_1 = n_2 = 2$, $n_3 = 1$. Thus the Laplace equation admits additive separation in these coordinates corresponding to a 5×5 D-Stäckel matrix.

ACKNOWLEDGMENT

This work has been partially supported by National Science Foundation Grant No. MCS 82-19847.

[1]E. G. Kalnins and W. Miller, "Intrinsic characterization of variable separation for the partial differential equations of mechanics," in *Proceedings of IUTAM-ISIMM Symposium on Modern Development in Analytical Mechanics*, Turin, 1982 (Atti della Academia delle Scienze di Torino, Turin, 1983), Supplement to Vol. 117, pp. 511–533.
[2]P. Stäckel, *Über die Integration der Hamilton-Jacobischen Differentialgleichung Mittels Separation der Variabelen* (Habitilatationsschrift, Halle, 1891).
[3]L. P. Eisenhart, Ann. Math. **35**, 284 (1934).
[4]T. Levi-Civita, Math. Ann. **59**, 383 (1904).
[5]T. W. Koornwinder, "A precise definition of separation of variables," in *Lecture Notes in Mathematics*, Vol. 810 (Springer, Berlin, 1980).
[6]L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, 1949), Appendix 13.
[7]P. Moon and D. E. Spencer, J. Franklin Inst. **253**, 585 (1952).
[8]V. N. Shapovalov, Differential Equations **16**, 1864 (1980); **10**, 1212 (1981).
[9]E. G. Kalnins and W. Miller, Trans. Am. Math. Soc. **244**, 241 (1978).

Multiple scales analysis of a nonlinear ordinary differential equation

Xu Gang and Shu Changqing
Institute of Physics, Chinese Academy of Sciences, Beijing, People's Republic of China

Lin Lei
Department of Physics, City College, City University of New York, New York 10031,[a] and Department of Physics, Queensborough Community College, City University of New York, Bayside, New York 11364, and Institute of Physics, Chinese Academy of Sciences, Beijing, People's Republic of China

(Received 27 August 1984; accepted for publication 21 December 1984)

Asymptotic solutions of the nonlinear ordinary differential equation $d^2\theta/dZ^2 + a\,d\theta/dZ + f(\theta) = 0$ for large a are obtained by the singular perturbation method of multiple scales analysis. They are in the form of $\theta(Z) = A(Z/a) + B(Z/a)\exp(-aZ)$. Initial and boundary value problems are discussed. The special case of $f(\theta) = \gamma + \cos 2\theta$ ($\gamma < 1$), encountered in shearing nematic liquid crystal soliton problems and other physical systems, is solved in detail. Previously obtained analytic solutions are recovered and justified. Our results are applicable to the unsteady shearing nematic problem.

I. INTRODUCTION

The nonlinear ordinary differential equation

$$\frac{d^2\theta}{dZ^2} + a\frac{d\theta}{dZ} + f(\theta) = 0, \tag{1.1}$$

where a is a constant and $f(\theta)$ is a nonlinear function of $\theta(Z)$, describes the damped nonlinear motion of a single particle. It appears frequently in mechanics and many physical problems. For example, in the discussion of solitons in steady shearing nematic liquid crystals such an equation was derived by Lin *et al.*[1] with

$$f(\theta) = \gamma + \cos 2\theta, \quad \gamma < 1. \tag{1.2}$$

Equation (1.1) is also studied in the traveling wave solutions of biological heredity models and chemical reaction-diffusion equations.[2]

In the specific problem studied in Refs. 1 and 3 (in which a is represented by η) two approximate analytic solutions, the so-called A and B solitons, are derived under the condition $a \gg 1$. In each of these solutions[1,3] there is only one constant of integration. However, since (1.1) is an ordinary differential equation of second order, two constants of integration in the general solution are expected (in order to satisfy arbitrary given initial or boundary conditions). This paper is partly motivated by the desire to clarify this issue. It turns out that the solutions given in Refs. 1 and 3 are special solutions (being solitons) of (1.1) and (1.2) and are fully justified (see Sec. IV).

Also, (1.1) appears in our study of unsteady shearing nematics.[4] An understanding of the asymptotic solutions of (1.1) is thus needed.

Equation (1.1) may be rewritten as

$$\epsilon\ddot{\theta} + \dot{\theta} + f(\theta) = 0, \tag{1.3}$$

where

$$\epsilon \equiv a^{-2}, \quad \tau \equiv \frac{Z}{a}, \quad \dot{\theta} \equiv \frac{d\theta}{d\tau}. \tag{1.4}$$

Alternatively, (1.3) is equivalent to

$$\dot{\theta} = z, \quad \epsilon\dot{z} = -z - f(\theta). \tag{1.5}$$

[a] Correspondence address.

Equation (1.5), and hence (1.1) or (1.3), is obviously a special case of the more general nonlinear ordinary differential equation

$$\dot{\theta} = g(\theta,z), \quad \epsilon\dot{z} = h(\theta,z). \tag{1.6}$$

The initial value problem of (1.6) has been discussed by O'Malley.[5] Under specific assumptions, asymptotic solutions have been constructed. There are also discussions on the properties of solutions of (1.6) under certain conditions.[6–8]

In this paper, multiple scales analysis[9] is applied to (1.1). An asymptotic solution for $a \to \infty$ ($\epsilon \to 0$) is constructed. Boundary and initial value problems are discussed, respectively (Sec. II). The general solutions obtained in Sec. II are specified to the case of (1.2) in Sec. III, in which the analytic solutions of A and B solitons of Ref. 3 are recovered. In Sec. IV, the mechanical interpretation of our solutions is presented. The expansion method used in Refs. 1 and 3 is discussed and justified.

II. PERTURBATIONAL EXPANSION

In (1.3) $f(\theta)$ is assumed to be differentiable an infinite number of times. In this section, the asymptotic expansion of the solution of (1.3) with $\epsilon \to 0$ will be derived using the singular perturbation method of multiple scales analysis.[9] Let

$$u \equiv \tau, \quad v \equiv g(\tau)/\epsilon, \tag{2.1}$$

where the functional form of $g(\tau)$ remains to be specified [see (2.18)]. For our purpose,

$$g'(\tau) \equiv \frac{dg}{d\tau} > 0 \tag{2.2}$$

is assumed. We then assume that $\theta(\tau) = \theta(u,v)$ can be expanded in powers of ϵ (when $\epsilon \to 0$):

$$\theta = \theta^{(0)} + \epsilon\theta^{(1)} + \epsilon^2\theta^{(2)} + \cdots, \tag{2.3}$$

where $\theta^{(0)}$, $\theta^{(1)}$,... are functions of τ through their dependence on $u(\tau)$ and $v(\tau)$.

Let

$$g' \equiv \frac{dg}{d\tau}, \quad g'' \equiv \frac{d^2g}{d\tau^2}, \quad \partial_u \equiv \frac{\partial}{\partial u}, \quad \partial_v \equiv \frac{\partial}{\partial v}. \tag{2.4}$$

 0022-2488/85/071566-04$02.50

By (2.1),

$$\dot{\theta} \equiv \frac{d\theta}{d\tau} = (\partial_u + \epsilon^{-1} g' \partial_v)\theta,$$
$$\ddot{\theta} \equiv \frac{d^2\theta}{d\tau^2} = [\partial_{uu} + \epsilon^{-2} g'^2 \partial_{vv}(2.5) + \epsilon^{-1}(2g' \partial_u + g'')\partial_v]\theta. \tag{2.5}$$

When $f(\theta)$ is expanded in a Taylor series, one obtains

$$f(\theta) = f(\theta^{(0)}) + f'(\theta^{(0)})\theta^{(1)}\epsilon + [f'(\theta^{(0)})\theta^{(2)} + \tfrac{1}{2}f''(\theta^{(0)})]\epsilon^2 + \cdots. \tag{2.6}$$

When (2.5) and (2.6) are substituted into (1.3) and the coefficient of each order of ϵ is equated to zero we obtain

$$g'(g' \partial_{vv} + \partial_v)\theta^{(n)} = C_n, \quad n = 0,1,2,\ldots, \tag{2.7}$$

with

$$\begin{aligned}
C_0 &= 0,\\
C_1 &= -(2g' \partial_{uv} + g'' \partial_v + \partial_u)\theta^{(0)} - f(\theta^{(0)}),\\
C_2 &= -(2g' \partial_{uv} + g'' \partial_v + \partial_u)\theta^{(1)} - \partial_{uu}\theta^{(0)} - f'(\theta^{(0)})\theta^{(1)},
\end{aligned} \tag{2.8}$$

$$\begin{aligned}
C_3 &= -(2g' \partial_{uv} + g''\partial_v + \partial_u)\theta^{(2)} - \partial_{uu}\theta^{(1)} - f'(\theta^{(0)})\theta^{(2)}\\
&\quad - \tfrac{1}{2}f''(\theta^{(0)})(\theta^{(1)})^2,\\
&\vdots\ .
\end{aligned}$$

Equation (2.7) is a second-order linear ordinary differential equation of $\theta^{(n)}$ with respect to v. For $n = 0$, (2.7) can be integrated giving

$$\theta^{(0)} = A_0(u) + B_0(u)\exp(-v/g'). \tag{2.9}$$

Putting (2.9) into (2.8) one has

$$C_1 = -\{[A_0' + f(\theta^{(0)})] + e^{-v/g'}[-(B_0' - (g''/g')B_0) - (g''B_0/(g')^2)v]\}, \tag{2.10}$$

where

$$A_0' \equiv \frac{dA_0}{du}, \quad B_0' \equiv \frac{dB_0}{du}. \tag{2.11}$$

Substituting $\theta^{(0)}$ of (2.9) into $f(\theta^{(0)})$ and expanding with respect to A_0 we have

$$f(\theta^{(0)}) = f(A_0) + f'(A_0)B_0 \exp(-v/g') + \cdots. \tag{2.12}$$

Putting (2.10) into (2.7) and using (2.12), Eq. (2.7) may be integrated for $n = 1$ to give

$$\theta^{(1)} = A_1(u) + B_1(u)\exp(-v/g') + \mathrm{ST} + \mathrm{NST}, \tag{2.13}$$

where the term NST comes from those in C_1 containing the factor $\exp(-nv/g')$ $(n \geqslant 2)$ and ST from C_1 containing the factor $\exp(-nv/g')$ with $n = 0$ and 1.

It can be shown easily that the ST terms violate the following condition for the expansion of (2.3) to hold, i.e.,

$$\lim_{v\to\infty} |\theta^{(n+1)}/\theta^{(n)}| < \infty, \quad n = 0,1,2,\ldots. \tag{2.14}$$

Consequently, ST are secular terms. To eliminate these secular terms and to determine the functions $A_0(u)$ and $B_0(u)$ in (2.9) we set these secular terms to zero resulting in

$$A_0' + f(A_0) = 0, \tag{2.15}$$
$$B_0' - (g''/g')B_0 - f'(A_0)B_0 = 0, \tag{2.16}$$
$$g''B_0 = 0. \tag{2.17}$$

In general $B_0 \neq 0$. By (2.17) one has

$$g''(u) = 0 \quad \text{or} \quad g' = \text{const.} \tag{2.18a}$$

Without loss of generality one may choose

$$g' = 1. \tag{2.18b}$$

Integrating (2.15) and (2.16) with the use of (2.18) we have

$$u - u_0 + \int \frac{dA_0}{f(A_0)} = 0, \tag{2.19}$$
$$B_0 = b_0/f(A_0), \tag{2.20}$$

where u_0 and b_0 are constants (independent of u and v).

By (2.8), (2.9), (2.13), and (2.18b) one has

$$\begin{aligned}
C_3 = &-[A_1' + A_0'' + f'(A_0)A_1]\\
&+ \exp(-v)[B_1' - B_0''\\
&- f'(A_0)B_1 - f''(A_0)A_1B_0] + \mathrm{NST}.
\end{aligned} \tag{2.21}$$

Similarly, the elimination of the secular terms gives rise to

$$A_1' + A_0'' + f'(A_0)A_1 = 0, \tag{2.22}$$
$$B_1' - B_0'' - f'(A_0)B_1 - f''(A_0)A_1B_0 = 0. \tag{2.23}$$

By integration,

$$A_1 = [a_1 + \ln|f(A_0)|]f(A_0), \tag{2.24}$$
$$B_1 = -B_0f'(A_0)[a_1 - 1 + \ln|f(A_0)|] + b_1/f(A_0), \tag{2.25}$$

where A_1, B_1, and A_0 are functions of u and a_1 and b_1 are constants of integration.

Correction terms of higher order can be calculated similarly. Therefore, the asymptotic expansion of the solution of (1.1) when $a \to \infty$ assumes the following form:

$$\theta = [A_0(u) + \epsilon A_1(u) + \cdots] + \exp(-v)[B_0(u) + \epsilon B_1(u) + \cdots] + \mathrm{NST}, \tag{2.26}$$

where $A_0(u)$ is given by (2.19) and A_1, B_0, B_1 are expressed as functions of A_0 through (2.24), (2.20), and (2.25), respectively, and $\epsilon = a^{-2}$

$$u = \tau = Z/a, \quad v = \tau/\epsilon = aZ. \tag{2.27}$$

The NST are nonsecular terms. They are the sum of terms containing the factor $\exp(-nv)$ $(n \geqslant 2)$ and do not involve any constants of integration. Moreover, they are of order ϵ and decay faster than the rest of the terms in (2.26) when $v \to \infty$.

For (1.1), let us assume that initial conditions

$$\theta(Z = 0) = D(\epsilon) \tag{2.28a}$$

and

$$\left.\frac{d\theta}{dZ}\right|_{Z=0} = \frac{E(\epsilon)}{a} \tag{2.28b}$$

are given such that $D(\epsilon)$ and $E(\epsilon)$ can be expanded in powers of ϵ

$$D(\epsilon) = \sum_{j=0}^{\infty} D_j\epsilon^j, \quad E(\epsilon) = \sum_{j=0}^{\infty} E_j\epsilon^j. \tag{2.29}$$

Note that

$$\frac{d\theta^{(0)}}{dZ} = \frac{1}{a}\left[A_0'\left(\frac{Z}{a}\right) + B_0'\left(\frac{Z}{a}\right)e^{-aZ} + (-a^2)B_0\left(\frac{Z}{a}\right)e^{-aZ}\right]. \tag{2.30}$$

The last term in (2.30) cannot satisfy the initial conditions of (2.28) and hence in (2.20) one should take

$$b_0 = 0, \tag{2.31}$$

resulting in

$$B_0(Z/a) = 0. \tag{2.32}$$

Therefore, the zeroth-order initial conditions should be satisfied by requiring

$$\theta^{(0)}(Z=0) = A_0(0) = D_0 \tag{2.33a}$$

and

$$\left.\frac{d\theta^{(0)}}{dZ}\right|_{Z=0} = E_0 + B_1(0) = A_0'(0). \tag{2.33b}$$

From (2.33a) one may determine the constant of integration $Z_0 \equiv u_0 a$. The constant b_1 in B_1 is determined by (2.33b). The initial conditions for orders of $j \geqslant 1$ are satisfied by taking

$$\theta^{(j)}(Z=0) = A_j(0) + B_j(0) = D_j, \tag{2.34a}$$

$$\left.\frac{d\theta^{(j)}}{dZ}\right|_{Z=0} = A_j'(0) + B_j'(0) = E_j - B_{j-1}. \tag{2.34b}$$

From these equations the different constants of integration in (2.3) may be determined. In the above calculations, for simplicity, the NST terms for $j \geqslant 1$ have not been included. These terms contain the factor $\exp(-jaZ)$ but no constants of integration. Although in principle these terms should be included, their contributions in establishing the constants a_i and b_i $(i \geqslant 1)$ are very small for j large.

If instead of the initial conditions of (2.28), the boundary conditions

$$\theta(Z=0) = F(\epsilon) \tag{2.35a}$$

and

$$\theta(Z=a) = G(\epsilon) \tag{2.35b}$$

are given for (1.1) such that

$$F(\epsilon) = \sum_{j=0}^{\infty} F_j \epsilon^j, \quad G(\epsilon) = \sum_{j=0}^{\infty} G_j \epsilon^j, \tag{2.36}$$

then

$$\theta^{(n)}|_{Z=0} = A_n(0) + B_n(0) = F_n, \tag{2.37a}$$

$$\theta^{(n)}|_{Z=a} = A_n(1) + B_n(1)\exp(-a^2)$$
$$\simeq A_n(1) = G_n. \tag{2.37b}$$

From (2.37) the constants of integration Z_0, a_i, and b_i $(i \geqslant 1)$ can then be determined.

III. SPECIAL CASE OF $f(\theta) = \gamma + \cos 2\theta$

The asymptotic solution of (1.1) obtained in Sec. II is expressed in terms of A_1, B_0, B_1,..., which are in turn expressed through the function $A_0 = A_0(u - u_0)$. In the special case of (1.2) we have

$$B_0(u) = b_0/(\gamma + \cos 2A_0), \tag{3.1}$$

$$A_1(u) = (\gamma + \cos 2A_0)(a_1 + \ln|\gamma + \cos 2A_0|), \tag{3.2}$$

$$B_1(u) = b_1/(\gamma + \cos 2A_0)$$
$$+ B_0(u) 2 \sin 2A_0 (a_1 - 1 + \ln|\gamma + \cos 2A_0|). \tag{3.3}$$

To find A_0 we put (1.2) into (2.19).

(a) For $\gamma + \cos 2A_0 > 0$ and $-\theta_0 \leqslant A_0 \leqslant \theta_0$, we obtain

$$A_0 = -\tan^{-1}\{W \tanh[(1-\gamma^2)^{1/2}(u - u_0)]\}, \tag{3.4}$$

where

$$W \equiv [(1+\gamma)/(1-\gamma)]^{1/2}, \quad \theta_0 \equiv \tan^{-1} W. \tag{3.5}$$

(b) For $\gamma + \cos 2A_0 < 0$ and $\theta_0 \leqslant A_0 \leqslant \pi - \theta_0$, we obtain

$$A_0 = -\cot^{-1}\{W^{-1} \tanh[(1-\gamma^2)^{1/2}(u - u_0)]\}. \tag{3.6}$$

The specific forms of A_1, B_0, B_1, ..., if needed (see Sec. IV), may be obtained from the formulas of Sec. II.

IV. DISCUSSION

From Sec. II the asymptotic solution of (1.1) may be written as

$$\theta = A(u,\epsilon) + B(u,\epsilon)\exp(-v), \tag{4.1}$$

where

$$A(u,\epsilon) = A_0(u) + \epsilon A_1(u) + \cdots, \tag{4.2}$$

$$B(u,\epsilon) = B_0(u) + \epsilon B_1(u) + \cdots, \tag{4.3}$$

and $A(u,\epsilon)$ is the so-called outer expansion.[5,9] It formally satisfies (1.1). However, $A(u,\epsilon)$ contains only one constant of integration of zeroth order while (1.1), being a differential equation of second order, requires two constants of integration (of zeroth order) to satisfy a general initial or boundary condition (see Sec. II). That is why the second term $B(u,\epsilon)\exp(-v)$ is important. It is called the boundary layer correction.[5] Note that because of the exponential factor this second term tends to zero when $\epsilon \to 0$ $(a \to \infty)$ for any finite Z (remembering $v = aZ$). In other words, in a mechanical analog, this term represents the rapidly decaying motion with respect to the "time" Z and is significant in the short "time" period $(0, 1/a)$ as far as the solution for θ is concerned. In contrast, the outer expansion $A(u,\epsilon)$ represents the long-time behavior (Z large) of θ.

Note that the second term in (4.1) diverges at $v \to -\infty$. For a solution which is finite for $v \to \pm\infty$ (or $Z \to \pm\infty$), as is the case of the soliton solutions considered in Refs. 1 and 3, one must take $B(u,\epsilon) = 0$. This is achieved by taking $b_0 = b_1 = \cdots = 0$. One then has

$$\theta = A(u,\epsilon) = A_0(u) + \epsilon A_1(u) + \cdots. \tag{4.4}$$

In Refs. 1 and 3, the A- and B-soliton solutions are obtained by a straightforward expansion of (1.1) or (1.3). By comparing (3.4) and (3.6) with the results of Refs. 1 and 3, we see that the zeroth-order solution obtained in the latter [Eq. (4) of Ref. 1, or Eqs. (15) and (18) of Ref. 3][10] are exactly the $A_0(u)$ $(u = Z/a)$ obtained here. It is also easily shown that to order ϵ, our result of (4.4) agrees with those in Eqs. (13) and (16) of Ref. 3.[10] In contrast, the multiple scales analysis employed in this paper enables one to obtain the solution (soliton solutions in particular) of (1.1) in a systematic way and to any order of ϵ desired.

Application of our results in Sec. II to the unsteady shearing nematic problem will be published elsewhere.[4]

[1]L. Lin, C. Q. Shu, J. L. Shen, P. M. Lam, and Y. Huang, Phys. Rev. Lett. **49**, 1335 (1982); **52**, 2190(E) (1984).
[2]P. C. Fife, *Mathematical Aspects of Reacting-Diffusing Systems* (Springer, New York, 1979).

[3]L. Lin and C. Q. Shu, Acta Phys. Sin. **33**, 165 (1984) (English translation to appear in Chin. Phys.).
[4]G. Xu, C. Q. Shu, and L. Lin (to be published).
[5]R. E. O'Malley, *Introduction to Singular Perturbations* (Academic, New York, 1974).
[6]F. A. Howes, J. Differential Eqs. **30**, 192 (1978).
[7]W. R. Wasow, Commun. Pure Appl. Math. **9**, 93 (1956).
[8]M. L. Vishik and L. A. Lyusternik, Usp. Mat. Nauk. **12**, 3 (1957) [Am. Math Soc. Transl. **20**, 239 (1961)].
[9]A. H. Nayfel, *Perturbation Methods* (Wiley, New York, 1973).
[10]In Refs. 1 and 3, for convenience, the centers of the solitons have been assumed to be at $Z = 0$. This corresponds to setting $u_0 = 0$. One only has to replace Z by $Z - Z_0$ in the results of Refs. 1 and 3 to obtain the general case of $u_0 \neq 0$.

Generalized moments applied to Fermi-type functions

M. H. Wendel and E. R. Hilf[a)]
Institut für Kernphysik, T. H. Darmstadt, D-6100 Darmstadt, Schlossgartenstrasse 9, Federal Republic of Germany

(Received 8 November 1983; accepted for publication 30 November 1984)

Generalized integral moments are defined for a general class of saturating functions $[f(0) = \rho_0, f(\infty) = 0]$. They are useful as independent variables for describing surface properties or macroscopic dynamics of finite systems. Applied especially to functions of the Fermi type, analytic solutions are given in terms of a semiconverging, and of a numerically semiunstable expansion, respectively, suitable for numerical evaluation. Results are compared to the semidivergent expansion as given by Åberg, of which some properties are exhibited here, and with the exact numerical solutions known for this special example.

I. MOMENTS OF SATURATING FUNCTIONS

Saturating distributions $\left(\lim_{x\to 0} f(x) = \rho_0, \lim_{x\to\infty} f(x) = 0\right)$ occur in physics in many different fields, such as the density of electrons in a piece of metal, of nucleons in a nucleus, and the distribution of thermal kinetic energies of particles in a quantum gas. We restrict ourselves here to one-dimensional problems.

To describe their macroscopic properties as well as surface properties one needs independent global dynamic variables, which may be defined by some measures of the saturating distribution.

Especially for the subclass of functions with known saturation value $\rho(0)$, Süssmann[1] defined the "surficial moments" to use as macroscopic variables, which were suggestively named as "surface-position," "-thickness," "flair," etc.

A different approach was given by Myers and Swiatecki[2] for the subclass of functions that can be written as a folding of a step function with a function localized to the surface.

A more general way was described by Ford and Wills,[3] defining moments by the integrals

$$(R_k)^k := [(k+3)/3]\langle r^{k+2}\rangle/\langle r^2\rangle, \tag{1}$$

with $\langle r^k\rangle := \int dr\, r^k \rho(r)$.

Much experimental information, such as form factors of scattering experiments to probe the properties of the saturating function, comes in terms of these Ford–Wills moments. Thus there is no direct experimental information on $\rho(0)$. Macroscopic variables, suitable to describe macroscopic dynamics as well as being good for physical intuition, have yet to be defined. They should be calculable using experimentally accessible information only such as the observed form factors. Thus we propose here to start with a series of surface moments directly derived from the Ford–Wills moments,

$$S_\mu^{(k)} := (\partial_k)^\mu R_k, \tag{2}$$

which might be approximated by difference expressions for integer k.

To gain quantities best for physical intuition we aim finally at definitions using the $S_\mu^{(k)}$ only but such that [for assumed to be known $\rho(0)$] they approach the surficial moments of Süssman—at least for sufficiently leptodermous distributions—and thus serve as their generalization. For this purpose we study the special example of a Fermi-type function allowing for an asymmetric falloff of $f(x)$ at the surface,

$$\rho(x) = \rho(0)\left(f(x) + \alpha \frac{d}{dx_1}(f(x_1))\bigg|_{x_1 = x}\right). \tag{3}$$

For $f(x) = (1 + \exp((x-c)/a))^{-1}$ it has been shown[4] that the surface position c, the surface thickness a, and the flair α can be inferred from the surface moments S_μ for a sufficiently thin surface to first order in $(1/k)$ by

$$s_0 := S_0 = c,\quad s_1 = (6/\pi)S_1/S_0 = (a/c)^2,$$
$$s_2 = -\tfrac{1}{2}S_2/S_1 = \alpha/c. \tag{4}$$

Generally, the $s_\mu := S_\mu/S_{\mu-1}$ are proposed as macroscopic variables.

We note that

$$s_\mu = \frac{S_\mu}{\prod_{\nu=1}^{\mu-1} s_\nu}. \tag{5}$$

They are surface moments of saturating functions of which only some Ford–Wills moments are assumed to be experimentally known. One may attach names in alliteration to the Süssmann moments, s_0 is the surface position, $\sqrt{s_1}$ is the relative surface thickness, and s_2 is the relative flair. The k dependence of the surface moments $s_\mu^{(k)}$ is a next-order effect and of interest if Ford–Wills moments for many different k are experimentally known.

This k dependence results in

$$s_0^{(k)} = S_0^{(k)} = R_k, \tag{6a}$$

$$\begin{aligned} s_1^{(k)} &= S_1^{(k)}/S_0^{(k)} \\ &= [\langle r^{k+2}\ln(r)\rangle - \ln((s_0^{(k)})^k)/k \\ &\quad + 1/(k+3)]/k, \end{aligned} \tag{6b}$$

$$\begin{aligned} s_2^{(k)} &= S_2^{(k)}/S_1^{(k)} \\ &= [\langle r^{k+2}\ln(r)\rangle(1 - 1/(k\cdot s_1^{(k)})) \\ &\quad + \langle r^{k+2}(\ln(r))^2\rangle/s_1^{(k)} \\ &\quad - \ln(s_0^{(k)})(1 - 2/(k\cdot s_1^{(k)}))/k - 1 + 1/(k+3) \\ &\quad - (2k+3)/(s_1^{(k)}\cdot(k^2+3k)^2)]/k. \end{aligned} \tag{6c}$$

In the remainder of the paper we shall study the gener-

[a)] Now at Fachbereich Physik, Universität Oldenburg, D-2900 Oldenburg, Carlv. Ossietskystrasse, Federal Republic of Germany.

 0022-2488/85/071570-06$02.50

alized surface moments S_μ for the special class of Fermi functions, to gain experience for the inverse problem of determining the bounds to the saturating function from the knowledge of a finite number of their moments. Mainly, this means to develop as a tool analytical solutions of the logarithmic moment occurring in Eqs. (6b) and (6c).

Different methods for the numerical solution of integrals of that kind—which have to be evaluated with respect to a parameter-dependent ansatz for $\rho(r)$—and their validity in different regions of the parameter space have already been discussed.[5]

In the following section we shall develop further an earlier—in this case—semidivergent[6] expansion by Åberg[7]; whereas in Sec. III we shall work out a semiconvergent series expansion as well as a semiunstable one, which seems to be more suitable for explicit analytical considerations.

The results of the semidivergent and the semiconvergent methods are presented and checked with the exact results for several sets of given parameters in Sec. IV.

We want to mention the paper of Gräf and Pabst,[8] who elegantly identified the moments of the Fermi–Dirac distribution with a contour integral and thus obtained a MacLaurin series expansion, suitable for numerical computation.

The mathematical tools used and developed here may be useful to other related problems as well; they are especially applicable to the problem of analytical evaluation of s_μ in general if the density is known, while for experimentally known form factors the generalized moments are given directly by Eqs. (1), (2), and (5).

II. THE ÅBERG METHOD

The method presented in Ref. 7 is defined by starting from an integration by parts

$$I_k = \int_0^\infty dr\, f(r)\rho^\nu(r) = [F(r)\rho^\nu(r)]_0^\infty + \int_0^\infty dr\, F(r)P(r), \tag{7}$$

where—for our purpose—we consider only the special case $\nu = 1$, $f(r) = r^k \ln(r)$, and $\rho(r)$ being the Fermi function $(1+\exp((r-c)/a))^{-1}$. Thus $P(r) = -d_r\,\rho(r)$.

Since the boundary term vanishes we first have to evaluate

$$F(r) = r^{k+1}/(k+1)(\ln(r) - 1/(k+1)) \quad \text{at } r = c. \tag{8}$$

Its Taylor expansion contains nth derivatives

$$d_r^n(r^k \ln r) = \sum_{i=0}^{n} \binom{n}{i} (d_r^i r^k)(d_r^{n-i} \ln r) = r^{k-n}\left\{ n! \sum_{i=0}^{n-1} \binom{k}{i} \frac{(-1)^{n-i+1}\Theta(k-i)}{n-i} + \frac{k!}{(k-n)!}(\ln r)\Theta(k-n) \right\}, \tag{9}$$

using the well-known step function Θ.

Therefore,

$$d_r^n F(r) = [(k+1)!/(k+1-n)!]\, r^{-n} F(r)\Theta(k+1-n) + \frac{n! + r^{k+1-n}}{k+1} \sum_{i=0}^{n-1} \binom{k+1}{i} \times \frac{(-1)^{n-i+1}}{n-i}\Theta(k+1-i), \tag{10}$$

and the series for $F(r)$ reads

$$\frac{c^{k+1}}{k+1} \sum_{m=0}^{\infty} f_m^{(k)} \left(\frac{r}{c} - 1\right)^m, \tag{11}$$

with

$$f_m^{(k)} = \left[(-1)^m \Theta_m^{(k)} + \binom{k+1}{m} \times \left(\ln c - \frac{1}{k+1} \right) \Theta(k+1-m) \right],$$

defining the shorthand

$$\Theta_m^{(k)} = \sum_{i=0}^{m-1} \binom{k+1}{i} \frac{(-1)^i \Theta(k+1-i)}{i-m}. \tag{12}$$

Substituting $s := (r-c)/a$ and $\alpha := -c/a$, we get

$$I_k(|\alpha|, c^{k+1}) = \frac{c^{k+1}}{k+1} \int_\alpha^\infty ds \sum_{m=0}^{\infty} f_m^{(k)} |\alpha|^{-m} \frac{s^m e^s}{(1+e^s)^2}. \tag{13}$$

In Ref. 7 the lower integration limit α was approximated by $-\infty$, then the sum and the integral were exchanged to ease the calculation.

However, in this special case, that last operation is a sensitive one because this series expansion of $F(r)$ is not absolutely convergent in the whole integration interval. Instead $F(r)$ is represented by a semidivergent series; a cutoff in the m summation is necessary. An analysis of the choice of this new parameter should yield an optimal value m_0 beyond which the series will start to diverge.

Thus

$$I_k(|\alpha|, c^{k+1}) = \frac{c^{k+1}}{k+1} \sum_{m=0}^{m_0} f_m^{(k)} |\alpha|^{-m} \times \int_{-\infty}^{\infty} ds \frac{s^m e^s}{(1+e^s)^2} + \sigma(m_0) \tag{14}$$

with a new term $\sigma(m_0)$ resulting both from the cutoff and also from the exchange of the lower integration limit.

The integral

$$J_m = \int_{-\infty}^{\infty} ds \frac{s^m e^s}{(1+e^s)^2}, \tag{15}$$

as already discussed by Åberg[7] as $J_{m,\nu}$ (here in the special case of $\nu = 1$ in his notation), is known to be related to the beta function, being the mth derivative of a function $g(\gamma)$ at $\gamma = 0$, where

$$g(\gamma) = \int_{-\infty}^{\infty} ds \frac{e^{(\gamma+1)s}}{(1+e^s)^2}; \tag{16}$$

$g(\gamma)$ is the integral representation of a special Euler integral of the first kind which can be evaluated to

$$g(\gamma) = \pi\gamma/\sin \pi\gamma. \tag{17}$$

Using its series expansion we finally arrive at

$$J_m = \left\{ d_\gamma^m \left[1 + \sum_{\mu=1}^{\infty} \frac{2(2^{2\mu-1}-1)|B_{2\mu}|}{(2\mu)!} (\pi\gamma)^{2\mu} \right] \right\}_{\gamma=0}, \tag{18}$$

with the Bernoulli numbers $B_{2\mu}$.

Obviously only the constant term of the mth derivative does not vanish as $\gamma \to 0$. Thus

$$J_m = \begin{cases} 1, & \text{for } m = 0, \\ 2(2^{m-1}-1)|B_m|\pi^m, & \text{for even } m, \\ 0 & \text{for odd } m. \end{cases} \tag{19}$$

The resulting final expression for the k th (logarithmic) moment I_k reads

$$I_k(|\alpha|,c^{k+1}) = \frac{c^{k+1}}{k+1} \left[2 \sum_{m'=1}^{M} f_{2m'}^{(k)} |\alpha|^{-2m'}(2^{2m'-1}-1) \times |B_{2m'}|\pi^{2m'} + \left(\ln c - \frac{1}{k+1} \right) \Theta(k+1) \right] + \sigma(M), \tag{20}$$

with $m' = m/2$.

III. A SEMICONVERGENT METHOD

In the last section we elaborated on the Åberg–Sommerfeld method, where, not knowing $\sigma(m)$, the optimal m_0 could be determined only empirically by comparison with the to-be-known exact result. Here we develop a different and, we think, profitable method which allows to estimate the integral with a negligible residual term. The methods used and developed here are useful also for analogous problems.

The k th logarithmic moment $I_k(|\alpha|,c)$, as given by (7) and (8), after substituting $z := r/a$, can easily be transformed to

$$I_k = a^{k+1}(\ln aA^{(k)} + B^{(k)}), \tag{21}$$

where, after substituting $y := 1 - z/|\alpha|$ and $y := -1 + z/|\alpha|$, respectively, the integrals A and B (the second is further transformed, using a functional relation for the logarithm) are given by

$$\begin{aligned} \frac{A^{(k)}}{|\alpha|^{k+1}} - \frac{1}{k+1} &= \int_0^1 dy \frac{(1+y)^k - (1-y)^k}{1+e^{|\alpha|y}} + \int_1^\infty dy \frac{(1+y)^k}{1+e^{|\alpha|y}} \\ &=: A_1(|\alpha|,k) + A_2(|\alpha|,k) \end{aligned} \tag{22}$$

and

$$\begin{aligned} \frac{B^{(k)}}{|\alpha|^{k+1}} - \frac{\ln|\alpha| - 1/(k+1)}{k+1} &= \ln|\alpha|(A_1 + A_2) \\ &\quad + \int_0^1 dy \frac{(1+y)^k \ln(1+y) - (1-y)^k \ln(1-y)}{1+e^{|\alpha|y}} \\ &\quad + \int_1^\infty dy \frac{(1+y)^k \ln(1+y)}{1+e^{|\alpha|y}} \\ &=: \ln|\alpha|(A_1 + A_2) + B_1(|\alpha|,k) + B_2(|\alpha|,k). \end{aligned} \tag{23}$$

Thus we gain for the k th logarithmic moments the more tractable form

$$\begin{aligned} I_k(|\alpha|,c^{k+1}) = c^{k+1}\{ &[\ln c - 1/(k+1)]/(k+1) \\ &+ \ln c(A_1(|\alpha|,k) + A_2(|\alpha|,k)) \\ &+ B_1(|\alpha|,k) + B_2(|\alpha|,k)\}. \end{aligned} \tag{24}$$

The four integrals in Eq. (24) have to be considered separately. We divide them into two groups, depending on whether they contain a logarithmic term or not.

(1) We first concentrate on B_2 as defined in Eq. (23). Since $y \in [1,\infty)$ and $|\alpha|$ at least greater than unity we are able to evaluate the denominator in terms of a geometric series. Therefore

$$B_2 = \sum_{\nu=0}^{\infty} (-1)^\nu \int_1^\infty dy\, (1+y)^k \ln(1+y)\, e^{-|\alpha|(\nu+1)y}. \tag{25}$$

Now substituting successively $z := y + 1$ and then $x := (\nu+1)|\alpha|z$, we obtain the more convenient form

$$B_2 = \sum_{\nu=0}^{\infty} (-1)^\nu e^{\alpha_\nu} \alpha_\nu^{-(k+1)} [\{\partial_x S_{\alpha_\nu}^{(x)}\}_{x=k} - \ln \alpha_\nu S_{\alpha_\nu}^{(k)}] \tag{26}$$

with

$$S_{\alpha_\nu}^{(k)} = \int_{2\alpha_\nu}^\infty dx\, x^k e^{-x}$$

and

$$\alpha_\nu := |\alpha|(\nu+1).$$

Now, $S_{\alpha_\nu}^{(k)}$ can be straightforwardly integrated ($\tilde{\alpha}_\nu := 2\alpha_\nu$)

$$e^{-\tilde{\alpha}_\nu} \tilde{\alpha}_\nu^k \left\{ 1 + \sum_{\mu=1}^{k} k(k-1) \times \cdots \times (k-\mu+1) \tilde{\alpha}_\nu^{-\mu} \right\}, \tag{27}$$

whereas—on the other hand—it is a representation of the Whittaker function $W_{k/2,(k+1)/2}$ multiplied by $\tilde{\alpha}_\nu^{k/2} e^{-\alpha_\nu}$.

For calculation of $\partial_x S_{\alpha_\nu}^{(x)}$ we apply an asymptotic formula for $W_{x/2,(x+1)/2}$, since we discuss $I_k(|\alpha|,c^{k+1})$ especially in the case $|\alpha| \gg 1$,

$$\partial_x [\tilde{\alpha}_\nu^k e^{-\tilde{\alpha}_\nu}(1 + D(x,|\alpha|,\nu))], \tag{28}$$

with the abbreviation

$$D(x,|\alpha|,\nu) := \sum_{j=1}^{\infty} \frac{\prod_{m=1}^{j} [m(x+1) - m^2]}{j!\, \tilde{\alpha}_\nu^j}. \tag{29}$$

Therefore $[\partial_x S_{\alpha_\nu}^{(x)}]_{x=k}$ reads

$$\begin{aligned} \tilde{\alpha}_\nu^k e^{-\tilde{\alpha}_\nu} \{ &\ln \tilde{\alpha}_\nu (1 + D(k,|\alpha|,\nu)) \\ &+ [\partial_x (1 + D(x,|\alpha|,\nu))]_{x=k} \}. \end{aligned} \tag{30}$$

Our next step is to evaluate D and $\partial_x D$, respectively.

We notice that both reduce to finite sums

$$D(k,|\alpha|,\nu) = \sum_{j=1}^{k} \frac{\prod_{m=1}^{j} [m(k+1) - m^2]}{j!\, \tilde{\alpha}_\nu^j} \tag{31}$$

and

$$[\partial_x D(x,|\alpha|,\nu)]_{x=k} = \sum_{j=1}^{k} \frac{1}{j! \tilde{\alpha}_\nu^j} \sum_{m=1}^{j} \frac{m}{[m(k+1) - m^2]} \times \prod_{l=1}^{j} [l(k+1) - l^2]. \tag{32}$$

Finally we get for the last integral in Eq. (24)

$$B_2 \cong \sum_{\nu=0}^{\infty} (-1)^\nu e^{-\alpha_\nu} \left\{ \frac{2^k}{\alpha_\nu} \left[\ln 2 + \ln \tilde{\alpha}_\nu D(k,|\alpha|,\nu) - \ln \alpha_\nu \sum_{\mu=1}^{k} k(k-1)\times\cdots\times(k-\mu+1)\tilde{\alpha}_\nu^{-\mu} \right] + [\partial_x D(x,|\alpha|,\nu)]_{x=k} \right\}. \tag{33}$$

(2) We now concentrate on the solution of the third integral in Eq. (24)

$$B_1 = \int_0^1 dy \frac{(1+y)^k \ln(1+y) - (1-y)^k \ln(1-y)}{1+e^{|\alpha|y}}. \tag{34}$$

Our aim is to do this integral by expansion of the numerator.

We first expand the polynomial terms

$$(1+y)^k \ln(1+y) - (1-y)^k \ln(1-y) = \sum_{i=0}^{k} \binom{k}{i} (\ln(1+y) - (-1)^i \ln(1-y)). \tag{35}$$

Now, since

$$\ln(1+y) = \sum_{k=0}^{\infty} (-1)^{k+1} \frac{y^k}{k}$$

is absolutely convergent for $-1 < y \leqslant 1$, we get

$$\sum_{i=0}^{k} \binom{k}{i} y^i \left(\sum_{\mu=1}^{\infty} (-1)^{\mu+1} \frac{y^\mu}{\mu} - (-1)^i \sum_{\mu=1}^{\infty} \frac{y^\mu}{\mu} \right), \tag{36}$$

for $-1 < y < 1$.

We add both infinite sums componentwise for $y \in [0,1)$ and exchange the resulting infinite sum with the finite (polynomial) sum:

$$\sum_{\mu=1}^{\infty} \sum_{i=0}^{k} \binom{k}{i} ((-1)^i - (-1)^\mu) \frac{y^{\mu+i}}{\mu}. \tag{37}$$

We note that this sum takes at $y = 1$ the value $2^k \ln(2)$. Since the numerator in Eq. (34) takes at $y = 1$ the same value, we can say that Eq. (37) is a sum representation for this term, being absolutely convergent in $[0,1)$ and continuous at $y = 1$.

Defining a new sum index $\gamma = \mu + i$ such that we get groups of constant power in y, Eq. (37) reads

$$\sum_{\gamma=1}^{\infty} c_\gamma y^\gamma, \tag{38}$$

where

$$c_\gamma := \sum_{\gamma = 1 + A_\gamma^{(k)}}^{\mu} \binom{k}{\gamma - \mu} ((-1)^{\gamma-\mu} - (-1)^\mu) \frac{1}{\mu} \tag{39}$$

vanishes for even γ, changing sign with μ. Here we introduce $A_\gamma^{(k)} = (\gamma - k - 1)\Theta(\gamma - k - 2)$ as a shift parameter assuring finiteness of i.

Therefore,

$$c_\gamma = \begin{cases} 0, & \text{for even } \gamma, \\ 2 \displaystyle\sum_{\mu = 1 + A_\gamma^{(k)}}^{\gamma} \binom{k}{\gamma - \mu} \frac{(-1)^{\mu+1}}{\mu}, & \text{for odd } \gamma. \end{cases} \tag{40}$$

So

$$B_1(|\alpha|,k) = \sum_{\gamma=0}^{m} c_{\gamma_2} \int_0^1 dy\, y^{\gamma_2} (1 + \exp(|\alpha|y))^{-1} + \tilde{R}(|\alpha|,k;m), \tag{41}$$

with the definition $\gamma_2 := 2\gamma + 1$, and a remainder

$$\tilde{R}(|\alpha|,k;m) := \int_0^1 dy \{ [(1+y)^k \ln(1+y) - (1-y)^k \ln(1-y)] - \sum_{\gamma=0}^{m} c_{\gamma_2} y^{\gamma_2} \} (1 + \exp(|\alpha|y))^{-1} \tag{42}$$

due to the cutoff in the summation (numerical calculations show that this last term vanishes as m goes to infinity).

The integral in Eq. (41) is done by us through geometric series expansion, exchange of sum and integral, and straightforward evaluation of the remaining integral:

$$\sum_{\gamma=0}^{m} c_{\gamma_2} |\alpha|^{-(\gamma_2+1)} \gamma_2! Z(\gamma_2 + 1) - \sum_{\mu=0}^{2m+1} \left(\sum_{\gamma=[\mu/2]}^{m} c_{\gamma_2} \frac{\gamma_2!}{(\gamma_2 - \mu)!} \right) |\alpha|^{-(\mu+1)} F_\mu(|\alpha|). \tag{43}$$

The result is a difference between two infinite sums; the first one is a representation of the Z function,

$$Z(\gamma_2 + 1) = \sum_{\nu=1}^{\infty} (-1)^{\nu-1} \nu^{-(\gamma_2+1)}, \tag{44}$$

the second one is abbreviated through

$$F_\mu(|\alpha|) := \sum_{\nu=1}^{\infty} (-1)^{\nu-1} \frac{\exp(-\nu|\alpha|)}{\nu^{\mu+1}}. \tag{45}$$

(3) Let us now concentrate on the two remaining integrals in Eq. (24):

$$A_1(|\alpha|,k) = \int_0^1 dy \frac{(1+y)^k - (1-y)^k}{1+e^{|\alpha|y}} \tag{46}$$

and

$$A_2(|\alpha|,k) = \int_1^\infty dy \frac{(1+y)^k}{1+e^{|\alpha|y}}. \tag{47}$$

Applying to $A_2(|\alpha|,k)$ the same transformations as given in part 1 of this section we reach

$$A_2(|\alpha|,k) = \sum_{\nu=1}^{\infty} (-1)^\nu 2^k \frac{e^{-\alpha_\nu}}{\alpha_\nu} \left\{ 1 + \sum_{\mu=1}^{k} k(k-1) \times\cdots\times(k-\mu+1)\tilde{\alpha}_\nu^{-\mu} \right\}. \tag{48}$$

In $A_1(|\alpha|,k)$ we first expand the binomials and then use Eq. (43), simply replacing

$$\sum_{\gamma=0}^{m} c_{2\gamma+1} \quad \text{by} \quad 2 \sum_{i=0}^{[(k-1)/2]} \binom{k}{2i+1};$$

therefore

$$A_1(|\alpha|,k) = 2 \sum_{i=0}^{[(k-1)/2]} \binom{k}{i_2} |\alpha|^{-(i_2+1)} i_2! Z(i_2+1) - 2 \sum_{\mu=0}^{2[(k-1)/2]+1} \sum_{i=[\mu/2]}^{[(k-1)/2]} \binom{k}{i_2} |\alpha|^{-(\mu+1)} \times \frac{i_2!}{(i_2-\mu)!} F_\mu(|\alpha|) \quad (i_2 := 2i+1). \tag{49}$$

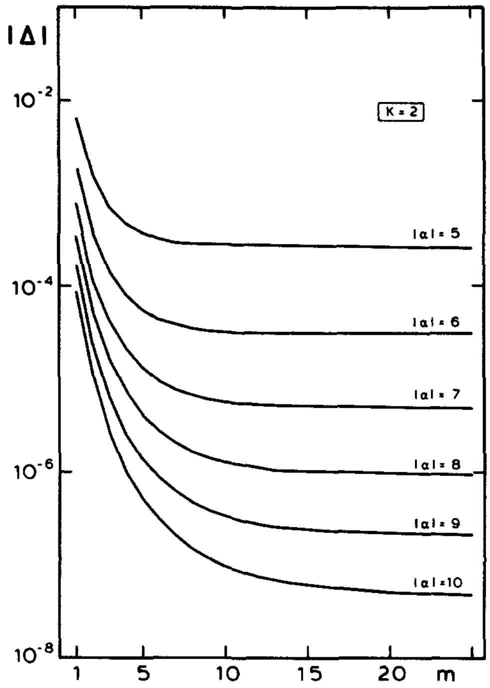


FIG. 1. The relative deviation Δ of our semiconvergent expansion of the Fermi moment $I_2(|\alpha|)$ to the exact result for increasing leptodermicity of the saturating function $(1+\exp(y-|\alpha|))^{-1}$.

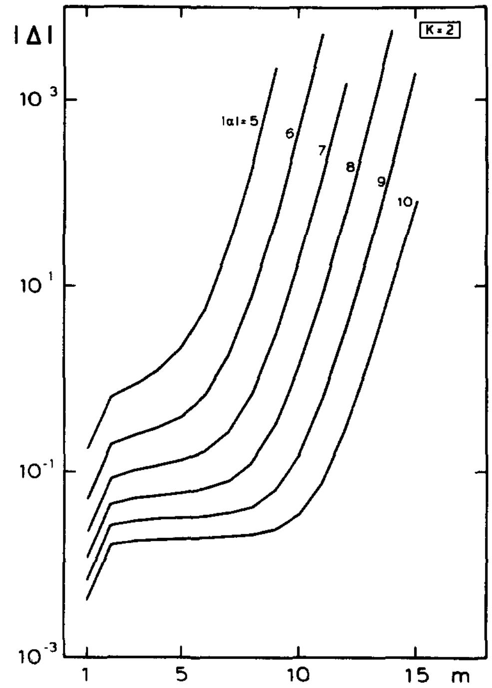


FIG. 3. Same as Fig. 2 but for the semidivergent expansion of Ref. 7.

IV. RESULTS

Testing the new method to evaluate the integrals $I_k(|\alpha|,c^{k+1})$ (Sec. III), and summing up to about $m=20$ (depending on the choice of the other parameters) the integral expansion appears to be semiconvergent. Beyond this the results start to be numerically unstable, because the expansion given for B_1 [see Eq. (43)] consists of a converging difference between two divergent sums. We will call this behavior semiunstable.

An alternative way of evaluating B_1 is now to identify it with an integral representation of the incomplete Γ function:

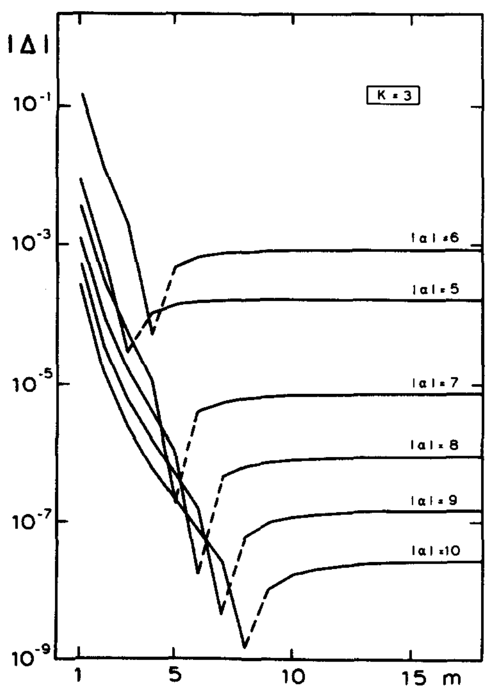


FIG. 2. Same as Fig. 1 but for the third moment $k=3$.

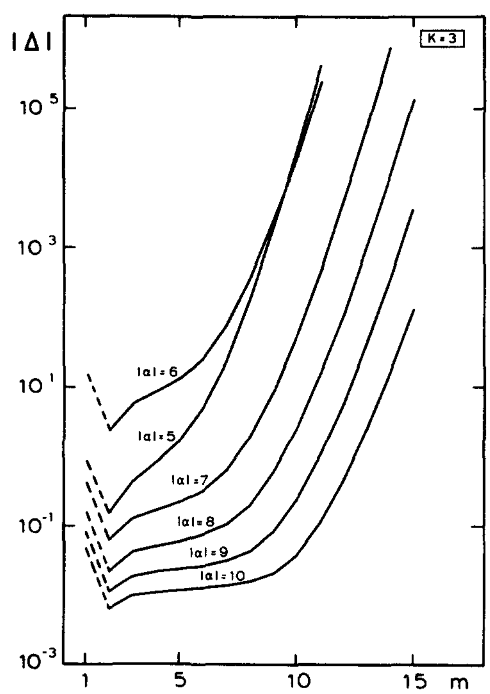


FIG. 4. Same as Fig. 3 but for the semidivergent expansion of Ref. 7.

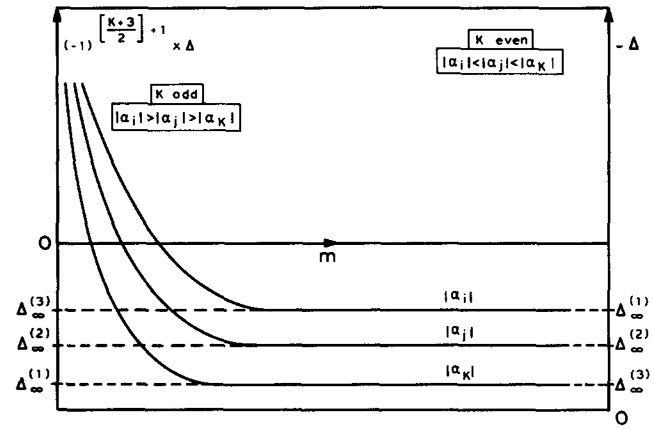


FIG. 5. The asymptotic behavior of the relative deviation Δ of our semiconvergent expansion of $I_k(|\alpha|)$. For all k and α a finite, small, but nonzero asymptotic limit is reached.

$$\sum_{\gamma=0}^{m} c_{2\gamma+1} \Bigg\{ \frac{1}{2\gamma+2} \sum_{\nu=0}^{\infty} (-1)^\nu e^{-\alpha_\nu} + \sum_{i=1}^{\infty} \Bigg\{ \Bigg(\sum_{\nu=0}^{\infty} \Bigg) \big[\Theta(-d_\alpha^{(i)}) - \Theta(d_\alpha^{(i)} - 1) \big] + \Bigg(\sum_{\nu=d_\alpha^{(i)}}^{\infty} + \sum_{\nu=d_\alpha^{(i)}-1}^{1} \Bigg) \Theta(d_\alpha^{(i)} - 1) \Bigg\} (-1)^\nu \times \frac{1}{2\gamma+2+i} \prod_{j=1}^{i} \frac{\alpha_\nu \exp(-\alpha_\nu/i)}{(2\gamma+j+1)!/(2\gamma+1)!} \Bigg\}, \qquad (43')$$

with $d_\alpha^{(i)} := [i/|\alpha| - \frac{1}{2}]$, (see Ref. 9) where we factorized the powers in $|\alpha|$ and the faculties, splitting the sum over ν near the maximum of the sum kernel in order to assure numerical stability for explicit calculation.

This new method for calculation of B_1 defines our second, semiconvergent version, which is numerically stable, but more tedious. Since the results of both methods are nearly identical over the usual range of use the user may choose the semiconvergent, but semiunstable version as given in Sec. III for fast numerical calculations and analytical considerations.

This last point is very important, because all expansions are there given in well-ordered powers in $|\alpha|$ and $e^{|\alpha|}$.

On the other hand, one may prefer the much more slow, semiconvergent but numerically stable version by substituting Eq. (43') for Eq. (43).

We calculated $I_k(|\alpha|, c^{k+1})$ in the case $c = 1$ for $k = 2,3$ and $|\alpha|$ varying between 5 and 10.

The function $\Delta_k(|\alpha|, m)$ represents the relative, the function $\delta_k(|\alpha|, m)$ the absolute deviation from the exact value of $I_k(|\alpha|, c^{k+1})$.

Figs. 1 and 2 show the results of the calculation with our semiconvergent, stable method and Figs. 3 and 4 the corresponding ones using the Åberg method.[7]

As sketched in Fig. 5 the term $\Delta_k(|\alpha|, m)$ depends on k in a characteristic way. It changes sign if k is an odd integer and reaches a specific nonvanishing value as m goes to infinity. That is what characterizes our second method as only semiconvergent, which is simply due to the substitution of $W_{k/2,(k+1)/2}$ by an asymptotic ($|\alpha| \to \infty$) approximation.

As Fig. 6 shows, the absolute deviation $\delta_k(|\alpha|, m \to \infty)$ is a monotonic decreasing function of $|\alpha|$, falling off almost exponentially [the absolute instead of the relative deviation is more advisable here because $I_{k=3}(|\alpha|, c=1)$ passes through zero for $|\alpha| = 5.961\ 10\ldots$; this is the reason as well for the "anomalous" shift of $\Delta_{k=3}(|\alpha| = 6, m)$ in Figs. 2 and 4].

The mathematical results given here are needed in the inverse problem of evaluating properties of the underlying saturating function if only some surface moments are experimentally known. The method of general surface moments and the result of a semiconvergent series for Fermi-type functions may prove useful in many fields of physics where saturating distributions occur, such as the density of electrons at a surface, the fermi distribution in momentum space, the occupation distribution in superconductors, etc.

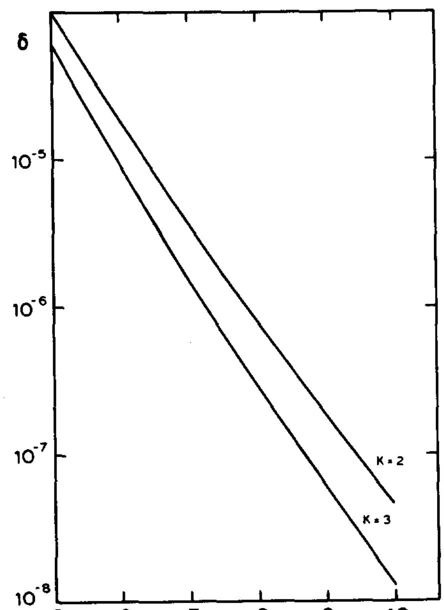


FIG. 6. The absolute deviation δ of the full semiconvergent series of the Fermi moments $I_k(|\alpha|)$ shows the efficient approximation of the exact result.

ACKNOWLEDGMENTS

We are thankful to W. Feix for some interesting and stimulating suggestions. We also want to thank the Gesellschaft für Schwerionenforschung (GSI) in Darmstadt for providing their excellent computer facilities.

This work was supported in part by the Deutsche Forschungsgemeinschaft, Grant No. Hi 164/14.

[1]G. Süssmann, Zeitschr. F.Phys. A **274**, 145 (1975).

[2]W. D. Myers, and W. J. Swiatecki, Ann. Phys. **55**, 395 (1969).

[3]K. W. Ford and J. G. Wills, Phys. Rev. **185**, 1429 (1969).

[4]E. R. Hilf and R. Wolff, in *Proceedings of the 4th International Conference on Nuclei Far from Stability*, Helsingør, 1981, CERN 81-09, p. 805.

[5]W. D. Bauer, E. R. Hilf, K. Koebke, and F. Schmitz, in *Modern Developments in Thermodynamics*, edited by B. Gal-or (Wiley, New York, 1974), pp. 371–390.

[6]The prefix semi- means asymptotically.

[7]S. Åberg, Nordita preprint 82/27, Sept. 1982.

[8]H. Gräf, "Note on the Numerical Calculation of the Fermi–Dirac integrals," Arbeitsbericht BMFT NV 2002 C, (1977); H. Gräf and M. Pabst, preprint Darmstadt IKDA 77/24, 1974.

[9]These brackets represent the operation of taking the integer part of the number they include.

Inequalities of Schwarz and Hölder type for random operators

Gerhard C. Hegerfeldt
Division of Physics, Mathematics, and Astronomy,[a] *California Institute of Technology, Pasadena, California 91125 and Institut für Theoretische Physik, Universität Göttingen, Göttingen, West Germany*

(Received 13 November 1984; accepted for publication 4 January 1985)

Let A and B be random operators on a Hilbert space, and let $\langle\ \rangle$ denote averages (expectations). We prove the inequality $\|\langle A^*B\rangle\|\leqslant\|\langle A^*A\rangle\|^{1/2}\|\langle B^*B\rangle\|^{1/2}$. A generalized Hölder inequality involving traces is also proved.

I. SCHWARZ INEQUALITY

In this paper we prove two inequalities, one of which was announced and extensively used before.[1] Despite the simple proof, the inequalities seem not to have been published elsewhere; only a special case of the Schwarz-type inequality for commuting operators has appeared.[2]

A random operator A is an operator-valued function $A(\cdot)$ on some space Ω, with a given probability measure μ on Ω. Averages or expectations are denoted interchangeably by $\langle\ \rangle$ or E. Thus

$$\langle A\rangle \equiv EA := \int_\Omega A(\omega)d\mu(\omega).$$

We have to be a little bit more precise. Let $\mathscr{H}$ be a separable Hilbert space and $B(\mathscr{H})$ the set of all bounded operators on $\mathscr{H}$. Let Σ be the σ algebra of sets in Ω on which μ is defined. We assume that the complex-valued function $(\varphi, A(\cdot)\psi)$ is Σ measurable for all $\varphi, \psi\in\mathscr{H}$. Thus, a random operator is a weakly measurable $B(\mathscr{H})$-valued function. A typical example is a random matrix.

It may happen that the expectation or average $E(\varphi, A(\cdot)\psi)$ exists for all $\varphi,\psi\in\mathscr{H}$; if in addition there is a bounded operator, denoted by EA or $\langle A\rangle$, such that

$$(\varphi, EA\psi) = E(\varphi,\ A(\cdot)\psi),$$

we say that the expectation of A exists (in the Pettis sense) and is given by EA, or $\langle A\rangle$. Clearly, $\langle A\rangle$ exists if $\langle\|A\|\rangle$ does, and then $\|\langle A\rangle\|\leqslant\langle\|A\|\rangle$.

Theorem 1: Let A and B be random operators on a Hilbert space $\mathscr{H}$. Then

$$\|\langle A^*B\rangle\|\leqslant\|\langle A^*A\rangle\|^{1/2}\|\langle B^*B\rangle\|^{1/2}, \tag{1.1}$$

where the existence of the right-hand side (rhs) implies the existence of the left-hand side (lhs) and where $\|\cdot\|$ is the usual operator norm on $B(\mathscr{H})$.

The consequences are analogous to those of Schwarz's inequality and are proved in a similar way.

Corollary 1: For a random operator A we define

$$\|A\|_\mu := \|\langle A^*A\rangle\|^{1/2},$$

if the rhs exists. Then

$$\|A+B\|_\mu\leqslant\|A\|_\mu+\|B\|_\mu \quad \text{(triangle inequality)},$$

$$|\|A\|_\mu-\|B\|_\mu|\leqslant\|A-B\|_\mu,$$

$$\|\langle A^*A-B^*B\rangle\|\leqslant\{\|A\|_\mu+\|B\|_\mu\}\|A-B\|_\mu.$$

If $\|A\|_\mu=0$, then $A=0$ with probability 1. Thus $\|\cdot\|_\mu$ is a norm on (equivalence classes of) random variables.

Proof of Theorem 1: We use the ordinary Schwarz inequality, first for the $d\mu$ intergral (i.e., for expectations) and then for the scalar product in $\mathscr{H}$. Let the rhs of Eq. (1.1) exist. Then

$$\langle\|A\varphi\|^2\rangle = E(\varphi, A^*A\varphi) = \|\langle A^*A\rangle^{1/2}\varphi\|^2 \tag{1.2}$$

and similarly for B. The two Schwarz inequalities now give

$$\langle\|A\varphi\|^2\rangle^{1/2}\langle\|B\psi\|^2\rangle^{1/2}\geqslant\langle\|A\varphi\|\|B\psi\|\rangle$$
$$\geqslant|E(\varphi, A^*A\psi)|, \tag{1.3}$$

with existence implied. We now take the sup over φ and ψ with $\|\varphi\|=\|\psi\|=1$. This shows that the rhs of Eq. (1.3) defines a bounded operator $\langle A^*B\rangle$. Since $\|\langle A^*A\rangle^{1/2}\|^2=\|\langle A^*A\rangle\|$, its norm is bounded by the rhs of Eq. (1.1). Q.E.D.

Using the existence statement of Theorem 1, the following analog of an inequality of Lieb and Ruskai[3] is an easy consequence.

Corollary 2: Let $\langle A^*A\rangle$ and $\langle B^*B\rangle$ exist. Then for any $\epsilon>0$

$$\langle A^*B\rangle\{\langle B^*B\rangle+\epsilon\}^{-1}\langle B^*A\rangle\leqslant\langle A^*A\rangle. \tag{1.4}$$

As a consequence

$$\langle A^*B\rangle\langle B^*A\rangle\leqslant\|\langle B^*B\rangle\|\langle A^*A\rangle. \tag{1.5}$$

Proof: Let $Q := \{\langle B^*B\rangle+\epsilon]^{-1}\langle B^*A\rangle$, which is a nonrandom operator. Then one has

$$0\leqslant(A-BQ)^*(A-BQ)+\epsilon Q^*Q.$$

Expanding and taking expectation gives Eq. (1.4). From

$$\|\langle B^*B\rangle+\epsilon\|^{-1}{}_1\leqslant\{\langle B^*B\rangle+\epsilon\}^{-1},$$

one then obtains Eq. (1.5). Both also follow directly from Eq. (1.3). Q.E.D.

An alternative proof of Theorem 1 was proposed to me, which is based on the observation that

$$\begin{pmatrix} A^*A & A^*B \\ B^*A & B^*B\end{pmatrix}\geqslant 0.$$

Sandwiching this with $\phi\equiv\begin{pmatrix}\lambda\varphi\\ \psi/\lambda\end{pmatrix}$ from both sides and then taking expectations and the sup over φ, ψ yields Eq. (1.1). In a similar way Corollary 2 can be derived directly, with B^*B replaced by $B^*B+\epsilon$.

We remark in passing that for $\dim\mathscr{H}<\infty$ the normed space $\{A; \|\langle A^*A\rangle\|^{1/2}\equiv\|A\|_\mu<\infty\}$ is complete, and the norms $\|A\|_\mu$ and $\|A^*\|_\mu$ are equivalent. For $\dim\mathscr{H}=\infty$ this is in general not true, and $\|\langle A^*A\rangle\|<\infty$ does not imply $\|\langle AA^*\rangle\|<\infty$.

[a] Address until October 1985.

 0022-2488/85/071576-02$02.50

II. HÖLDER INEQUALITY

Defining $|A|^p := (A^*A)^{p/2}$, Eq. (1.1) can be written as

$$\|\langle A^*B\rangle\| \leqslant \|\langle |A|^2\rangle\|^{1/2}\|\langle |B|^2\rangle\|^{1/2}.$$

The corresponding Hölder-type inequality for $p \neq 2$ does not hold in general if dim $\mathscr{H} \geqslant 2$. This can be shown by counterexamples. For trace norms, however, one has the following.

Theorem 2: Let A and B be random operators on a Hilbert space $\mathscr{H}$. Then, for $r \geqslant 1$ and $1/p + 1/q = 1/r, p, q > 0$,

$$\{\mathrm{Tr}|\langle A^*B\rangle|^r\}^{1/r} \leqslant \mathrm{Tr}\{\langle |A^*B|\rangle^r\}^{1/r} \leqslant \{\mathrm{Tr}\langle |A|^p\rangle\}^{1/p}\{\mathrm{Tr}\langle |B|^q\rangle\}^{1/q}, \quad (2.1)$$

where existence of the rhs implies existence of the rest. Here A^* may be replaced by A in the middle and the lhs.

Similarly as before, we use Hölder's inequality for integrals and then for trace norms. But first we note a simple fact.

Lemma 1: On positive random operators, trace and expectations commute,

$$E\,\mathrm{Tr}|A| = \mathrm{Tr}\,E\,|A|, \quad (2.2)$$

and existence of either side implies that of the other. In this case A is trace class almost surely, EA exists and is trace class, and

$$\mathrm{Tr}\,EA = E\,\mathrm{Tr}\,A. \quad (2.3)$$

Proof: Let $\{\varphi_n\}$ be an orthonormal basis in $\mathscr{H}$. Then

$$E\,\|A\| \leqslant E\,\mathrm{Tr}|A| = \sum_n E\,(\varphi_n, |A|\varphi_n),$$

by positivity. Hence, if the rhs is finite then $E\,|A|$ and EA exist as bounded operators and the rhs equals $\mathrm{Tr}\,E\,|A|$. Equation (2.3) then follows from Lebesgue's bounded convergence. Q.E.D.

Proof of Theorem: By Hölder's inequality,[4] first for integrals and then for trace norms, we have

$$\{E\,\mathrm{Tr}|A|^p\}^{1/p}\{E\,\mathrm{Tr}|B|^q\}^{1/q} \geqslant \{E\,[(\mathrm{Tr}|A|^p)^{1/p}(\mathrm{Tr}|B|^q)^{1/q}]^r\}^{1/r} \geqslant \{E\,\mathrm{Tr}|A^*B|^r\}^{1/r}. \quad (2.4)$$

By Lemma 1, this proves the second part of Eq. (2.1), together with existence. The remainder follows from Lemma 2.

Lemma 2: Let A be a random operator and let $p \geqslant 1$. Then

$$\mathrm{Tr}|EA|^p \leqslant \mathrm{Tr}\,E\,|A|^p, \quad (2.5)$$

where existence of the rhs implies that of the lhs.

Proof: Let the rhs exist. By Lemma 1, $|A|^p$ is trace class almost surely. Since $\|A\|^p \leqslant \mathrm{Tr}|A|^p$ one has

$$E\,\|A\| \leqslant \{E\,\|A\|^p\}^{1/p} < \infty.$$

Hence EA exists as a bounded operator.

Now let X be any nonrandom operator with $|X|^q$ trace class,[5] $1/p + 1/q = 1$. Then, by the second half of Eq. (2.1), $E\,|XA|$ exists and is trace class. Thus, by Lemma 1, $EXA = XEA$ is also trace class. By duality one now has[5]

$$\{\mathrm{Tr}|EA|^p\}^{1/p} = \sup_{\mathrm{Tr}|X|^q = 1} |\mathrm{Tr}\,XEA| \leqslant E \sup_{\mathrm{Tr}\,|X|^q = 1} |\mathrm{Tr}\,XA| = E\,\{\mathrm{Tr}|A|^p\}^{1/p} \leqslant \{E\,\mathrm{Tr}|A|^p\}^{1/p}. \quad (2.6)$$

Q.E.D.

It was pointed out to me that the argument in Eq. (2.4) can be replaced by an equivalent linear version of Hölder's inequality, i.e.,

$$r^{-1}\,\mathrm{Tr}|A^*B|^r \leqslant p^{-1}\lambda^p\,\mathrm{Tr}|A|^p + q^{-1}\lambda^{-q}\,\mathrm{Tr}|B|^q,$$

for all $\lambda > 0$. Taking expectation and using Lemma 1 also gives the second part of Eq. (2.1).

Remark: Finiteness of the measure μ has only entered in the proof of Lemma 2. For nonfinite μ, Theorem 1, Theorem 2 with $r = 1$, and Lemma 2 with $p = 1$ still hold, as does the second inequality of Theorem 2 for any $r \geqslant 1$.

ACKNOWLEDGMENTS

I would like to thank Mary Beth Ruskai for pointing out Ref. 2. My thanks also go to W. A. J. Luxemburg and B. Simon for the hospitality extended to me at Caltech.

This work was supported in part by the Stiftung Volkswagenwerk.

[1]G. C. Hegerfeldt and R. Reibold, J. Stat. Phys. **32**, 313 (1983).

[2]M. Hasell Rizvi and R. W. Shorrock, in *Studies in Probability and Ergodic Theory, Advances in Mathematics Supplementary Studies*, edited by G-C. Rota (Academic, New York, 1978), Vol. 2, p. 113.

[3]E. H. Lieb and M. B. Ruskai, Adv. Math. **12**, 269 (1974).

[4]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vols. I and II.

[5]If $p = 1$, we take X bounded and replace $\{\mathrm{Tr}|X|^q\}^{1/q}$ by $\|X\|$.

Graded manifolds and vector bundles: A functorial correspondence

M. Batchelor
Cambridge University, Department of Pure Mathematics and Mathematical Statistics, 16 Mill Lane, Cambridge CB2 1SB, England

(Received 21 August 1984; accepted for publication 21 December 1984)

A functorial correspondence between the category of graded manifolds and the category of vector bundles is given. Given a graded manifold (X,A), a vector bundle G over X is given as a subset of the product $X \times A^0$, where A^0 is the dual coalgebra of A. This bundle has an obvious coalgebra structure on each fiber. The correspondence is achieved by showing that the sheaf A is isomorphic as a sheaf of algebras to the sheaf of sections of the vector bundle G' dual to G.

I. INTRODUCTION

Graded manifolds are introduced in Refs. 1 and 2 to provide an analog of "graded Lie group" to recapture explicitly the geometry implicit in the algebraic structure of Lie superalgebras. In Ref. 3, graded manifolds are classified by isomorphism classes of vector bundles. The purpose of this paper is to strengthen that result, giving a completely different proof. The new result is that there are functors between the category of graded manifolds and the category of vector bundles which establish an equivalence of categories.

There are three advantages to the new proof. First, the proof is constructive and canonical. Second, the existence of an equivalence between the two categories should allow one to exploit the machinery developed for vector bundles. Finally, the proof illustrates the potential of coalgebra theory, an underemployed technique in this field. The vector bundle in question is constructed from a subset of the dual coalgebra using a natural topology on the dual coalgebra.

Basic definitions and notation are given in Sec. II. Section III summarizes results about coalgebras and dual coalgebras which are needed. Section IV develops the main tool—the construction of vector bundles from the dual coalgebra. The main result is proved in Sec. V.

II. NOTATION AND BASIC DEFINITIONS

Vector spaces: All algebras and vector spaces are algebras and vector spaces over the real numbers. The vector space spanned by elements $v_1,\dots,v_k$ will often be denoted $\langle v_1,\dots,v_k\rangle$.

Manifolds: All manifolds considered are smooth paracompact connected Hausdorff manifolds.

Sheaves: If A is a sheaf of vector spaces over a manifold X, A restricts to a sheaf $A|_U$ over an open subset U in A. Usually $A|_U$ will simply be written as A. If B is another sheaf over X, a map of sheaves $\sigma: A\to B$ is a collection of maps $\sigma(U): A(U)\to B(U)$ for each open set U in X such that the maps $\sigma(U)$ commute with restriction to subsets. Often the map $\sigma(U)$ will simply be written σ.

Dual spaces: If V is a vector space its dual space will be denoted by V'. If v is in V and ξ is in V', evaluation of v at ξ will be denoted by $\langle v,\xi\rangle$.

Exterior algebras: If V is a vector space the exterior algebra on V, denoted ΛV, is the graded commutative algebra freely generated by V contained in $(\Lambda V)_1$. If V has a basis $\{v_1,\dots,v_n\}$, ΛV has a basis

$$\{v_1^{j_1}\wedge\cdots\wedge v_n^{j_n}: j_i\in\{0,1\}\}.$$

In fact ΛV is a Z graded commutative algebra, with ΛV_k spanned by basis elements $v_1^{j_1}\wedge\cdots\wedge v_n^{j_n}$, $\Sigma j_i = k$.

Graded manifolds: A graded manifold is a pair (X,A), where X is a manifold and A is a sheaf of Z_2 graded commutative algebras such that we have the following.

(i) There is a surjective map of sheaves

$$\epsilon: A\to C^\infty,$$

where C^∞ is the sheaf of smooth real valued functions on X.

(ii) There is an open cover $\{U_\alpha\}$ of X and isomorphisms of sheaves of Z_2 graded commutative algebras

$$\tau_\alpha: A|_U \to C^\infty \otimes \Lambda R^s|_U.$$

The maps τ_α are called trivializations, and A is said to be trivial over the sets U_α.

Lemma 2.1: If (X,A) is a graded manifold, the sheaf A is a fine sheaf over x. That is, for any open set U with an open subset V, with V equal to the interior of its closure in V, the restriction map

$$\rho(V,U): A(U)\to A(V)$$

is onto.

Proof: Let $\{U_\alpha\}$ be an open cover of U with trivializations τ_α, and suppose that U_α is the interior of its closure in U for each α. Also, since X is a nice manifold, the collection $\{U_\alpha\}$ can be indexed by the positive integers. Write

$$V_i = V\cup U_1\cup U_2\cup\cdots\cup U_i.$$

Now let a_0 be in $A(V)$. The trivialization τ_1 provides a commutative square

$$\begin{tikzcd} A(U_1) \arrow[r, "\tau_1"] \arrow[d] & C^\infty(U)\otimes \Lambda R^s \arrow[d] \\ A(U_1\cap V) \arrow[r, "\tau_1"] & C^\infty(U_1\cap V)\otimes \Lambda R^s. \end{tikzcd}$$

Since the sheaf C^∞ is fine and both rows are isomorphisms, a_0 restricted to $U_1\cap V$ is the restriction of some b_1 in $A(U_1)$. Now $\{V,U_1\}$ is an open cover of V_1, and $b_1 = a_0$ on $U_1\cap V$. Thus there exists a_1 in $A(V_1)$ with $a_1 = b$ on U_1, $a_1 = a_0$ on V.

Proceed inductively to find elements a_i in $A(V_i)$ $a_i = a_j$ on $V_i\cap V_j$. Thus there exists a in $A(\cup V_j)$ with $a = a_0$ on V, as desired.

III. DUAL COALGEBRAS

The dual coalgebras: Let A be an algebra. Define

$$A^0 = \{\alpha\in A': \ker\alpha > I,\ I \text{ an ideal, } \dim A/I<\infty\}.$$

 0022-2488/85/071578-05$02.50

The vector space A^0 has the following properties.

(i) A^0 is a coalgebra with comultiplication

$$\Delta: A^0 \to A^0 \otimes A^0$$

induced by multiplication in A.

(ii) If A and B are algebras, $(A \otimes B)^0 = A^0 \otimes B^0$.

(iii) If $f: A \to B$ is a map of algebras, f induces a map $f^0: A^0 \to B^0$ of coalgebras.

The coalgebra A^0 is called the dual coalgebra. See Sweedler[4] for the basic properties of the dual coalgebra.

Examples: (i) The exterior algebra ΛV. Suppose V is a vector space of dimension s. Then ΛV is itself finite dimensional and $(\Lambda V)^0 = (\Lambda V)' = (\Lambda V')$. Let $\{\xi_i\}$ be a basis for V'. Let $\mathbf{h} = (h_1,...,h_s)$, $\mathbf{k} = (k_1,...,k_s)$. Comultiplication is given by the formula

$$\Delta \xi_1^{j_1} \wedge \cdots \wedge \xi_s^{j_s}$$
$$= \sum \sigma(\mathbf{h},\mathbf{k}) \xi_1^{h_1} \wedge \cdots \wedge \xi_s^{h_s} \otimes \xi_1^{k_1} \wedge \cdots \wedge \xi_s^{k_s}$$

where $h_i + k_i = j_i$ and $\sigma(\mathbf{h},\mathbf{k})$ is defined by

$$\xi_1^{j_1} \wedge \cdots \wedge \xi_s^{j_s} = \sigma(\mathbf{h},\mathbf{k}) \xi_1^{j_1} \wedge \cdots \xi_s^{h_s} \xi_1^{k_1} \wedge \cdots \wedge \xi_s^{k_s}.$$

(ii) The algebra of smooth functions: Let U be an open set in R^r. Then

$$C^\infty(U)^0 = \left\langle \left. \frac{\partial^{\Sigma j_i}}{\partial x_1^{j_1} \cdots \partial x_r^{j_r}} \right|_u : u \in U, \quad j_i = 0,1,2,... \right\rangle$$

and comultiplication is given by the generalized Leibnitz rule

$$\Delta \left. \frac{\partial^{\Sigma j_i}}{\partial x_1^{j_1} \cdots \partial x_r^{j_r}} \right|_u = \sum \frac{c(\mathbf{h},\mathbf{k}) \partial^{\Sigma k_i}}{\partial x_1^{k_1} \cdots \partial x_r^{k_r}} \otimes \frac{\partial^{\Sigma h_i}}{\partial x_1^{h_1} \cdots \partial x_r^{h_r}},$$

$$h_i + k_i = j_i, \quad c(\mathbf{h},\mathbf{k}) = \binom{j_1}{k_1} \cdots \binom{j_r}{k_r}.$$

(iii) $C^\infty(U) \otimes \Lambda R^s$: This is the example that will be most useful for analyzing graded manifolds. Using property (ii) of the dual coalgebra,

$$(C^\infty(U) \otimes \Lambda R^s)^0$$
$$= \left\langle \left. \frac{\partial^{\Sigma k_i}}{\partial x_1^{k_1} \cdots \partial x_e^{k_r}} \right|_u \otimes \xi_1^{j_1} \wedge \cdots \wedge \xi_s^{j_s}: \right.$$
$$\left. -u \in U, \quad j_i \in \{0,1\}, \quad k_i = 0,1,2,... \right\rangle.$$

For a discussion of these facts, see Kostant.[1]

The following proposition concerning ΛV^0 will be needed in Sec. IV.

Proposition: Let $\sigma: \Lambda V^0 \to \Lambda V^0$ be a map of coalgebras which is the identity on V'. Then σ is the identity.

Proof: The coalgebra map σ induces an endomorphism of algebras

$$\sigma': (\Lambda V^0)' \to (\Lambda V^0)'.$$

There is a canonical isomorphism identifying $(\Lambda V^0)'$ with ΛV. The statement that σ is the identity on V' implies that σ' is the identity on V. Thus σ' must be the identity, and hence σ must be the identity.

IV. VECTOR BUNDLES CONTAINED IN THE DUAL COALGEBRA

Let (X,A) be a graded manifold and let p be a point in X. Let $A^0(p)$ be the maximal subcoalgebra of $A(X)^0$ containing Rp as its unique simple subcoalgebra. Suppose for each p one chooses a finite-dimensional subspace K_p of $A^0(p)$. Form a "bundle" over X by setting

$$K = \cup K_p, \quad \pi: K \to X, \quad \pi(k) = p, \quad \text{if } k \text{ is in } K_p.$$

Put a topology on A^0 by defining basic sets as follows. For $\epsilon > 0$, n a positive integer, and elements $a_1,...,a_n$, define

$$B(\epsilon; a_1,...a_n) = \{\alpha \in A^0: |\langle a_i,\alpha \rangle| < \epsilon, i = 1,...,n\}.$$

With the topology generated by this base, A^0 becomes a Hausdorff topological vector space.

Lemma: The inclusion

$$\theta: X \to A^0$$

is a homeomorphism of X onto its image.

Proof: First show that any open set U in X contains a set of the form $B(a_1,...,a_n; \epsilon) \cap X$. To do this pick an element u in U and a suitable neighborhood W of u having the following properties: (i) A has a trivialization τ over W, and (ii) W contains a neighborhood Z of u with the closure of Z contained in U.

Let ψ be a function on X with

$$\psi(x) = 1, \quad x \text{ not in } Z,$$
$$1 \geqslant \psi(x) > 0, \quad x \neq u,$$
$$\psi(u) = 0.$$

The trivialization provides a commutative square

$$\begin{array}{ccc} A(W) & \longrightarrow & C^\infty(W) \otimes \Lambda R^s \\ \downarrow & & \downarrow \\ A(W - Z) & \longrightarrow & C^\infty(W - Z) \otimes \Lambda R^s. \end{array}$$

Let $a = \tau^{-1}(\psi \otimes 1)$ in $A(W)$, and let $b = 1$ in $A(X - Z)$. Since $a = b$ in $A(W \cap (X - Z))$, there exists some element c in $A(X)$ with $a = c$ on W and $a = b$ on $X - Z$. Consider the set $B(c, \frac{1}{2}) \cap X$.

For x in X,

$$|\langle c,x \rangle| = |\epsilon c(x)| = \psi(x),$$

and hence $|\langle c,x \rangle| < \frac{1}{2}$ implies that x is in $Z \leqslant U$.

Now show the converse, that given a set

$$S = (\alpha + B(a_1,...,a_n,\delta)) \cap X,$$

there exists an open set V in X with V contained in S. Suppose that s is in S. Since $\langle a_i,x \rangle = \epsilon a_i(x)$, for x in X, and ϵa_i is smooth, there exist open sets V_i with s in V_i and

$$|\langle a_i,s \rangle - \langle a_i,x \rangle| \leqslant \delta - |\langle a_i,s \rangle - \langle a_i,\alpha \rangle|,$$

for all x in V_i. Let $V = \cap V_i$. This is an open set in X containing s, and for x in V,

$$|\langle a_i,x \rangle - \langle a_i,\alpha \rangle|$$
$$\leqslant |\langle a_i,x \rangle - \langle a_i,s \rangle| + |\langle a_i,s \rangle - \langle a_i,\alpha \rangle| \leqslant \delta$$

as desired.

Now put a topology on K as follows. The space $X \times A^0$ carries the product topology. Identify K with the subset

$$K = \{(u,a^0): a^0 \in K_u\}$$

and give K the subset topology. The topology on the fibers gives each fiber the structure of a finite-dimensional topological vector space, and hence must coincide with the usual topology on finite-dimensional vector spaces.

One can recognize "algebraically smooth" sections of K by saying a map $\sigma: X \to K$ is an algebraically smooth section if $\pi\sigma = 1$ and for all a in $A(X)$, the map

$$\langle \sigma, a \rangle : X \to R, \quad \langle \sigma, a \rangle(p) = \langle \sigma(p), a \rangle$$

is a smooth map.

For an open set U in X define $\Gamma(U,K)$ to be the sheaf of algebraically smooth sections.

Lemma: The assignment $U \to \Gamma(U,K)$ defines a sheaf of C^∞ modules on X.

Proof: First check that $\Gamma(U,K)$ is a presheaf: that is, if V is an open set contained in U, check that there is a restriction homomorphism

$$\rho(V,U): \Gamma(U,K) \to \Gamma(V,K).$$

Certainly if σ is in $\Gamma(U,K)$, σ restricts to a section τ of K over V. The problem is to check that it is algebraically smooth.

Let u be in V and choose a neighborhood Y of u in V such that the closure of Y is contained in V. By Lemma 2.1, given a in $A(V)$ there exists b in $A(U)$ and $\langle \tau, a \rangle = \langle \sigma, b \rangle$ on Y. Thus $\langle \tau, a \rangle$ is smooth on a neighborhood of u in V, for any a, and hence τ is algebraically smooth.

The remaining sheaf theoretic properties are not difficult to verify. The C^∞ module structure comes from pointwise multiplication.

Proposition: Suppose K is constructed as above and has the property that for all p in X there exists a neighborhood U of p in X such that (i) $\Gamma(U,K)$ is a free $C^\infty(U)$ module on some generators $m_1,\ldots,m_k$; (ii) the elements $m_1(u),\ldots,m_k(u)$ form a basis for K_u, for all u in U; and (iii) there exist $a_1,\ldots,a_k$ in $A(U)$ such that $\langle m_i, a_j \rangle = \delta_{ij}$. Then K can be given a unique smooth structure making K a C^∞ vector bundle, with $\Gamma(\ ,K)$ coinciding with the usual sheaf of smooth functions.

Proof: Construct an atlas for K as follows. For z in Kp, choose a neighborhood U of p such that U satisfies properties (i)–(iii) above, and there is a diffeomorphism ϕ from U to an open set in R^r. Define a map

$$\rho: K|_u \to U \times R^k,$$
$$\rho(e) = (\pi e, r_1\pi(e),\ldots,r_k\pi(e)),$$

where the r_i are functions on U defined by the equation

$$e = \sum r_i(e) m_i(\pi e).$$

Thus $(\phi \times 1)\rho$ is a map from $K|_u$ to an open set in R^{r+k}.

First check that $(\phi \times 1)\rho$ is a homeomorphism. This will be a homeomorphism if ρ is. Notice that ρ has an inverse, explicitly the map sending

$$(u,r_1,\ldots,r_k) \mapsto \left(u, \sum r_i m_i(u)\right).$$

Check that ρ is continuous. Certainly π is continuous. For the functions r_i, consider the inverse image of an open set,

$$r_i^{-1}(r-\delta, r+\delta) = \{(u,\alpha):\ u \in U,\ \ \alpha \in K_u,\ \ |r_i(\alpha) - r| < \delta\}.$$

The open set

$$S = K \cap (U \times (r m_i(w) + B(a_i;\delta))),$$

where w is any element of U and a_i is the element of $A(U)$ with $\langle m_j, a_i \rangle = \delta_{ij}$, has the property that for (x,α) in S,

$$|\langle a_i, r m_i(w) - \alpha \rangle| = \left| \left\langle a_i, r m_i(w) - \sum r_i(\alpha) m_i(x) \right\rangle \right| = |r - r_i(\alpha)|.$$

Thus $S = r_i^{-1}(r-\delta, r+\delta)$ as desired.

Now check that ρ is an open map. Given an open set

$$S = K \cap (U \times (\alpha + B(a_1,\ldots,a_p;\delta))),$$

consider the set

$$\rho(S) = \left\{(u,r_1,\ldots,r_k): \sum r_i m_i(u) \in \alpha + B(a_1,\ldots,a_p;\delta)\right\}$$
$$= \left\{(u,r_1,\ldots,r_k): \left| \left\langle \alpha - \sum r_i m_i(u), a_j \right\rangle \right| < \delta, \text{ for all } j\right\}.$$

Suppose $\rho(S)$ is nonempty. Then there exists $(u, r_1,\ldots,r_k)$ with

$$\left| \left\langle \alpha - \sum r_i m_i(u), a_j \right\rangle \right| < \nu < \delta,$$

for all j. Each function $\langle \Sigma r_i m_i(w), a_j \rangle$ is a smooth function on U so there exists an open subset W of U such that its closure is compact, u is in W, and

$$|r_i| \, |\langle m_i, a_j \rangle(u) - \langle m_i, a_j \rangle(w)| < \delta - \nu/2k.$$

Now suppose that $(s_1,\ldots,s_k)$ satisfies

$$|r_i - s_i| \max |\langle m_i(w), a_j \rangle| < \delta - \nu/2k, \tag{1}$$

where the maximum is taken over all j and all w in W. Then

$$\left| \left\langle \alpha - \sum s_i m_i(w), a_j \right\rangle \right| \leqslant \left| \left\langle \alpha - \sum r_i m_i(u), a_j \right\rangle \right| + \left| \sum \langle r_i m_i(u) - s_i m_i(w), a_j \rangle \right|$$
$$\leqslant \nu + \sum |\langle r_i m_i(u) - r_i m_i(w) + r_i m_i(w) - s_i m_i(w), a_j \rangle|$$
$$\leqslant \nu + \sum |r_i| \, |\langle m_i(u) - m_i(w), a_j \rangle| + |r_i - s_i| \, |\langle m_i(w), a_j \rangle|.$$
$$\leqslant \delta.$$

Thus $Q = \{(w,s_1,\ldots,s_k):\ w \in W,\ (s_1,\ldots s_k) \text{ satisfies (1) above}\}$ is an open neighborhood of $(u,r_1,\ldots,r_k)$ in $U \times R^k$ as desired.

Now check that any two such charts are compatible. Suppose U' is another neighborhood of p with $\Gamma(U',K)$ free on generators $m_1',\ldots m_{k'}'$, as above, affording the homeomorphism

$$\rho': K|_U \to U \times R^{k'}.$$

Suppose that ϕ' is a diffeomorphism of U' with an open set in R^r. Let $W = U \cap U'$. Check that the composite $(\phi' \times 1)\rho'\rho^{-1} \times (\phi^{-1} \times 1)$ is a diffeomorphism from $(\phi \times 1)W \times R^k$ to $(\phi' \times 1)W \times R^k$.

This will be a diffeomorphism if and only if $\rho'\rho^{-1}$ is a diffeomorphism. Since the sets $\{m_i(u)\}$, $\{m_i'(u)\}$ are both bases for the vector space K_u, for all u in W, write

$$m_i(u) = \sum g_{ij}(u) m_i'(u).$$

Thus the g_{ij} are functions on U. Now $\rho'\rho^{-1}$ can be written explicitly as

$$\rho'\rho^{-1}(u,w) = (u,w'), \quad w_i' = \sum w_i g_{ij}(u).$$

It is enough therefore to check that the functions g_{ij} are smooth.

Consider m_i on $U \cap U'$. Pick u in $U \cap U'$ and neighborhoods V, W of u in U with the closure of V contained in W. Let χ be a function on U' with $\chi(x) = 1$ for x in V and $\chi(x) = 0$ on the complement of W in U'. Now χm_i is an element of $\Gamma(U \cap U', K)$, and $\chi m_i = m_i$ on U. Define an element m in $\Gamma(U', K)$ by setting $m = \chi m_i$ on $U \cap U'$ and $m = 0$ on $U' - W$. Since $\chi m_i = 0$ on $(U \cap U') \cap (U' \cap W)$, m is well defined. Then since $\Gamma(U', K)$ is a free $C^\infty(U')$ module generated by the m_i', write

$$m = \sum s_j m_j,$$

where the s_j are smooth functions on U'. Now observe that $s_j = g_{ij}$ on U. Hence the g_{ij} are smooth as desired.

Certainly with this smooth structure K is a vector bundle: the maps ρ provide local diffeomorphisms of $K|_U$ with $U \times R^k$ preserving the vector space structure on the fibers. That the smooth structure on K is unique depends on the following facts.

First, any atlas A of charts for K must contain an atlas of charts of the form $(K|_U, (\phi \times 1)\nu)$ where ν is a trivialization of $K|_U$ and (U, ϕ) is a chart on X as above.

Second, if $(K|_{U'}, (\phi' \times 1)\nu')$ is a chart of the above form in a second atlas A', and $U \cap U'$ is not empty, the assumption that $\Gamma(\ , K)$ is the sheaf of smooth functions allows one to conclude that the charts are in fact compatible. Hence A and A' must give rise to the same differentiable structure on K.

Examples.

(i) The odd tangent space $T(X,\mathcal{A})_1$. For p in X define

$$T_p(X,\mathcal{A})_1 = \{\alpha \in \mathcal{A}^0(p): \\ \alpha(ab) = \alpha(a)\epsilon b(p) + (-1)^{|a|}\epsilon a(p)\alpha(b)\}.$$

If U is an open set over which $\mathcal{A}$ has a trivialization τ, then

$$\tau(U): \mathcal{A}(U) \to C^\infty(U) \otimes \Lambda\langle\theta_1,\ldots,\theta_s\rangle$$

is an algebra isomorphism, where $\{\theta_i\}$ is a basis for R^s. Thus, using example (iii) of Sec. III

$$\tau(U)(T_p(X,\mathcal{A})_1) = p \otimes \Lambda R^{s'}$$

and $\Gamma(U, T(X,\mathcal{A})_1)$ is isomorphic to $C^\infty(U) \otimes (R^s)'$. This is clearly locally a free C^∞ module on generators $1 \otimes \lambda_i$, where the $\{\lambda_i\}$ is the basis of $R^{s\prime}$ dual to the basis $\{\theta_i\}$. With this choice of generators conditions (ii) and (iii) are clearly satisfied.

(ii) Suppose $\mathcal{A}$ is the sheaf $C^\infty \otimes \Lambda R^s$. Define

$$K_p = p \otimes \Lambda(R^s)'.$$

The sheaf $\Gamma(\ , K)$ is in fact globally free.

(iii) Now get an analog for example (ii) in the event that $\mathcal{A}$ is not globally trivial. Define

$$G_p(0) = Rp, \quad G_p(1) = Rp + Tp(X,\mathcal{A})_1,$$

$$G_p(k) = \{\alpha \in \mathcal{A}^0(p): \\ \Delta\alpha - \alpha \otimes p - p \otimes \alpha \in G_p(k-1) \otimes G_p(k-1)\}.$$

Using induction one can show that $G_p(k)$ is a subcoalgebra of $G_p(l)$ for all $k < l$. Define

$$G_p = \cup G_p(k).$$

Notice that if U is an open set over which $\mathcal{A}$ is trivial, then $\Gamma(U,G)$ is C^∞ free. The necessary verifications proceed as for $T(X,\mathcal{A})_1$.

Proposition: (i) G_p is isomorphic to $\Lambda T_p(X,\mathcal{A})_1$ as coalgebras.

(ii) G is isomorphic to $\Lambda T(X,\mathcal{A})_1$ as vector bundles.

Proof: Pick an open set U with $\mathcal{A}$ trivial over U. A trivialization

$$\tau: \mathcal{A}(U) \to C^\infty(U) \otimes \Lambda R^s$$

provides a coalgebra isomorphism

$$\tau^0: (C^\infty(U) \otimes \Lambda R^s)^0 \to \mathcal{A}(U)^0.$$

By inspection, τ^0 sends $p \otimes (R^s)'$ isomorphically onto $T_p(X,\mathcal{A})_1$ and $p \otimes \Lambda(R^s)'$ isomorphically onto G_p. Hence (i).

Now let $\boldsymbol{\tau}^0$ denote the restriction of τ^0 to $p \otimes R^{s\prime}$. Consider the map

$$\tau^0 \Lambda(\boldsymbol{\tau}^{0-1}): \Lambda T_p(X,\mathcal{A})_1 \to G_p.$$

This map restricts to the identity map on $T_p(X,\mathcal{A})_1$, and is a coalgebra isomorphism. Thus it is independent of choice of trivialization τ used to define it (see the examples in Sec. III), and one can define a bundle map

$$\xi: \Lambda T(X,\mathcal{A})_1 \to G,$$

$$\xi = \tau^0 \Lambda(\boldsymbol{\tau}^{0-1}) \quad \text{on } \Lambda T_p(X,\mathcal{A})_1$$

for some trivialization τ in a neighborhood of p. Thus ξ provides the bundle isomorphism required in (ii).

V. THE MAIN THEOREM

Lemma: Let G' be the dual bundle of G. There is a canonical isomorphism of sheaves of algebras

$$\sigma: \mathcal{A} \to \Gamma(\ , G').$$

Proof: Let U be an open set in $\mathcal{A}$. Define

$$\sigma(U): \mathcal{A}(U) \to \Gamma(U, G')$$

by setting

$$\langle \sigma(U)a(u), g \rangle = \langle a, g(u) \rangle,$$

for a in $\mathcal{A}(U)$, $g(u)$ in G_u. If g is a smooth section of G over U, then the function $\langle \sigma(U)a, g \rangle = \langle a, g \rangle$ from U to R is smooth, and hence $\sigma(U)a$ is a smooth section of G' over U. It is not hard to see that σ commutes with restriction maps appropriately.

Now check that σ is an isomorphism. It is enough to check that $\sigma(U_\alpha)$ is an isomorphism for a collection of open sets U_α which cover X. If U is a set over which $\mathcal{A}$ has a trivialization τ, then the following diagram commutes:

$$\begin{array}{ccc} \mathcal{A}(U) & \xrightarrow{\tau} & C^\infty(U) \otimes \Lambda\langle\theta_1,\ldots,\theta_s\rangle \\ \sigma(U) \downarrow & & \downarrow \eta \\ \Gamma(U,G') & \xrightarrow{\nu} & C^\infty(U) \otimes \Lambda\langle\lambda_1,\ldots,\lambda_s\rangle'. \end{array}$$

Here η is the map determined by the canonical identification of $\Lambda(\theta_1,...,\theta_s)$ and $\Lambda(\lambda_1,...,\lambda_s)'$. The choice of sections

$$\lambda_1^{i_1}\wedge\cdots\wedge\lambda_s^{i_s}:x\to x\otimes\lambda_1^{i_1}\wedge\cdots\wedge\lambda_s^{i_s}$$

fixes a C^∞ linear isomorphism

$$\nu':\Gamma(U,G)\to C^\infty(U)\otimes\Lambda\langle\lambda_1,...,\lambda_s\rangle.$$

Since ν' is an isomorphism, by choosing sections dual to $\lambda_1,...,\lambda_s$, one gets the C^∞ linear isomorphism

$$\nu:\Gamma(U,G)\to C^\infty(U)\otimes(\Lambda\langle\lambda_1,...,\lambda_s\rangle)'.$$

Thus $\sigma(U)$ is an isomorphism for all open sets U over which A is trivial. Since X is covered by such open sets, σ is an isomorphism of sheaves.

Multiplication in $\Gamma(U,G')$ is given by

$$\langle st(u),g\rangle=\sum(-1)^{\mu(i)}\langle s(u),g_{1i}\rangle\langle t(u),g_{2i}\rangle,$$

where g is in G_u, $\Delta g=\Sigma g_{1i}\otimes g_{zi}$, and $\mu(i)=|t(u)||g_{2i}|$. For a, b in $A(U)$,

$$\langle\sigma(U)(ab)(u),g\rangle=\langle ab,g\rangle=\sum(-1)^{|g_{1i}||b|}\langle a,g_{1i}\rangle\langle b,g_{zi}\rangle$$

$$=\langle\sigma(U)a\sigma(U)b,g\rangle.$$

Hence σ is a map of sheaves of algebras.

Lemma: There is a canonical isomorphism of vector bundles

$$G'\to\Lambda E,$$

where E is the dual bundle of $T(X,A)_1$.

Proof: This follows from the examples in Sec. IV. Since G is isomorphic to $\Lambda T(X,A)_1$ as a bundle of coalgebras via the canonical isomorphism ξ, the dual map

$$\xi':G'\to(\Lambda T(X,A)_1)'=\Lambda E$$

is a bundle isomorphism which preserves multiplication in the fibers.

Theorem 5.1: Let (X,A) be a graded manifold, and let $T(X,A)_1$ denote the odd tangent bundle of (X,A). Let E be the dual bundle to $T(X,A)_1$. Then there is a canonical isomorphism of sheaves of algebras

$$\eta:A\to\Gamma(\ ,\Lambda E).$$

Proof: The isomorphism of vector bundles in the examples in Sec. IV provides an isomorphism of sheaves

$$\kappa:\Gamma(\ ,G')\to\Gamma(\ ,\Lambda E),$$

which preserves the multiplicative structure. Then define $\eta=\kappa\sigma$.

Corollary: Let V denote the category of finite-dimensional real vector bundles over nice manifolds and let GM be the category of graded manifolds. Then there are functions

$$h:\mathrm{GM}\to V,\quad k:V\to\mathrm{GM},$$

which are adjoint functors.

Proof: Define h as follows. If (X,A) is a graded manifold define

$$h(X,A)=E,$$

where E is the bundle dual to the odd tangent bundle of (X,A). If f is a map $f:(Y,B)\to(X,A)$ of graded manifolds, then by definition, f is an algebra homomorphism $f:A(X)\to B(Y)$, and f gives rise to a map of coalgebras

$$f^0:B(Y)^0\to A(X)^0.$$

It is not hard to check that $T_y(Y,B)_1$ gets sent to $T_{f^0(y)}(X,A)_1$, and thus gives a map of vector bundles

$$Tf_1:T(Y,B)_1\to T(X,A)_1.$$

Define $hf=Tf_1$. Thus h is a covariant functor.

Define k by setting

$$kF=(S,\Gamma(\ ,\Lambda F')),$$

where F is a vector bundle over a manifold S and F' is its dual bundle. If H is another vector bundle over a manifold T, and

$$\chi:F\to H$$

is a vector bundle map, χ induces a vector bundle map

$$\Lambda\chi:\Lambda F\to\Lambda H.$$

Define $k\chi$ by setting

$$k\chi:\Gamma(T,\Lambda H')\to\Gamma(S,\Lambda F'),$$

$$\langle k\chi(\sigma)(s),f\rangle=\langle\sigma(\beta(s)),\ \chi(f)\rangle,$$

where s is in S, f is in ΛF, and σ is in $\Gamma(T,\Lambda H')$.

It is not hard to check that

$$hk(F)=F$$

for a vector bundle F, and if (X,A) is a graded manifold the canonical isomorphism

$$\eta(X):A(X)\to\Gamma(X,\Lambda E)$$

of Theorem 5.1 provides a natural transformation from kh to the identity.

[1]B. Kostant, "Graded manifolds, graded Lie theory and prequantization," in *Differential Geometric Methods in Mathematical Physics, Lecture Notes in Mathematics*, Vol. 570 (Springer, Berlin, 1977), pp. 177–306.
[2]F. Berezin and G. Kac, Math. USSR Sb. **11**, 311 (1970).
[3]M. Batchelor, Trans. Math. Soc. **253**, 329 (1979).
[4]M. Sweedler, *Hopf Algebras* (Benjamin, New York, 1969).

On the space-times admitting a synchronization of constant curvature

Carlos Bona
Departament de Física Teórica, Universidad de Palma de Mallorca, Palma de Mallorca, Spain
Bartolomé Coll
Chaire de Physique Mathématique, Collège de France, Paris, France

(Received 17 July 1984; accepted for publication 16 November 1984)

A synchronization S on the space-time is a foliation by spacelike hypersurfaces. We study here the vector fields, tangent to S, which are Killing fields for the induced metric on every instant of S but which are not necessarily Killing fields of the whole space-time metric; they are called S-Killing vector fields. We analyze the multiplicity of the maximal symmetry or complete integrability case, that is the case for which the space-times admit a synchronization S with the maximum number of S-Killing vector fields. In particular, the important case where S is umbilical is treated in detail.

I. INTRODUCTION

The symmetries defined by Killing vector fields on the space-time of the relativity theory have been thoroughly studied. But, in spite of their great interest, these symmetries often imply too drastic restrictions on space-time and are not, therefore, always welcome.

The hyperbolic geometry of space-time is seen by any system of observers as a Riemannian three-dimensional time-dependent geometry: the geometry of every instant of the synchronization defined by the $t =$ const hypersurfaces. The physical situations of interest are then often related directly to the symmetries of this evolving three-dimensional geometry rather than to the symmetries of the four-dimensional space-time as a whole.

The object of this paper is to study these kinds of symmetries. More precisely, we are interested in vector fields of space-time which are Killing fields for the Riemannian metric induced on every instant of a synchronization S but are not necessarily Killing fields of the hyperbolic four-dimensional metric. These vector fields will be called, in short, *S-Killing vector fields.*

The S-Killing vector fields describe adequately spatial symmetries of the gravitational field associated to special matter distributions, corresponding to compact objects as well as to the continuous medium related to cosmological models. These applications would suffice to justify the study of such vector fields, but their knowledge is also essential in other domains of relativity theory: for example, the integration (analytical or numerical) of the Einstein constraint equations (and then of the Einstein evolution equations),[1] the analysis, from the Cauchy data, of the dimension of the isometry group admitted by the space-time,[2] the existence of rigid motions in relativity,[3] and, in general, the study of the integrability of the differential systems whose associated homogeneous system is precisely that of the Killing equations.

In this paper we shall study in detail the "maximal symmetry" case, that is the case in which the space-times admit a synchronization S with the maximum number of S-Killing vector fields. It follows that the induced metric on every instant $t =$ const of the synchronization is of constant curvature. We shall call them, for brevity, *synchronizations of constant curvature.*

The existence of S-Killing vector fields which are not Killing fields seems to have been first considered only very recently, in the works of Krasinski[4] and Collins and Szafron.[5]

The paper by Krasinski begins with the study of spherically symmetric synchronizations, but the second half of his work is devoted to the synchronizations of constant curvature. He considers some space-times with such synchronizations by taking two inequivalent extensions of the Friedmann–Robertson–Walker models, based on the two coordinate systems commonly used to deal with these metrics. Here, we are interested directly in the whole class of those space-times, and we think that a good understanding of their multiplicity needs a good geometric formulation. We will show that the evolution formalism of general relativity is perfectly adapted to this goal.

The papers by Collins and Szafron analyze the problem in a fairly general way, by considering the more general case of conformally flat synchronizations. Nevertheless, they use in an essential way the assumption that the matter content is a perfect fluid whose flow lines form a geodesic congruence orthogonal to the synchronization. In our work we do not make any hypothesis on the matter content of the space-time.

In Sec. II we make precise the notion of synchronization; the geometric elements associated with it are presented and the evolution formalism of the Einstein equations is recalled. In Sec. III, we obtain the equations defining the S-Killing vector fields and give the necessary and sufficient conditions for an S-Killing vector field to be a Killing vector field.

The general study of the constant curvature synchronizations is carried out in Sec. IV. We introduce the so-called standard and conform-standard forms of the induced metric and we obtain the necessary and sufficient conditions for these forms to be "normal" for the space-time metric. This leads us finally to consider, in Sec. V, the important case of the umbilical synchronizations.

II. SYNCHRONIZATIONS

All the results presented here (except those related to the Petrov–Bel types) are qualitatively independent of the

 0022-2488/85/071583-09$02.50

dimension of space-time. For this reason, and also because it can throw light on many interesting questions related to other dimensions,[6] we shall here generically call *space-time* a manifold V_{n+1} sufficiently differentiable and of dimension $n+1$.

Tensor elements defined on V_{n+1} will be noted with a caret in order to distinguish them from those, defined on submanifolds, which will be introduced below. Also, throughout this work, we shall note in general with the same letter tensors and cotensors associated with the metric; when a confusion is possible, tensors will be distinguished with an asterisk "*."

We shall suppose that the space-time is endowed with a Lorentzian structure $(V_{n+1},\hat{g})$ (see Ref. 7) and, in order to make precise some expressions, we will choose the metric $\hat{g}$ to have signature $-(n-1)$, although the results will be given in a signature-independent form.

In $(V_{n+1},\hat{g})$ we shall call an *energy tensor* $\hat{T}$ the symmetric two-cotensor field

$$\hat{T} = \mathrm{Ric}(\hat{g}) - \tfrac{1}{2}R(\hat{g})\cdot\hat{g}, \tag{1}$$

where $\mathrm{Ric}(\hat{g})$ is the Ricci tensor associated to $\hat{g}$ and $R(\hat{g}) \equiv \mathrm{tr}\,\mathrm{Ric}(\hat{g})$, "tr" being the trace operator relative to $\hat{g}$. By virtue of the Bianchi identities, $\hat{T}$ is divergence-free,[8] say $\delta\hat{T} = 0$.

As usual in general relativity, we shall suppose that the energy tensor $\hat{T}$ is known *a priori*. Equations (1) are then differential equations on $\hat{g}$, and will also be called the *Einstein equations* of $(V_{n+1},\hat{g})$.

In a domain Ω of the space-time $(V_{n+1},\hat{g})$, a *synchronization* S is, by definition, a foliation of Ω by spatial hypersurfaces. Every hypersurface Σ of the foliation is an *instant* of the synchronization S and two points of Ω are said to be *synchronous* relatively to S if they belong to the same instant Σ of S (see Ref. 9).

To every synchronization S is associated the unitary one-form $\hat{n}$, normal at any point to the corresponding instant of S; with our signature, we have

$$\hat{g}^*(\hat{n},\hat{n}) = 1. \tag{2}$$

Conversely, any *locally integrable* unitary one-form $\hat{n}$, that is such that[10]

$$\hat{n} \wedge d\hat{n} = 0, \tag{3}$$

defines locally a synchronization S.

From (3) and (2) we have

$$\hat{n} = |d\phi|^{-1}\, d\phi, \tag{4}$$

with $|d\phi|^2 \equiv \hat{g}^*(d\phi,d\phi)$. A local representation of (the different instants of) S is thus given by the local equation

$$\phi(x) = t,$$

for different t. The parametrization t defines the time induced on S by the representation ϕ.

A vector field $\hat{s}^*$ is said to leave the synchronization S invariant if the local pseudogroup of transformations defined by $\hat{s}^*$ transforms hypersurfaces of S into hypersurfaces of S. The set of vector fields which leave a synchronization S invariant is a Lie algebra, which we shall note by A_S. It contains the Lie algebra T_S of the vector fields tangent to the synchronization S (which leave invariant separately every instant of S).

A *motion* Γ is defined, in the domain Ω, by a congruence of timelike curves, called *trajectories*. Let us note by $A(\Gamma)$ the Lie algebra of vector fields tangent to Γ and define $A_S(\Gamma)$ by

$$A_S(\Gamma) \equiv A_S \cap A(\Gamma).$$

The elements of $A_S(\Gamma)$ are the vector fields tangent to Γ which leave invariant the synchronization S.

In the present context the pair $\{\Gamma,S\}$, of a motion Γ and a synchronization S, will be called an *evolution basis* on Ω. The evolution bases so defined are the structural elements of the evolution formalisms[11] of general relativity.

If the vector field $\hat{s}^*$ is an element of $A_S(\Gamma)$ and ϕ is a local representation of S, we have $L(\hat{s}^*)\phi = \Phi$, where Φ is a (nonzero) constant function on every instant of S. Thus, the relation

$$L(\hat{s}^*)\phi = 1 \tag{5}$$

defines a one-to-one correspondence between the elements $\hat{s}^*$ of $A_S(\Gamma)$ and the representations ϕ of S. Every pair $\{\hat{s}^*,\phi\}$ verifying the equation (5) will be called a *local representation* of the evolution basis $\{\Gamma,S\}$.

If $\{\hat{s}^*,\phi\}$ and $\{\hat{r}^*,\psi\}$ are two local representations of $\{\Gamma,S\}$ with domains U and V respectively, $U\cup V \subset \Omega$ and $U\cap V \neq 0$, then there exists f: $\mathbb{R}\to\mathbb{R}$ such that, in $U\cap V$,

$$\psi = f\circ\phi, \quad \hat{r}^* = (f'_\phi)^{-1}\hat{s}^*, \tag{6}$$

where $f'_\phi \equiv df\circ\phi$.

To every evolution basis $\{\Gamma,S\}$ we associate in a natural manner, the class $C(\Gamma,S)$ of local charts $\{x^\alpha\}$ simultaneously adapted to Γ and to S; that is, such that $x^i = $ const on every trajectory of Γ and $x^0 = $ const on every instant of S. The condition (5), verified by any local representation $\{\hat{s}^*,\phi\}$, is necessary and sufficient for the existence of elements of $C(\Gamma,S)$ such that $\phi(x) \equiv x^0$ and $\hat{s}^* \equiv \partial/\partial x^0$.

The evolution formalism of general relativity is relative in the sense that the geometric objects and space-time itself do not appear in a covariant way but in reference to (or seen by) an evolution basis $\{\Gamma,S\}$.

On every instant Σ of the synchronization S, every (covariant) tensor of $(V_{n+1},\hat{g})$ induces a unique (covariant) tensor of the same rank; in particular, the tensor g induced by $\hat{g}$ endows every instant Σ with a Riemannian structure (Σ,g). The intrinsic Lorentzian geometry of the space-time V_{n+1} defined by the metric $\hat{g}$ is replaced in the evolution formalism by a variable n-dimensional Riemannian geometry (Σ,g) and the geometric objects considered are those defined on that geometry in evolution.

Let $\hat{n}$ by the unit normal one-form of the synchronization S. It is easy to see that, on every instant Σ of S, any p-tensor $\hat{Q}$ (tensor of rank p) of $(V_{n+1},\hat{g})$ is biunivocally characterized by the following set of 2^p tensors of (Σ,g): the $\binom{p}{A}$ A-tensors $(A = 0,1,\ldots,p)$ induced by the A-tensors of $(V_{n+1},\hat{g})$ obtained by taking all the possible $p-A$ interior products of $\hat{n}^*$ with $\hat{Q}$. Such a set will be called the *Σ characterization* of $\hat{Q}$ (see Ref. 12).

Thus, the Σ characterization of a vector $\hat{s}^*$ is the set $\{\sigma;s\}$, where the scalar σ is the interior product of $\hat{n}^*$ and $\hat{s}$ (see Ref. 13), $\sigma \equiv i(\hat{n}^*)\hat{s}$, and s is the covector induced by $\hat{s}$.

The strict elements of the Σ characterization of a second-rank symmetric tensor $\hat{A}$ are $\{\alpha;a;A\}$ where $\alpha \equiv i^2(\hat{n}^*)\hat{A}$ and a and A are, respectively, the vector and the second-rank (symmetric) tensor induced by $i(\hat{n}^*)\hat{A}$ and $\hat{A}$ (see Ref. 14). In particular, the Σ characterizations of $\hat{n}$ and $\hat{g}$ are, respectively, $\{1;0\}$ and $\{1;0;g\}$.

It is also interesting to consider the tensor $2\hat{K} \equiv L(\hat{n}^*)\hat{g}$. As $\hat{n}$ is unitary, one has $i^2(\hat{n}^*)L(\hat{n}^*)\hat{g} = 0$ and it follows from Eq. (4) that

$$i(\hat{n}^*)L(\hat{n}^*)\hat{g} = 2\bot(\hat{n})d\ln\sigma,$$

where $\bot(\hat{n}) \equiv \mathrm{Id} - \hat{n}\otimes\hat{n}$ is the orthogonal (to $\hat{n}$) projector and $\sigma \equiv |d\phi|^{-1}$ is an integration factor of $\hat{n}$; then, the strict elements of the Σ characterization of $\hat{K}$ are $\{0; -d\sigma;K\}$, where $d\sigma$ is now the differential of σ on every instant Σ and K is the cotensor induced by $\hat{K}$.

The n-dimensional one-form $d\sigma$ measures the non-Gaussian character of the synchronization S ($d\sigma = 0 \Leftrightarrow S$ is Gaussian). Nevertheless, the symmetric two-cotensor K depends only on the individual instant Σ: K is called the *second fundamental form*, or *extrinsic curvature*, of Σ.

In our domain Ω, let us take an evolution basis $\{\Gamma,S\}$. On every instant Σ of S, let $\{\tau;t;T\}$ and $\{\rho;r;R\}$ be the (strict components of the) Σ characterizations of the energy tensor $\hat{T}$ of the medium and of the Ricci tensor $\mathrm{Ric}(\hat{g})$, respectively. We have, by definition,

$$\tau \equiv i^2(\hat{n}^*)\hat{T}, \quad t \equiv \mathrm{induc}_\Sigma\, i(\hat{n}^*)\hat{T}, \quad T \equiv \mathrm{induc}_\Sigma\, \hat{T}, \tag{7}$$

and analogous relations for ρ, r, and R.

The quantities τ, t, and T represent, respectively, the energy density, the momentum density, and the stress tensor of the medium relatively to the synchronization S. The Σ characterization of the Einstein equations is then

$$\begin{aligned} \rho &= (1/(n-1))((n-2)\tau - \mathrm{tr}\,T), \\ r &= t, \\ R &= T - (1/(n-1))(\tau + \mathrm{tr}\,T)g. \end{aligned} \tag{8}$$

Now, let us consider any local representation $\{\hat{s}^*,\phi\}$ of $\{\Gamma,S\}$. In every local chart $\{x^\alpha\}$, $(\alpha = 0,1,\ldots,n)$, of $C(\Gamma,S)$ adapted to the representation $\{\hat{s}^*,\phi\}$ one has $\phi(x) \equiv x^0$, and thus the time t induced by the representation ϕ is simply given by the timelike coordinate $t = x^0$; in addition, as we have $\hat{s}^* \equiv \partial/\partial x^0$, t is also a canonical parameter of the integral curves of $\hat{s}^*$.

Denoting by $\{\sigma;s\}$ the Σ characterization of $\hat{s}^*$ (see Ref. 14), the extrinsic curvature K may be written

$$K = (1/2\sigma)\,(\partial_t g - L(\hat{s}^*)g), \tag{9}$$

where ∂_t is the derivation operator with respect to the parameter t. From the local coordinate expression of $\mathrm{Ric}(\hat{g})$ as a function of $\hat{g}$ and its partial derivatives, we obtain, for its Σ characterization $\{\rho,r,R\}$ (see Ref. 15)

$$\begin{aligned} \rho &= -\mathrm{tr}\,K^2 + (1/\sigma)(\Delta\sigma - \partial_t\,\mathrm{tr}\,K + L(s^*)\mathrm{tr}\,K), \\ r &= -\delta K - d\,\mathrm{tr}\,K, \\ R &= \mathrm{Ric}(g) + 2S - (1/\sigma)(\nabla\, d\sigma + \partial_t K - L(s^*)K), \end{aligned} \tag{10}$$

where $K^2 \equiv K \times K$ and $\times$ is the cross product (contraction over the adjacent indices of the tensorial product). All the operators, except ∂_t, are defined on every instant Σ of S and the de Rham Laplacian Δ, the covariant derivative ∇, the trace tr and the divergence δ are all taken with respect to the induced metric g on every instant Σ. We have noted by S the covariant symmetric quadratic form in K:

$$S \equiv K \times K - \tfrac{1}{2}\mathrm{tr}\,K\cdot K \tag{11}$$

If we consider the energy tensor $\{\tau;t;T\}$ as a known source and the quantities g and K as unknowns, the Einstein equations (8) may be written, using (9) and (10), in the equivalent form

$$\begin{aligned} &C_1: \quad \mathrm{tr}\,K^2 - \mathrm{tr}^2 K + \mathrm{tr}\,\mathrm{Ric}(g) = -2\tau, \\ &C_2: \quad \delta K + d\,\mathrm{tr}\,K = -t, \\ &E_1: \quad \partial_t g = 2\sigma K + L(s^*)g, \\ &E_2: \quad \partial_t K = \sigma(\mathrm{Ric}(g) + 2S - R) - \nabla\, d\sigma + L(s^*)K, \end{aligned} \tag{12}$$

where R is now the quantity, depending of τ and T, defined by (8).

These are the so-called *Einstein equations of the evolution* (or 3 + 1) *formalism*. In the Gaussian evolution bases ($\sigma = \sigma(t)$) they were given by Lichnerowicz[16,17] and, under the general form (12), by Choquet-Bruhat.[18,19] These equations constitute the starting point for the Hamiltonian formulation of general relativity obtained, in an important work, by Arnowitt, Deser, and Misner[20]; due to this fact, the Einstein equations of the evolution formalism (12) are sometimes called ADM equations. We shall reserve this last appellation only to situations in which the Hamiltonian formalism is explicitly concerned.

III. *S*-KILLING VECTOR FIELDS

Let us take an evolution basis $\{\Gamma,S\}$ in Ω and consider a local representation $\{\hat{s}^*,\phi\}$ of it, with domain $U \subset \Omega$. Let us define in U the local operator on one-forms

$$\Pi_{|U} \equiv \mathrm{Id} - d\phi\otimes\hat{s}^*. \tag{13}$$

If $\{\hat{r}^*,\psi\}$ is another local representation of $\{\Gamma,S\}$ with domain $V \subset \Omega$, $U \cap V \neq 0$, and $\Pi_{|V}$ is the corresponding local operator, it follows from the relation (6) that $(\Pi_{|U})_{|U\cap V} = (\Pi_{|V})_{|U\cap V}$. Thus, the $\Pi_{|U}$'s define in Ω a unique operator Π which depends only on the evolution basis $\{\Gamma,S\}$, and not on its local representations.

From (13) and (5), it is obvious that $L(\hat{s}^*)\Pi_{|U} = 0$, and then

$$L(\hat{r}^*)\Pi = 0, \quad \forall \hat{r}^* \in A_S(\Gamma). \tag{14}$$

The operator Π can be extended, in the usual way, to the cotensor fields. In particular, for the two-cotensors $\hat{Q}$ one has $\Pi(\hat{Q}) = \Pi \times \hat{Q} \times \overline{\Pi}$, where $\overline{\Pi}$ stands for the local adjoint of Π: $\overline{\Pi} \equiv \mathrm{Id} - \hat{s}^*\otimes d\phi$.

Let $\hat{Q}$ and Q be, respectively, a cotensor field on V_{n+1} and its induced part on Σ. With respect to any orthonormal frame containing $\hat{n}$ (the unit normal to Σ), one has

$$Q = \bot(\hat{n})\hat{Q}, \tag{15}$$

where $\bot(\hat{n}) \equiv \mathrm{Id} - \hat{n}\otimes\hat{n}^*$ is the projector orthogonal to $\hat{n}$. However, with respect to the natural frame associated to an arbitrary local chart of $C(\Gamma,S)$, Eq. (15) is no longer valid; instead of (15) we have

$$Q = \Pi(\hat{Q}), \tag{16}$$

where Π is the operator defined by (13) in the local representation attached to the given chart. It is evident that, when the motion Γ is normal to the synchronization S, we have $\Pi = \perp(\hat{n})$ and the relations (15) and (16) coincide.

Thus, the tensor field $\hat{g}_I$ of V_{n+1} which, in any local chart of $C(\Gamma,S)$, reduces to the induced metric g on any instant Σ of S, is given by

$$\hat{g}_I \equiv \Pi(\hat{g}). \tag{17}$$

For this reason, we shall call $\hat{g}_I$ the *induced metric tensor* of the evolution basis $\{\Gamma,S\}$ (see Ref. 21).

In the same line, it is natural to define, for example, the *rigid motions* Γ with respect to a synchronization S as the motions for which

$$L(\hat{r}^*)\hat{g}_I = 0, \quad \forall \hat{r}^* \in A_s(\Gamma). \tag{18}$$

This new notion of rigidity has been recently considered by us elsewhere.[3,9]

In the present work, we are interested in vector fields $\hat{m}^*$ which are Killing vector fields for the induced metric at any instant Σ of S. The condition for $\hat{m}^*$ to be tangent to Σ is obviously $i(\hat{m}^*)\hat{n} = 0$, but the condition of metric invariance is not $L(\hat{m}^*)\hat{g}_I = 0$; this is due to the fact that $\hat{m}^*$ is not contained in $A_S(\Gamma)$ so that $i(\hat{n}^*)L(\hat{m}^*)\hat{g}_I$ does not vanish identically. Thus, the condition for $\hat{m}^*$ to be a Killing field for the induced metric on every instant Σ is

$$\perp(\hat{n})L(\hat{m}^*)\hat{g}_I = 0. \tag{19}$$

In spite of the fact that $\hat{g}_I$ depends [by (17) and (13)] on the evolution basis $\{\Gamma,S\}$, Eq. (19) depends *only* on the synchronization S as it was to be expected. To see it, it is sufficient to observe that, like $\hat{g}_I$ and $\hat{g}$, the tensors $L(\hat{m}^*)\hat{g}_I$ and $L(\hat{m}^*)\hat{g}$ differ only by terms containing $d\phi$ because, $\hat{m}^*$ being tangent to S [$\Leftrightarrow L(\hat{m}^*)\phi = 0$], one has $L(\hat{m}^*)d\phi = 0$; the projector $\perp(\hat{n})$ cancels out that difference and (19) turns out to be equivalent to

$$\perp(\hat{n})L(\hat{m}^*)\hat{g} = 0. \tag{20}$$

Let us develop this expression. It follows from Eq. (4) that

$$L(\hat{m}^*)\hat{n} = L(\hat{m}^*)(|d\phi|^{-1}d\phi) = -(L(\hat{m}^*)\ln|d\phi|)\cdot\hat{n}, \tag{21}$$

and we have also

$$i(\hat{n}^*)L(\hat{m}^*)\hat{g} = L(\hat{m}^*)\hat{n} - \hat{g}([\hat{m}^*,\hat{n}^*]), \tag{22}$$

$$\begin{aligned} i^2(\hat{n}^*)L(\hat{m}^*)\hat{g} &= |d\phi|^{-2}i^2(d^*\phi)L(\hat{m}^*)\hat{g} \\ &= -|d\phi|^{-2}i^2(d\phi)L(\hat{m}^*)\hat{g} \\ &= -2L(\hat{m}^*)\ln|d\phi|, \end{aligned} \tag{23}$$

where we have used the relation

$$\{L(\hat{m}^*)\hat{g}\}^* = -L(\hat{m}^*)\hat{g}^*. \tag{24}$$

If we substitute now Eqs. (21)–(23) in the development of (20), we obtain

$$L(\hat{m}^*)\hat{g} + \hat{n}\otimes[\hat{m}^*,\hat{n}^*]_* + [\hat{m}^*,\hat{n}^*]_*\otimes\hat{n} = 0, \tag{25}$$

where we have noted by $[\hat{m}^*,\hat{n}^*]_* \equiv \hat{g}([\hat{m}^*,\hat{n}^*])$ the covariant form of the Lie bracket $[\hat{m}^*,\hat{n}^*]$. Allowing for (24), the contravariant form of (25) is

$$-L(\hat{m}^*)\hat{g}^* + \hat{n}^*\otimes[\hat{m}^*,\hat{n}^*] + [\hat{m}^*,\hat{n}^*]\otimes\hat{n}^* = 0, \tag{26}$$

which is nothing else but

$$L(\hat{m}^*)(\hat{g}^* - \hat{n}^*\otimes\hat{n}^*) = 0. \tag{26'}$$

In any local chart $\{x^\alpha\}$ adapted to S, that is so that the instants Σ of S are defined by the local equation $x^0 =$ const, $\hat{n}$ is given by $\hat{n}_\alpha = (\hat{g}^{00})^{-1/2}\delta^0_\alpha$ and the condition $i(\hat{m}^*)\hat{n} = 0$ by $\hat{m}^0 = 0$. The components g_{ij} $(i,j = 1,2,\ldots,n)$, of the induced metric are $g_{ij} = \hat{g}_{ij}$ and, as it is well known,[22] the inverse g^{ij} of g_{ij} is $g^{ij} = \hat{g}^{ij} - (\hat{g}^{00})^{-1}\hat{g}^{0i}\hat{g}^{0j}$. On every instant Σ, $\hat{m}^*$ induces a covector m with components $m_i = \hat{m}_i$, and the vector $m^i = g^{il}m_l$ is such that $m^i = \hat{m}^i$. With these elements, it is easy to verify that Eqs. (26) are strictly equivalent, on every instant, to $L(\hat{m}^*)g^* = 0$ and thus to

$$L(m^*)g = 0. \tag{27}$$

Definition 1: We shall call *S-Killing vector fields* the fields of V_{n+1} which are, on every instant of a synchronization S, Killing vector fields of the induced metric [that is, they verify Eq. (27)].

Gathering the previous results, one has the following proposition.

Proposition 1: The S-Killing vector fields $\hat{m}^*$ are determined by any one of the following equivalent differential systems:

$$L(\hat{m}^*)\hat{g} = [\hat{n},\hat{m}^*]_*\otimes\hat{n} + \hat{n}\otimes[\hat{n}^*,\hat{m}^*]_*, \tag{28a}$$

$$L(\hat{m}^*)(\hat{g}^* - \hat{n}^*\otimes\hat{n}^*) = 0, \tag{28b}$$

$$\perp(\hat{n})L(\hat{m}^*)\hat{g} = 0, \tag{28c}$$

$$\perp(\hat{n})L(\hat{m}^*)\hat{g}_I = 0, \tag{28d}$$

together with the condition $i(\hat{m}^*)\hat{n} = 0$.

As a corollary of (28), Proposition 2 follows immediately.

Proposition 2: The necessary and sufficient condition for an S-Killing vector field $\hat{m}^*$ to be a Killing vector field of $(V_{n+1},\hat{g})$ is that $\hat{m}^*$ commutes with the unit normal $\hat{n}^*$ to S:

$$[\hat{m}^*,\hat{n}^*] = 0. \tag{29}$$

It is interesting to translate relation (29) into the language of the evolution formalism. Let $\{\Gamma,S\}$ be an evolution basis and $\{\hat{s}^*,\phi\}$ a local representation of it, and let $\{\sigma;s\}$ and $\{0;m\}$ be, respectively, the Σ characterizations of $\hat{s}^*$ and $\hat{m}^*$; in every local chart of $C(\Gamma,S)$ adapted to $\{\hat{s}^*,\phi\}$, we have $\hat{n}_\alpha = \sigma\delta^0_\alpha, \hat{n}^\alpha = \sigma^{-1}(\delta^\alpha_0 - s^l\delta^\alpha_l), \hat{m}^0 = 0$, and $\hat{m}^l = m^l$. Using these expressions, the scalar term of the Σ characterization of (29), $i(\hat{n}^*)[\hat{m}^*,\hat{n}^*] = 0$, may be written $L(\hat{m}^*)\sigma = 0$, whereas the vectorial term gives $\partial_t m^* = [s^*,m^*]$ or, in covariant form, $\partial_t m = 2\sigma i(m^*)K - L(s^*)m$, where we have used the Einstein evolution equation E_1 given in (12). With these results, we have the following proposition.

Proposition 3: In the evolution formalism, the necessary and sufficient conditions for an S-Killing vector field $\hat{m} \sim \{0;m^*\}$ to be a Killing vector field of $(V_{n+1},\hat{g})$ are

$$L(m^*)\sigma = 0, \quad \partial_0 m^* = [s^*,m^*], \tag{30a}$$

or, in covariant form,

$$L(m^*)\sigma = 0, \quad \partial_0 m = 2\sigma i(m)K - L(s^*)m, \tag{30b}$$

where $\{\sigma;s\} \sim \hat{s}^*$ characterizes the local representation of the chosen evolution basis and K is the extrinsic curvature of the instants of S.

The converse of Proposition 2 is well known: any Killing vector field $\hat{m}^*$ of $(V_{n+1},\hat{g})$ tangent to S is an S-Killing vector field. Thus, it commutes with the unit normal $\hat{n}^*$ to S and then, on account of the general relation $L([\hat{m}_1^*,\hat{m}_2^*]) = [L(\hat{m}_1^*),L(\hat{m}_2^*)]$, it follows that the Lie bracket of two Killing vector fields tangent to S is another Killing vector field tangent to S.

The relations (28) being linear and homogeneous in $\hat{m}^*$, it is clear that, for a given synchronization S, the set of S-Killing vector fields has a $\mathbb{R}$-vector space structure. In addition, from the general relation $L(\lambda\hat{m}^*)\hat{g} = \lambda L(\hat{m}^*)\hat{g} + d\lambda \otimes \hat{m} + \hat{m} \otimes d\lambda$ and (28c) it is obvious that, $\hat{m}^*$ being an S-Killing field, $\lambda\hat{m}^*$ is also an S-Killing field iff $\perp(\hat{n})d\lambda = 0$. Thus, the S-Killing vector space has a module structure over the ring F_S of the constant functions on every instant of S. Finally, from the commutator relation and (28a) it follows that the S-Killing module has a Lie algebra structure.

If $\{0;m_1^*\}$ and $\{0;m_2^*\}$ are, respectively, the Σ characterizations of two Killing fields $\hat{m}_1^*$, and $\hat{m}_2^*$, tangent to S, the Σ characterization of the Lie bracket $[\hat{m}_1^*,\hat{m}_2^*]$ is then $\{0;[m_1^*,m_2^*]\}$. It is thus clear that, when Eq. (29) holds, the Killing and S-Killing Lie algebra structures are identical. This fact is the basis for a classification of the S-isometry groups according to their isometry subgroups.[23]

IV. SYNCHRONIZATIONS OF CONSTANT CURVATURE

In the preceding section, we have given a $(n+1)$-dimensional formulation of the S-Killing vector fields, which allows us to find easily their basic structures and is certainly helpful to solve many open problems, the most important one being, perhaps, that of the existence, in a given space-time, of synchronizations S admitting S-Killing vector fields.

Nevertheless, in what follows we shall suppose that our space-time admits such a synchronization and that we know it. In such a case, it is useful to consider S as the synchronization of an evolution basis $\{\Gamma,S\}$ and to work directly in the evolution formalism sketched out in Sec. II.

Moreover, we shall suppose that the synchronization S is completely integrable or, in other words, that S admits the maximum number of S-Killing vector fields. From their Definition 1, it is clear that every instant Σ of S, considered as a Riemannian manifold (Σ,g) admits then $n(n+1)/2$ Killing vector fields: every Σ is of constant curvature. We shall call for short synchronizations of constant curvature those for which every instant is of constant curvature.

Let S be our synchronization and let us take a normal evolution basis $\{\Gamma_N,S\}$, that is to say, such that Γ_N is the congruence of normal curves to every instant of S. For all the local representations $\{\hat{s}^*,\phi\}$ of $\{\Gamma_N,S\}$ we have $\hat{s} = |d\phi|^{-2}d\phi = |d\phi|^{-1}\hat{n}$ and the Σ characterization of $\hat{s}^*$ is of the form $\{\sigma;0\}$ with $\sigma \equiv i(\hat{s}^*)\hat{n} = |d\phi|^{-1}$. In any local chart $\{x^\alpha\}$ of $C(\Gamma_N,S)$ adapted to $\{\hat{s}^*,\phi\}$, these relations give $x^0 = t, \hat{g}^{0i} = 0, \sigma = (\hat{g}^{00})^{-1/2}$ and the metric in $(V_{n+1},\hat{g})$ may be written

$$\hat{g} = \sigma^2(x,t)dt \otimes dt + g_{ij}(x,t)dx^i \otimes dx^j, \tag{31}$$

where $g \equiv g_{ij}(x,t)dx^i \otimes dx^j$ is the induced metric on every instant Σ of local equation $t = \text{const}$.

The S-Killing vector fields $\hat{m}^*$ are evidently of the form $\{0,m^*\}$ with $m \equiv m_i(x,t)dx^i$ and such that $L(m^*)g = 0$. Also, the fact that g is, at every instant, of constant curvature is expressed by

$$R_{ij,kl} = \kappa(g_{ik}g_{jl} - g_{il}g_{jk}), \tag{32}$$

where $R_{ij,kl}$ are the components of the Riemann tensor $\mathrm{Riem}(g)$. The metrics (32) are conformally flat,

$$g = \alpha^2\eta, \tag{33}$$

and it is well known that, for every fixed t, there exist local coordinates $\{y^i\}$ ("Cartesian" for η) such that α is the standard function

$$\alpha(y) = 1/[1 + (\kappa/4)y^2], \tag{34}$$

with

$$y^2 \equiv y \circ y \equiv c_{ij}y^iy^j \tag{35}$$

and

$$\eta = c_{ij}\,dy^i \otimes dy^j, \tag{36}$$

where c_{ij} is the signature symbol of g:

$$c_{ij} = 0, \quad \text{if } i \neq j, \quad c_{ii} = e_i, \tag{37}$$

$e_i = \pm 1$ depending on the signature of g (see Ref. 24).

Because the metrics of same constant curvature and same signature are locally isometric, there exists, for each t, a diffeomorphism Y_t such that the reciprocal image of its inverse, X_t^*, carries g to its standard form

$$X_t^*: g_{ij}(x,t) \rightarrow \alpha^2(y,t)c_{ij}, \tag{38}$$

where $\alpha(y,t)$ is the standard function (34) with $\kappa = \kappa(t)$.

In the space-time, Y_t preserves the synchronization but transforms the timelike motion Γ_N into another, not normal, one. The local expression of the transformed motion Γ is given by the inverse X_t of Y_t, with local equation $x^i = x^i(y,t)$ and whose tangent vector field is[25]

$$s^i(x,t) \equiv \left(\frac{\partial x^i(y,t)}{\partial t}\right)_{|y=y(x,t)}. \tag{39}$$

In the new local chart of V_{n+1} defined by $\{y^0 = t, y^i = y^i(x,t)\}$, the other components of $\hat{g}$ are transformed to $\hat{g}_{0i}(y,t) = s_i(y,t)$, where $s \equiv s_i(y,t)dy^i$ is the covector associated to (39) by g, and to $\hat{g}_{00}(y,t) = \sigma^2(y,t) + s \cdot s(y,t)$, where $s \cdot s$ stands for the g-scalar product and, σ being a scalar, we have put $\sigma(y,t) \equiv \sigma(x(y,t),t)$. It follows that in every space-time $(V_{n+1},\hat{g})$ admitting a constant curvature synchronization, the metric can be written in the form

$$\hat{g} = (\sigma^2 + s \cdot s)dt \otimes dt + s \otimes dt + dt \otimes s + \alpha^2 c_{ij}\,dy^i \otimes dy^j. \tag{40}$$

The possibility of working in local charts in which the space-time dependence of the induced metric is condensed in a single scalar function, as in (40), offers a great computational advantage. For this reason, we give the following definition.

Definition 2: We shall call *standard charts of* V_n the charts $\{t,y^i\}$, adapted to S, in which the induced metric g takes the standard form $g_{ij} = \alpha^2 c_{ij}$, with α and c_{ij} given, respectively, by (34) and (37). Similarly, we shall call *conform-standard charts* those for which the induced metric g takes a form conformal to the standard one. The corresponding forms of the space-time metric $\hat{g}$ will also be called standard and conform-standard forms, respectively.

Let us note that, starting from the standard form (40), the motion associated to an arbitrary transformation Z_t, defined by $z^i = z^i(y,t)$, $y^i = y^i(z,t)$, is given, at every instant, by

$$v \sim v^i(y,t) \equiv \left(\frac{\partial y^i(z,t)}{\partial t}\right)_{|z = z(y,t)}. \tag{41}$$

The induced metric takes then the general form $g = g_{ij}(z,t) dz^i \otimes dz^j$, and the space-time metric may be expressed by

$$\hat{g} = (\sigma^2 + (s+v)\cdot(s+v))dt \otimes dt + (s+v) \otimes dt + dt \otimes (s+v) + g_{ij}\, dz^i \otimes dz^j, \tag{42}$$

where $\sigma^2 = \sigma^2(z,t) \equiv \sigma^2(y(z,t),t)$, and $s \sim s_i(z,t)$ and $v \sim v_i(z,t)$ being, respectively, the transformed vector fields of $s_i(y,t)$ and of that defined by (41).

The n-dimensional conformal group C $(n>2)$ is constituted by the isometries of the flat metric $\eta \sim c_{ij}$,

$$z^i = y^i + r^i, \quad z^i = \Lambda^i{}_l y^l, \quad \Lambda^i{}_l \Lambda^j{}_m c_{ij} = c_{lm}; \tag{43}$$

together with the dilatations,

$$z^i = \lambda y^i; \tag{44}$$

and the acceleration transformations

$$z^i = (y^i + y^2 a^i)/H(a,y), \tag{45}$$

where

$$H(a,y) \equiv 1 + 2a\circ y + a^2 y^2; \tag{46}$$

and, according to (35), we have noted $a\circ y \equiv c_{ij} a^i y^j$, $a^2 \equiv a\circ a$. Thus, the general transformation of the conformal group may be written as

$$z^i = [\lambda \Lambda^i{}_l (y^l + r^l) + \lambda^2 (y+r)^2 a^i]/H(\lambda a\circ\Lambda, y+r), \tag{47}$$

its inverse being

$$y^i = -r^i + \overline{\Lambda}^i{}_l (z^l - z^2 a^l)/\lambda H(-a,z), \tag{48}$$

with $\overline{\Lambda} \equiv \Lambda^{-1}$, $(a\circ\Lambda)_l \equiv c_{ij} a^i \Lambda^j{}_l$, and $H(\ ,\ )$ is the function defined by (46).

The Lie algebra A_C of C is represented by[26]

$$p^i_{(a)} = \delta^i{}_a, \quad m^i_{(ab)} = \delta^i_a y_b - \delta^i_b y_a, \qquad d^i = y^i, \quad a^i_{(a)} = y^2 \delta^i_a - 2y^i y_a, \tag{49}$$

with $y_l \equiv c_{ls} y^s$. The general element of A_C is then of the form

$$s = A^{(a)} p_{(a)} + B^{(ab)} m_{(ab)} + Cd + D^{(a)} a_{(a)}, \tag{50}$$

where $A^{(a)}$, $B^{(ab)}$, C, and $D^{(a)}$ are arbitrary constants.

A general transformation (47) of C transforms the metric g from its standard form

$$g_{ij} = \alpha^2(y) c_{ij}, \tag{51}$$

with $\alpha(y)$ given by (34), to the conform-standard form

$$g_{ij} = \rho^2(z) c_{ij}, \tag{52}$$

with

$$\rho(z) = \theta/[H(-c,z) + (\kappa/4)\theta^2 x^2], \tag{53}$$

where

$$\theta \equiv \alpha(r)/\lambda, \quad c^l \equiv a^l + (\kappa/4)\theta \Lambda^l{}_s r^s. \tag{54}$$

The transformed form (52) is yet a standard form iff $\theta = 1$ and $c^l = 0$, that is, iff

$$\lambda = \alpha(r), \quad a^l = -(\kappa/4)\Lambda^l{}_s r^s. \tag{55}$$

Thus, the g-isometry group I_g is defined by

$$z^i = \frac{\alpha(r)\Lambda^i{}_l((y+r)^l - (\kappa/4)\alpha(r)(y+r)^2 r^l)}{H(-(\kappa/4)\alpha(r)r, y+r)}, \tag{56}$$

the inverse transformation being given by

$$y^i = -r^i + [\overline{\Lambda}^i{}_l z^l + (\kappa/4) z^2 r^i]/\alpha(r) H((\kappa/4)\Lambda\circ r, z). \tag{57}$$

It follows that the Lie algebra I_g of I_g may be represented by $m^i_{(ab)}$ and

$$t^i{}_{(a)} \equiv p^i_{(a)} - (\kappa/4) a^i_{(a)}, \tag{58}$$

where $p^i_{(a)}$, $m^i_{(ab)}$, and $a^i_{(a)}$ are given by (49). The g-Killing fields $t^i{}_{(a)}$ and their corresponding finite transformations [obtained from (56) for $\Lambda^i{}_j = \delta^i{}_j$] are called *quasitranslations*.

We have now all the elements needed to discuss the "generality" of the standard and conform-standard forms of $\hat{g}$. Evaluating (41) for the general transformation (47) and taking into account the definitions (49), we obtain the following result.

Proposition 4: The motion $v(y,t)$ associated, in a standard chart, to a general conformal transformation with time-dependent parameters $(r^a, \Lambda^a{}_l, \lambda, a^a)$ is of the form

$$v(y,t) = A^{(a)} p_{(a)} + B^{(ab)} m_{(ab)} + Cd + D^{(a)} a_{(a)}, \tag{59}$$

where the coefficients are given by

$$\begin{aligned} D^{(a)} &= -\lambda \dot{a}_l \Lambda^{la}, \\ C &= -(\dot{\lambda}/\lambda + 2D^{(s)} r_{(s)}), \\ B^{(ab)} &= \tfrac{1}{2}(\dot{\overline{\Lambda}}{}^a{}_l \Lambda^{lb} + 2(D \wedge r)^{ab}), \\ A^{(a)} &= -\dot{r}^a + 2B^{(as)} r_s + Cr^a + 2D^{(s)} r_s r^a - r^2 D^{(a)}. \end{aligned} \tag{60}$$

Evaluating now (60) for the values (55) of the group parameters, a straightforward calculation shows the following.

Proposition 5: The motion associated to a g-isometric transformation with time-dependent parameters r^a, $\Lambda^{(a)}{}_l$ is of the form (59) with the following values of the coefficients:

$$\begin{aligned} A^{(a)} &= -\alpha(r)(\overline{\Lambda}^a{}_l \dot{\Lambda}^l{}_s r^s + \dot{r}^a), \\ B^{(ab)} &= \tfrac{1}{2}\dot{\overline{\Lambda}}{}^a{}_l \Lambda^{lb} - (\kappa/4)(A^{(a)} r^b - r^a A^{(b)}), \\ C &= -(\dot{\kappa}/4)\alpha(r) r^2, \\ D^{(a)} &= -(\kappa/4) A^{(a)} + (\dot{\kappa}/4)\alpha(r) r^a. \end{aligned} \tag{61}$$

It is interesting to note that the fact that the transformation is a g isometry does not imply that its associated motion v is a g-Killing vector field unless $\dot{\kappa} = 0$ (when the time-dependence occurs only via the group parameters r,Λ).

Taking into account Eqs. (40)–(42), we have the following proposition.

Proposition 6: The most general transformation which changes the standard form of the space-time metric $\hat{g}$,

$$\hat{g} = (\sigma^2 + s\cdot s)dt \otimes dt + s \otimes dt + dt \otimes s + \alpha^2 c_{ij} dy^i \otimes dy^j, \tag{62}$$

into its conform-standard form

$$\hat{g} = (\sigma^2 + (s+v)\cdot(s+v))dt \otimes dt + (s+v) \otimes dt + dt \otimes (s+v) + \rho^2 c_{ij} dz^i \otimes dz^j \tag{63}$$

is a time-dependent g-conformal transformation, and v is necessarily a g-conformal Killing vector field.

Remembering the dimensions of C and I, it is clear from the above results that two conform-standard (resp. standard) forms of the metric $\hat{g}$ are related by, at most, $(n+1)(n+2)/2$ [resp. $n(n+1)/2$] functions of the sole time.

V. THE UMBILICAL CASE

We have seen in the preceding section that the metrics of the space-times admitting a constant curvature synchronization depend [for every prescription of the curvature $\kappa(t)$] on an arbitrary function σ and on the equivalence classes $\{s\}$ of vector fields differing by conformal fields. We shall analyze here a particularly interesting class of such space-times: those for which every instant of the synchronization is umbilical. The synchronizations having this property will be called, for short, umbilical synchronizations.[28]

Let P be a point of an instant Σ. For every g-unitary vector field u^*, the quantity $K(u,u) \equiv i^2(u^*)K$ is called the *extrinsic* (or *normal*) *curvature* of Σ at P in the direction u^*; it measures the separation between the geodesic of V_{n+1} and the geodesic of Σ having at P the same direction u^*, and its maxima and minima values correspond to the principal directions of K with respect to g. P is called an *umbilical point* if $K(u,u)$ does not depend on the direction u^*, and Σ is called an *umbilical instant* if all its points are umbilical. Therefore, it is clear that S is umbilical if, and only if,

$$K = \beta g. \tag{64}$$

Let us consider our space-time written in a conform-standard chart

$$\hat{g} = (\sigma^2 + s\cdot s)dt \otimes dt + s \otimes dt + dt \otimes s + \rho^2(z)c_{ij}dz^i \otimes dz^j. \tag{65}$$

In such a chart, the time derivative of the induced metric $g_{ij} = \rho^2 c_{ij}$, is given by

$$\partial_t g_{ij} = 2(\dot{\rho}/\rho)g_{ij}, \tag{66}$$

so that, if we substitute now this value of $\partial_t g_{ij}$ into the definition (9) of K, we have the following proposition.

Proposition 7: A synchronization of constant curvature is *umbilical* if, and only if, the motion s associated to any conform-standard chart is a g-conformal field.

According to Propositions 6 and 7, if we start from the standard form (62) and perform a conformal transformation with associated motion $v = -s$ [v being of the form (59)], we shall arrive at the expression (63) for $\hat{g}$, with $s + v = 0$. Taking into account the relations (53) and (54) for $\rho(z)$, we have the following proposition.

Proposition 8: The space-times with umbilical and constant curvature synchronizations are those that can be written in the normal form

$$\hat{g} = \sigma^2\, dt \otimes dt + \rho^2 c_{ij}\, dx^i \otimes dx^j, \tag{67}$$

with

$$\rho = \theta/[1 - 2c\circ x + (c^2 + (\kappa/4)\theta^2)x^2], \tag{68}$$

where $\sigma = \sigma(x,t)$, $\kappa = \kappa(t)$, $\theta = \theta(t)$, $c_l = c_l(t)$.

Let us consider, for a moment, one of these normal conform-standard charts. According to the definition (9) of K, we have $2\sigma K = \partial_t g$, so that, taking (66) into account, we obtain

$$\beta = (1/\sigma)(\ln \rho)^{\cdot} \tag{69}$$

for the conformal factor β in (64).

From (64), the direct evaluation of the second Einstein constraint equation C_2 gives

$$-t = (n-1)d\beta. \tag{70}$$

On the other hand, it follows from (32) and (64) that

$$\mathrm{Ric}(g) = (n-1)\kappa g, \quad K \times K = \beta^2 g,$$

so that, taking traces and substituting into the first Einstein constrain equation C_1, we have

$$-2\tau = n(n-1)(\kappa - \beta^2). \tag{71}$$

On account of (66) and (64), the time derivative of K is

$$\partial_t K = (\dot{\beta} + 2\sigma\beta^2)g$$

and the evaluation of the tensor S defined in (11) gives

$$S = -[(n-2)/2]\beta^2 g.$$

Therefore, the tensor R defined in (10) results in

$$R = -(1/\sigma)\nabla\, d\sigma + \{(n-1)\kappa - n\beta^2 - \dot{\beta}/\sigma\}g, \tag{72}$$

and the third equation in (8) allows us to evaluate T. Let us note that, for a given prescription of $\kappa(t)$, σ and β are intrinsic scalars on every instant. Thus, as $\dot{\beta}/\sigma = L(\hat{n}^*)\beta$, we can state the following proposition.

Proposition 9: The S-characterization $\{\tau,t,T\}$ of the energy momentum tensor $\hat{T}$ of a space-time with an umbilical and constant curvature synchronization, is given, in any local chart adapted to S, by

$$\begin{aligned}
\tau &= [n(n-1)/2](\beta^2 - \kappa),\\
t &= -(n-1)d\beta,\\
T &= -(1/\sigma)(\nabla\, d\sigma + \Delta\sigma\cdot g)\\
&\quad + [(n-1)/2](2L(\hat{n}^*)\beta + n\beta^2 - (n-2)\kappa)g.
\end{aligned} \tag{73}$$

The Riemann tensor $\mathrm{Riem}(\hat{g})$ and the Weyl tensor $\mathrm{Conf}(\hat{g})$ of $(V_{n+1}, \hat{g})$ are related by

$$\mathrm{Conf}(\hat{g}) = \mathrm{Riem}(\hat{g}) - l(\hat{T}),$$

where $l(\hat{T})$ is a linear function of $\hat{T}$. Let us note by $(\hat{n}\hat{n}\hat{A})$ the two-tensor field with components $(\hat{n}\hat{n}\hat{A})_{\beta\delta} \equiv \hat{n}^\mu \hat{n}^\nu \hat{A}_{\mu\beta,\nu\delta}$, where $\hat{n}$ and $\hat{A}$ are, respectively, an unit vector field and a symmetric double two-form. Let us note by $*$ the duality operator acting on forms. It is well known that the two two-tensor fields

$$\hat{G} \equiv (\hat{n}\hat{n}\ \mathrm{Conf}(\hat{g})), \quad \hat{C} \equiv (\hat{n}\hat{n}{*}\mathrm{Conf}(\hat{g})) \tag{74}$$

determine biunivocally $\mathrm{Conf}(\hat{g})$; they are usually called electriclike and magneticlike components of the free gravitational field with respect to the observers of unitary velocity $\hat{n}$.

Now, let $\hat{n}$ be the unit normal to a synchronization S and let us note by G and C, respectively, the strict components of the Σ characterization of $\hat{G}$ and $\hat{C}$ on every instant Σ of S: G and C are nothing else but the restriction to Σ of $\hat{G}$ and $\hat{C}$, respectively. A straightforward computation in Gaussian charts relatively to Σ gives, taking into account the Einstein evolution equations (12),

$$G = \mathrm{Ric}(g) + K \times K - \mathrm{tr}\, K\cdot K - \tfrac{1}{2}T + \tfrac{1}{6}(4\tau + \mathrm{tr}\, T)g,$$
$$C = \delta\widetilde{K} - \tfrac{1}{2}{*t}, \tag{75}$$

where $*$ stands now for the duality operator on (Σ,g) and where we have noted by $\widetilde{K}$ the three-tensor obtained by duality from K considered as a double one-form, that is,

$$(\widetilde{K})_{ijk} = \eta_{ijl} K^l_{\ k}.$$

Equations (75) relate quantities defined on Σ by operations intrinsic to Σ. For this reason, they are valid in any local chart adapted to Σ, that is, for *every* instant Σ of S and *any* motion Γ.

In the flat case, $\mathrm{Riem}(\hat{g}) = 0 \Rightarrow \hat{T} = 0$, we have $G = C = 0$. The equation $G = 0$ shows then that the square of the extrinsic curvature K of Σ depends only on the induced metric g on Σ (Gauss theorem). The equation $C = 0$ gives the variation over Σ of the extrinsic curvature K (Codazzi theorem). For this reason G and C will be called,[28] respectively, the *Gauss* and the *Codazzi tensors of the instant* Σ.

On every instant Σ, G and C may be constructed from the geometrical data g and K and the physical data τ, t, and T. In the generic case, $\mathrm{Riem}(\hat{g}) \neq 0$, their nullity is the necessary and sufficient condition for $\hat{g}$ to be conformally flat.

Let us define

$$Y \equiv T - \tfrac{1}{3}(4\tau + \mathrm{tr}\, T)g, \quad Z \equiv G + \tfrac{1}{2}Y.$$

From (73) it is obvious that

$$T - \tfrac{1}{3}\mathrm{tr}\, T\cdot g = -(1/\sigma)(\nabla\, d\sigma + \tfrac{1}{3}\Delta\sigma\cdot g)$$

and taking into account the value of τ, it follows that

$$Y = -(1/\sigma)(\nabla\, d\sigma + \tfrac{1}{3}\Delta\sigma\cdot g) - 4(\beta^2 - \kappa)g.$$

On the other hand, if we substitute into (75) this result and the expressions for $\mathrm{Ric}(g)$ and $K \times K$ from the preceding paragraph, we have

$$Z = 2(\kappa - \beta^2)g.$$

Let us finally compute C. Accounting for the Einstein constraint equation C_2 of (12), (75) gives

$$C_{ij} = -\eta^l_{\ is}\nabla_l K^s_{\ j} - \tfrac{1}{2}\eta_{ijl}t^l$$
$$= -\eta^l_{\ is}\beta_l\delta^s_j - \tfrac{1}{2}\eta_{ijl}(1-n)\beta^l = [(n-3)/2]\eta_{ijl}\beta^l,$$

where β_l stands for $\partial\beta/\partial x^l$. We can then state the following proposition.

Proposition 10: In the physical space-times ($n = 3$), the Gauss and Codazzi tensors of the umbilical and constant curvature synchronizations are given, respectively, by

$$G = (1/2\sigma)(\nabla\, d\sigma + \tfrac{1}{3}\Delta\sigma\cdot g), \quad C = 0. \tag{76}$$

Thus, we see that, relatively to the system of observers normal to the synchronization (with unitary velocity $\hat{n}^*$), the space-time is of the "electric" type. Furthermore, from (73) and (76) it is clear that, apart from a numerical factor, G is nothing but the deviator (traceless part) of the stress tensor T:

$$G = -\tfrac{1}{2}(T - \tfrac{1}{3}\mathrm{tr}\, T\cdot g). \tag{77}$$

Proposition 11 follows.

Proposition 11: The only vacuum physical space-time admitting an umbilical and constant curvature synchronization is Minkowski space-time.

It is well known that the Petrov type of space-time is related to the number of independent eigenvectors of the tensor $P \equiv G + iC$. In our case ($P = G$), P is diagonalizable and, then, it is a Petrov type T_3. Thus, depending on the fact that the eigenvalues of G be all different, one double or the three null, we have, in Penrose notation, the following three cases of the Petrov–Bel[29] classification.

Proposition 12: The physical space-times admitting umbilical and constant curvature synchronizations are of the type I, D, or O of the Petrov–Bel classification.

In the case of a perfect fluid with unitary velocity $\hat{u}^*$, proper energy density ρ, and pressure p, the stress tensor T is

$$T = (\rho + p)u \otimes u - pg,$$

where u is the induced of $\hat{u}$ on every instant of S. It is obvious that p will be at least a double eigenvalue of T; thus, on account of (77) we have the following proposition.

Proposition 13: The perfect fluid space-times admitting umbilical and constant curvature synchronizations are of the Petrov–Bel type D or O. They are of type O if, and only if, the fluid is not tilted ($\hat{u} = \hat{n} \Leftrightarrow u = 0$).

The not-tilted case is conformally flat, so that the space-time is contained either in the class of the generalized Schwarzschild interiors (if $\beta = 0$) or in the class of generalized Robertson–Walker space-times (if $\beta \neq 0$) (see Ref. 30).

The Debever vectors[31] associated to the Petrov–Bel types are given by the Sachs equations.[32] The Σ characterization of these equations in terms of the Gauss and Codazzi tensors has been given elsewhere.[12]

Let us note by $\{1, l\}$ the Σ characterization of the Debever vectors $\hat{l}$, $g(l,l) = -1$ (see Ref. 33). The general Sachs equation

$$\delta^{\lambda\mu}_{\alpha\beta}\delta^{\pi\rho}_{\gamma\delta}\hat{l}_\lambda\hat{l}_\pi\hat{l}^\sigma\hat{l}^\tau(\mathrm{Conf}\,(\hat{g}))_{\sigma\mu,\tau\rho} = 0,$$

where $\delta^{\lambda\mu}_{\alpha\beta} = \delta^\lambda_\alpha\delta^\mu_\beta - \delta^\mu_\alpha\delta^\lambda_\beta$, is strictly equivalent to[12]

$$2\perp(l)G - i^2(l)G\cdot(g + l \otimes l)$$
$$= (*l)\times\perp(l)C + {}^t((*l)\times\perp(l)C)$$

where $\perp(l)$ is the projector orthogonal to l, $*l$ is the two-form dual of l, and ${}^tM$ stands for the transpose of the two-tensor M. In our case, it reduces to

$$2\perp(l)G - i^2(l)G\cdot(g + l \otimes l) = 0. \tag{78}$$

Let p_i ($i = 1,2,3$) be the eigenvalues of T (monotonically ordered), $v_{(i)}$ the corresponding eigenvectors,

$$T = -\sum_i p_{(i)}v_{(i)} \otimes v_{(i)},$$

and let us note by $p_{(ij)} \equiv p_{(i)} - p_{(j)}$ the relative increments of the partial pressures $p_{(i)}$; we have from (77)

$$G = \tfrac{1}{6}\sum_{i,j} p_{(ij)}v_{(i)} \otimes v_{(i)}. \tag{79}$$

If we substitute this expression for G into (78), a simple algebraic computation gives the following result.

Proposition 15: The Debever vectors of the space-times admitting umbilical and constant curvature synchronizations are of the form $\{1, \pm l_\pm\}$, with $l_\pm$ given by

$$l_{\pm} = \left(\frac{p_{(12)}}{p_{(13)}}\right)^{1/2} v_{(1)} \pm \left(\frac{p_{(23)}}{p_{(13)}}\right)^{1/2} v_{(3)}.$$

Let us note that if the space-time is of type I, $l_{\pm}$ is contained in the plane determined by the proper direction corresponding to the extreme pressures and if, instead, it is of type D, $l_{\pm}$ is in the direction corresponding to the simple pressure. Of course, in the case O all the $p_{(ij)}$ are null and the $l_{\pm}$ are undetermined.

ACKNOWLEDGMENTS

We are indebted to Dr. Natalie Derruelle for her valuable aid in preparing the final English version of the manuscript.

One of us (C. B.) wants to acknowledge the Spanish Comision Asesora para la investigación científica y técnica (CAICYT) for partial financial support.

[1]In fact, this is what we shall make here: the analytical integration of Einstein's constraint and evolution equations in the complete integrability case.

[2]The necessary and sufficient constraints that the Cauchy data of the Einstein equations must verify in order that the space-time admits a r-dimensional isometry group have been given by B. Coll, J. Math. Phys. **18**, 1918 (1977); C. R. Acad. Sci. Paris **292**, 461 (1981). But the integrability conditions for these constraints, not yet studied, depend on the isometry group admitted by the intrinsic metric on the initial hypersurface Σ, that is, in other words, on the Σ-Killing vector fields.

[3]The notion of rigidity related to a synchronization has been considered by B. Coll, C. R. Acad. Sci. Paris **295**, 103 (1982); and C. Bona, Phys. Rev. D **27**, 1243 (1983).

[4]A. Krasinski, Gen. Relativ. Gravit. **13**, 1021 (1981).

[5]C. B. Collins and D. A. Szafron, J. Math. Phys. **20**, 2347, 2354, 2362 (1979).

[6]For $n > 3$, the pentadimensional case ($n = 4$) is particularly adapted to a formulation in terms of synchronizations. Some interesting aspects of the cases $n < 3$ have been pointed out recently by S. Giddings, J. Abbott, and K. Kuchar, Gen. Relativ. Gravit. **16**, 751 (1984).

[7]For our purposes, we only need that $\hat{g}$ be nondegenerate. Nevertheless, in order to have a direct physical interpretation on V_4, we shall suppose $\hat{g}$ to be also Lorentzian.

[8]δ denotes, to within a sign, the divergence operator. In local charts $(\delta\hat{T})_A = -\nabla_\rho \hat{T}^\rho{}_A$, A being some set of tensorial indices.

[9]For the physical meaning of a synchronization, see B. Coll, "A relativistic notion of rigidity," to be published.

[10]d is the usual operator of exterior differentiation on forms, and $\wedge$ denotes the exterior product.

[11]The evolution formalisms of general relativity are *relative* in the sense that the space-time is replaced by a time-dependent three-dimensional spatial geometry, and the geometrical objects considered are defined on this evolutive geometry. The *standard* evolution formalism, the only one considered here, is the one in which every instant of the synchronization is characterized by its first and second fundamental forms. However, many other formalisms may be considered, in which the induced metric is replaced by the quotient metric relative to the given motion, or by a conformal one, the extrinsic curvature and the characterization of the motion being replaced by "more or less" adapted quantities. For some of these different possibilities see, for example, L. Bel and J. C. Escard, Rend. Cl. Sci. Fis. Mat. Nat., Accad. Nazio. Lincei, **41**, 476 (1966); J. Stachel, J. Math. Phys. **21**, 1776 (1980); R. A. Nelson, Gen. Relativ. Gravit. **14**, 647 (1982).

[12]B. Coll, Thèse d'état, Univ. Pierre et Marie Curie, Paris, 1980.

[13]In local charts, the *interior product* is $(i(\hat{n}^*)\hat{T})_A \equiv \hat{n}^\rho \hat{T}_{\rho A}$, where A represents some set of tensorial indices.

[14]The canonical isomorphisms between tensors and cotensors defined by the metrics $\hat{g}$ and g on $(V_{n+1}, \hat{g})$ and (Σ, g), respectively, allow us to extend the notion of Σ characterization to the contravariant tensors of $(V_{n+1}, \hat{g})$: the *Σ characterization E^* of a p tensor* $\hat{E}^*$ on $(V_{n+1}, \hat{g})$ is the p-extensor field of (Σ, g) associated by g to the Σ characterization E of the p-cotensor field $\hat{E}$ associated by $\hat{g}$ to $\hat{E}^*$.

[15]The direct calculation is long and tedious. A short way, based on the geometric properties of the Σ characterizations, is given in Ref. 12.

[16]A. Lichnerowicz, Thèse d'état, Univ. de Paris, Paris, 1939.

[17]A. Lichnerowicz, J. Math. Pures Appl. **23**, 37 (1944).

[18]Y. Choquet-Bruhat, C. R. Acad. Sci. Paris **226**, 1071 (1948).

[19]Y. Choquet-Bruhat, J. Rat. Mec. Anal. **5**, 951 (1956).

[20]R. Arnowitt, S. Deser, and C. W. Misner, "The dynamics of General Relativity," in *Gravitation: An Introduction to Current Research* (Wiley, New York, 1962).

[21]In fact, being a tensor field *on* $(V_{n+1}, \hat{g})$, $\hat{g}_I$ is a convenient one-parameter family of induced metrics.

[22]See, for example, A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnetisme* (Masson, Paris, 1955).

[23]C. Bona and B. Coll, "Classification of the space-times admitting a constant curvature synchronization," in preparation.

[24]In the Lorentzian case and signature $-(n-1)$, one has $c_{ij} = -\delta_{ij}$.

[25]That is, Y_t transforms the local representation $\{\sigma;0\}$ of $\{\Gamma_N, S\}$ in the local representation $\{\sigma;s\}$ of $\{\Gamma, S\}$, where s is given by (39).

[26]Of course, $a,b = 1,\ldots,n$ and $m_{(ab)} = -m_{(ba)}$.

[27]For the notion of umbilical points see, for example, J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954).

[28]For the introduction of the Gauss and Codazzi tensors in $(V_4, \hat{g})$, see Ref. 12.

[29]The Petrov classification [A. Z. Petrov, Sci. Not. Kazan State Univ. **114**, 55 (1954)] involves only three types of space-times. The complete one, involving the five well-known (nontrivial) types, was obtained by L. Bel, Thèse d'état, Univ. de Paris, Paris, 1959; for this reason, we call it the Petrov–Bel classification.

[30]H. Stephani, Commun. Math. Phys. **4**, 137 (1967). See also, D. Kramer *et al.*, *Exact Solutions of Einstein's Field Equations* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1980), Chap. 32.

[31]R. Debever, C. R. Acad. Sci. Paris **249**, 1324 (1959).

[32]R. Sachs, Proc. R. Soc. London, Ser. A **264**, 309 (1961).

[33]We have used the signature $(+ - - -)$; thus, g is negative definite.

A gravitational lens produces an odd number of images

Ross H. McKenzie
Joseph Henry Laboratories of Physics, Princeton University, Princeton, New Jersey 08544

(Received 9 August 1984; accepted for publication 18 January 1985)

Rigorous results are given to the effect that a transparent gravitational lens produces an odd number of images. Suppose that p is an event and T the history of a light source in a globally hyperbolic space-time (M,g). Uhlenbeck's Morse theory of null geodesics is used to show under quite general conditions that if there are at most a finite number n of future-directed null geodesics from T to p, then M is contractible to a point. Moreover, n is odd and $\frac{1}{2}(n-1)$ of the images of the source seen by an observer at p have the opposite orientation to the source. An analogous result is noted for Riemannian manifolds with positive definite metric.

I. INTRODUCTION

If a galaxy is located between a light source, such as a quasar, and an observer then the presence of the gravitational field of the galaxy may imply the existence of more than one light path from the source to the observer. Consequently, the observer sees more than one image of the source. Five different cases of this multiple imaging effect, sometimes called the gravitational lens effect, have now been observed by astronomers.[1–5] Dyer and Roeder have shown that a transparent static spherically symmetric gravitational lens (i.e., galaxy) must produce an odd number of images.[6] Burke has shown that this is also true for lenses without spherical symmetry.[7]

There is a relatively simple demonstration of why there are an odd number of images. Although it seems to be well known among astronomers it does not apper to have been published before and so is given here. Consider the situation shown in Fig. 1. A light source is located at S and an observer at O. There is a transparent galaxy G somewhere between S and O. A map f from the small sphere A to the sphere B is defined as follows. The map f maps a point x on A to the point on B where the light ray through O and x intersects B. The number of images of S seen by O is the number of points on A mapped onto S.

Suppose $g:M \to N$ is a smooth map between manifolds of the same dimension and that M is compact. If y is a regular value of g then we define

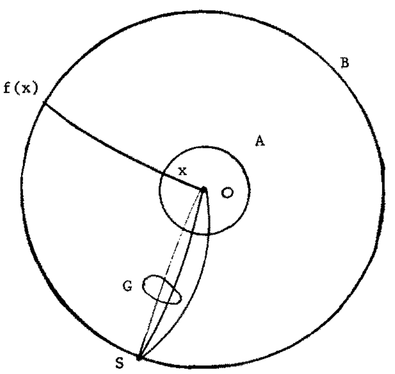


FIG. 1. A galaxy G is located somewhere between a light source S and an observer O. Because of the gravitational field of the galaxy there may be more than one light ray from S to O. f maps the sphere A onto the sphere B. If x is on A then $f(x)$ is defined to be the point on B where the ray through O and x intersects B.

$$\deg(g,y) = \sum_{x \in g^{-1}(y)} \operatorname{sgn} dg_x,$$

where $\operatorname{sgn} dg_x = +1(-1)$ if $dg_x : T_x(M) \to T_y(N)$ preserves (reverses) orientations. It turns out that $\deg(g,y)$ is the same for all regular y; it is called the degree of g and denoted $\deg(g)$. A complete discussion is given in Ref. 8.

In an actual physical situation it is reasonable to expect that there will be a point y on B such that $f^{-1}(y)$ is a single point, i.e., there is only one ray from O to y. Thus, $\deg(f) = 1$.

Let n_+ (n_-) be the number of points x in $f^{-1}(S)$ such that $\operatorname{sgn} df_x = +1(-1)$. Thus, n_+ (n_-) is the number of images of S, seen by O, which have the same (opposite) orientation as the source, and

$$n_+ - n_- = \deg(f,S) = \deg(f) = 1.$$

Thus, if O sees $n = n_+ + n_-$ images of S then $n = 2n_- + 1$ and so n is odd, and the demonstration is complete. Note that this argument also yields $n_- = \frac{1}{2}(n-1)$, i.e., the number of images with reversed orientation is $\frac{1}{2}(n-1)$.

There are several difficulties encountered when an effort is made to make this argument into a rigorous proof. First, A and B must be chosen so that f is a smooth single-valued map. Second, there does not necessarily exist a point y such that $f^{-1}(y)$ is a single point. (It might be possible to show that the degree of f is 1 because f is homotopic to the identity since the interior of B is contractible to a point.) More importantly, it has been implicitly assumed that the casual structure and topology of the underlying space-time are "very reasonable." One of the aims of this paper is to go part of the way toward establishing the most general causal structure and topology that the space-time can have, while still retaining an odd number of images.

It should be noted that an astronomer might not observe an odd number of images due to resolution or intensity limitations. In fact, in four of the five known cases of multiple imaging only two images have been observed. Recently several high-resolution studies of these quasars did not detect a third image.[9]

Although no original mathematics is involved, a rigorous treatment of multiple imaging is justified on three grounds. First, gravitational lenses are receiving a great deal of attention from astronomers because they might help solve two of the outstanding problems in astronomy: the values of the three cosmological constants H_0, q_0, and Λ, and the distribution of mass in galaxies and clusters of galaxies. Because

 0022-2488/85/071592-05$02.50

a considerable effort is being made to find third images it is worth knowing under precisely what conditions there should be an odd number of images. Second, the treatment given here shows that the fact that there are an odd number of images is basically a topological result and does not depend on any particular property of gravity. Third, it is an interesting application of global analysis.

The model used here to investigate the gravitational lens effect is the following: the instant of observation is an event p and the history of the source is a timelike curve T in a space-time (M,g) (a manifold M with Lorentz metric g). The presence of lenses or galaxies implies that the space-time is not flat. If the lens is not transparent then the space-time has the region where the lens is removed. The formalism used here will implicitly assume that the lens is transparent. Each image seen at p corresponds to a future-directed null geodesic from T to p.

It is important to note that no assumptions are made concerning the location of the lens, matter distribution in the lens, or how the metric g depends on the energy momentum tensor (in general relativity this is governed by Einstein's field equations). Thus the model also represents the propagation of light in an inhomogeneous medium.

II. A MORSE THEORY OF NULL GEODESICS

Showing there are an odd number of images is a direct application of the Morse theory of null geodesics due to Uhlenbeck.[10] An excellent introduction to Morse theory is Milnor's book.[11] Morse theory describes the relationship between the topology of a differentiable manifold and the critical point behavior of generic smooth functions on the manifold. A point p in a manifold M is called a critical point of a real-valued function f on M if $(df)_p = 0$; that is if M is a finite-dimensional manifold and $(x^1,...,x^n)$ is a local coordinate system in a neighborhood of p then

$$\frac{\partial f}{\partial x^1}(p) = ... = \frac{\partial f}{\partial x^n}(p) = 0.$$

Uhlenbeck constructs a functional on a space of curves such that the critical points of the functional are in one-to-one correspondence with the null geodesics from T to p. She assumes that (M,g) is globally hyperbolic. Conjectures concerning the necessity of this assumption are made in Sec. IV. There are several definitions of global hyperbolicity which are all equivalent. A complete discussion is given in Ref. 12. One definition is as follows: a space-time M is said to be *globally hyperbolic* if it contains a Cauchy surface S (Ref. 13). It can then be shown that M is homeomorphic to $S \times R$.

A Cauchy surface is basically a spacelike surface such that specification of the value of a field on the surface suffices to determine the value of the field throughout the entire space-time. Examples of globally hyperbolic space-times are the Friedman, maximally extended Schwarzschild, and all static space-times. Examples of space-times which are not globally hyperbolic are the Kerr, Taub–NUT (Newman–Unti–Tamburino), plane wave, anti-de Sitter, and the maximally extended Reissner–Nordstrom space-times.

Two events p and q are said to be conjugate along a given geodesic if there exists a nonzero variation of geodesics (or Jacobi field) along the geodesic which vanishes at p and q. The *index* of the given geodesic is the number of conjugate points, counting multiplicities, along the geodesic. A curve T and an event p are said to be nonconjugate if (a) all the null geodesics from p to T are nonconjugate and (b) if a geodesic from p to T is conjugate then $\exp_p : T_p(M) \to M$ is transverse to T.

Theorem 1[14]: Let (M,g) be a globally hyperbolic space-time and T a smooth timelike curve in M. Then for all points p in M not belonging to a set of measure zero, T and p are nonconjugate.

If T and p are conjugate along a given null geodesic, then the corresponding image of the source will have infinite intensity. However, Theorem 1 can be interpreted as saying that the probability of this occurring is zero.[15]

It should be noted that Uhlenbeck considered null geodesics from p to T, whereas the problem considered here involves null geodesics from T to p. Consequently, in the theorems due to Uhlenbeck stated below the direction of the time has been reversed.

Suppose that in the orthogonal splitting $M = S \times R$, $p = (r,t_0)$. Let q be the point on S where the timelike curve T intersects $S \times \{t_0\}$. If $\alpha:[0,1] \to S$ is a path in S then let $P_\alpha(t)$ be the solution to the differential equation

$$\left[\frac{d}{dt}P_\alpha(t)\right]^2 = g(\alpha(t),P_\alpha(t))\left(\frac{d\alpha}{dt},\frac{d\alpha}{dt}\right),$$

with the initial condition $P_\alpha(0) = t_0$.

Uhlenbeck defines a functional J on the set of paths in S from q to r by

$$J(\alpha) = \int_0^1 \left(\frac{d}{dt}P_\alpha(t)\right)^2 dt.$$

If M is a static space-time then, up to a constant factor, $J(\alpha)$ is just the value of the energy integral for α. In loose terms, $J(\alpha)$ is a measure of the "time" taken for a light signal to travel along the curve α. Consequently, the first half of the following theorem can be interpreted as a generalization of Fermat's principle for space-times.

Theorem 2[16]: Let (M,g) be a globally hyperbolic space-time and $M = S \times (a,b)$. Then, the critical points of the functional J are the unique projections onto S of null geodesics from T to p. Let $\Omega(q,r)^c$ be the set of piecewise differentiable curves in S, with endpoints q and r, and such that $J(\alpha) < c \leq b$. If T and p are nonconjugate then

$$M_k \geq B_k(\Omega(q,r)^c),$$

$$\sum_{k=0}^{\infty}(-1)^k M_k = \sum_{k=0}^{\infty}(-1)^k B_k(\Omega(q,r)^c) = \chi(\Omega(q,r)^c),$$

where $B_k(\Omega(q,r)^c)$ is the Euler characteristic of $\Omega(q,r)^c$, and M_k is the number of null geodesics, $(\alpha(t),P_\alpha(t))$, with $J(\alpha) < c$, of index k, from T to p.

This is essentially a local theorem. Uhlenbeck also developed a result which related the number of null geodesics to the global topology of the space-time.

A smooth orthogonal splitting $M = S \times R$ is said to satisfy the metric growth condition if in the coordinates (x^i,t) the metric is

$$ds^2 = g_{ij}(x,t)dx^i dx^j - r(x,t)dt^2$$

and for every compact $K \subset S$ there exists a function $F(t)$ with $\int_{-\infty}^{0} [d\tau/F(\tau)] = \infty$ such that for $t \leq 0$ and a fixed Riemannian metric $\hat{g}$ on S

$$[1/r(x,t)]g_{ij}(x,t)v^i v^j \leqslant F^2(t)\hat{g}_{ij}(x)v^i v^j,$$

for all $v^i \in T_x(S)$ and for all $x \in K$. This condition holds trivially for conformally static space-times. As we shall see this condition ensures that a large class of timelike curves do not have particle horizons. With respect to Uhlenbeck's theory this condition ensures that each curve in S which is a critical point of J is realized as a null geodesic from T to p.

Theorem 3[17]: Suppose that T is a timelike curve and p is an event in a space-time (M,g) and that the following conditions hold: (1) (M,g) is a globally hyperbolic space-time; (2) there is an orthogonal splitting $S \times R$ of M which satisfies the metric growth condition; (3) if in this splitting $T = (y(t),t)$ then $\lim_{t \to -\infty} y(t)$ exists; and (4) p and T are nonconjugate. Then

$$M_k \geq B_k(\Omega(M)),$$

where M_k is the number of null geodesics from T to p of index k and $\Omega(M)$ is the loop space of M, the set of all paths in M starting and ending at a given point.

This result differs from the local theorem (Theorem 2) in that no restriction is placed on the "energy" of the geodesics and the path space, whose topology is related to the number of null geodesics of different index from T to p, does not depend on p or T but only the topology of M.

Corollary 1: There exists at least one null geodesic from T to p.

This means that $J^+(T) = M$ and so T has no particle horizon. Similarly, if the metric growth condition holds for $t \geq 0$ then T has no event horizons.

III. TOPOLOGICAL CONDITIONS WHICH ENSURE AN ODD NUMBER OF IMAGES

In order to show that there are an odd number of null geodesics from T to p assumptions are now made which imply that $\Omega(q,r)^c$ and $\Omega(M)$ are contractible to a point. It then follows that both $\chi(\Omega(q,r)^c)$ and $\chi(\Omega(M))$ are one and so $M_+ - M_- = 1$, where $M_+(M_-)$ is the number of null geodesics of even (odd) index from T to p. Finding appropriate assumptions is greatly simplified by the fact that the topology of $\Omega(q,r,B)$, the space of paths in a set B joining q and r, two points in B, is closely related to the topology of B. In particular, if B is contractible to a point then so is $\Omega(q,r,B)$.

The ith homotopy group $\pi_i(\Omega(q,r,B))$ is isomorphic to $\pi_i(\Omega(B))$ which is in turn isomorphic to $\pi_{i+1}(B)$ for $i \geq 0$ (Ref. 18). If B is contractible to a point then all the homotopy groups of B are trivial. Thus, the homotopy groups of $\Omega(q,r,B)$ are also trivial and so it must also be contractible to a point. This can also be shown directly by constructing an explicit deformation retraction.[19]

In the local version of the multiple image theorem it will be assumed that $\Omega(q,r)^c$ is a deformation retract of $\Omega(q,r,B)$ [i.e., any curve in B from q to r can be deformed into a curve α with $J(\alpha) < c$] where B is a subset of S which is contractible to a point.

It turns out that no additional assumptions are needed to show that $\Omega(M)$ is contractible to a point. The following proposition, which is of interest in its own right, shows that if there are only a finite number of null geodesics from T to p then M is contractible to a point. If the Poincaré conjecture is correct, this means that M is homeomorphic to R^4.

Proposition 1: Suppose that the assumptions in Theorem 3 hold. If M is not contractible to a point, then there is an infinite number of null geodesics from T to p.

This is the analog of the following result due to Serre[20] for Riemannian manifolds.

Proposition 2: Suppose that M is a complete connected Riemannian manifold which is not contractible to a point. Then any pair of nonconjugate points in M is joined by an infinite number of geodesics.

The proof of Proposition 1 given below follows Serre's proof closely.

Proof of Proposition 1: Let U be the universal covering manifold for S. Then $N = U \times R$ can be given the structure of a Lorentz manifold by the canonical projection $\pi: N \to M$, which is then a local isomorphism.[21] It can be readily shown that N satisfies conditions (1) and (2) in Theorem 3. Let T' be a curve in N such that $\pi(T') = T$. Then since π is a local isomorphism, T' is also timelike and condition (3) is satisfied. Let $\{p_i'\}_{i \in I} = \pi^{-1}(p)$, where the index set I is an arbitrary ordering of the elements of the fundamental group, be the set of points in N which are projected onto p. Again, because it is a local isomorphism, the projection defines a one-to-one correspondence between the null geodesics in N joining T' and $\{p_i\}_{i \in I}$ and the null godesics in M joining T and p. Two different cases are now considered.

(A) *The fundamental group of M, $\pi_1(M)$, is infinite*: The set $\{p_i'\}_{i \in I}$ is infinite. From Corollary 1, there is at least one future-directed null geodesic from T' to each p_i. Projecting these null geodesics into M gives an infinite number of future-directed null geodesics from T and p.

(B) *M has finite fundamental group*: Let p' and T' in N be such that $\pi(T') = T, \pi(p') = p$. It is then sufficient to show that there is an infinite number of null geodesics from T' to p'. It is here that the hard part of Serre's argument comes into play. He shows that if N is the universal covering manifold of a noncontractible manifold with finite fundamental group then N is not contractible to a point and moreover there are an infinite number of integers i such that $B_i(\Omega(N)) \neq 0$. Theorem 3 can be applied to the null geodesics in N from T' to p'. The inequalities given there imply that there are an infinite number of null geodesics from T' to p'.

It is important to be clear about what is meant when it is stated that there is an infinite number of geodesics between two points. A geodesic is a curve $\alpha(t)$ in M (i.e., a mapping $\alpha:[0,1] \to M$) whose tangent vector $d\alpha/dt$ is parallel transported along α. The distinction between the mapping α and the point set which is its image (i.e., the set of points traced out by the curve α) must be made to avoid confusion when discussing Propositions 1 and 2. For example, if two points on a sphere are not antipodal then there are an infinite number of distinct geodesics joining them.[22] However, all but two of these geodesics trace out the same set of points: the great circle containing the two points.

If there are an even (odd) number of conjugate points along a null geodesic between T and p then the corresponding image of the source seen by an observer at p has the same (opposite) orientation as the source.[23]

The previous discussion is now summarized in the following theorems.

Theorem 4: Suppose that T is the history of a light source S and p is an event in a globally hyperbolic space-time (M,g). Suppose that p and T are nonconjugate and $\Omega(q,r)^c$ is a deformation retract of $\Omega(q,r,B)$ where B is contractible to a point. Then, there are an odd number $(2n+1)$ of null geodesics $(\alpha(t), P_\alpha(t))$ with $J(\alpha) < c$, from T to p. Exactly n of the corresponding images of S seen at p will have the opposite orientation to the source.

Theorem 5: Suppose that the assumptions in Theorem 3 hold and that there are only a finite number m of null geodesics from T to p. Then M is contractible to a point and m is odd. Exactly $\frac{1}{2}(m+1)$ of the images, seen at p, of a light source whose history is S will have the same orientation as S.

There are results analogous to Theorems 2 and 3, but which concern geodesics on Riemannian rather than Lorentzian manifolds. Since these results do not appear to have been noted before they are now stated. They follow directly from Proposition 2 and the Morse inequalities for geodesics proven by Palais.[24]

Let p and q be two conjugate points in a complete Riemannian manifold and $\Omega(p,q)$ be the set of paths from p to q. The energy functional $E:\Omega(p,q) \to R$ is defined as follows. If $\sigma:[0,1] \to M$ then

$$E(\sigma) = \int_0^1 g\left(\frac{d\sigma}{dt}, \frac{d\sigma}{dt}\right) dt.$$

If $\Omega(p,q)^c = \{\sigma\,\Omega(p,q) | E(\sigma) < c\}$ is contractible to a point then there are an odd number of geodesics, of energy less than c, from p to q. Futhermore, it follows from Proposition 2 that if there are only a finite number n of geodesics (of all energies) from p to q then M is contractible to a point and by the usual argument n is odd. In each case exactly $\frac{1}{2}(n+1)$ $[\frac{1}{2}(n-1)]$ of the geodesics have an even (odd) number of conjugate points on them between p and q.

IV. APPLICATION OF THE RESULTS TO MULTIPLE IMAGING BY GRAVITATIONAL LENSES

Several astronomers have calculated numerically the positions and intensities of the different images that would be seen by an observer in a multiple imaging situation.[1,25,26] It has been noted that if the lensing parameters (for example, the distribution of mass in the lens, the source and observer positions) are varied so that two images coalesce, then as the images move closer together their intensity increases without bound. This behavior is explained by Theorem 5. The space-times corresponding to the lenses considered are static. Thus, they are globally hyperbolic and satisfy the metric growth condition. The source is at a fixed point in space so $\lim_{t\to-\infty} y(t)$ exists. If there are a finite number of images n then by Theorem 5 n can only be even if p and T are conjugate along some null geodesics γ. As mentioned previously this means that the image corresponding to γ has infinite intensity and in a certain sense there is a probability of zero of this situation occurring. If the intensity of each image is a continuous function of the lensing parameters, then as the two images coalesce their intensity must increase without bound.

The result concerning the number of images of each orientation might be of practical significance. Quasar images in the radio frequency range sometimes have sufficient structure that the relative orientation of the images might be determined. Thus, if only two images are detected and they have the same orientation, then the third image must have the opposite orientation. If the lens galaxy is spherically symmetric, then the images of the galaxy and quasar lie on a single line and there are constraints on the possible location of a third image. Let L be the line segment between the two images and $\bar{L}$ its complement. If the two observed images have the same orientation, then there must be an image with the opposite orientation on L. If the two images have the opposite orientation, then there must be an image on $\bar{L}$. These conclusions can be drawn from the diagram and discussion in Ref. 27. Unfortunately, if the lens is not spherically symmetric, it does not seem to be possible to make deductions about the possible positions of a third image.

The applicability of these theorems to the observed cases of gravitational lensing is now discussed. Despite its somewhat technical nature the local theorem is more relevant than the global result to the situations under study of astronomers. First, the observation of multiple images is very local in the following sense. Astronomers do not look for the multiple images over the whole sky but rather only over a very small solid angle. Also, only light rays from the source which take approximately the same "time" to reach the observer can be attributed to be from a single quasar. This is because multiple images are assumed to represent the same source if they have the same spectra. If an image corresponding to an earlier epoch of the quasar was observed, its red shift and spectra would be different and so it would not be known whether or not it was from a different source. Thus consideration of $\Omega(q,r)^c$ is more physical than consideration of the whole path space.

Second, the space-time representing the universe is not contractible to a point because of the singularity in the past (the "big bang"). Also the Cauchy surface might be compact (e.g., S^3) and so not be contractible to a point. The metric growth condition [in particular, $\int_{-\infty}^0 dt/F(t) = \infty$] is not satisfied for most realistic cosmological models. It is only satisfied for a Robertson–Walker space-time if the scale factor $R(t)$ tends to zero more rapidly than t. This is not the case for physically realistic models.

The assumption, made in both Theorems 4 and 5, that the space-time is globally hyperbolic is now considered: is it physically realistic and is it necessary for there to be an odd number of images? At least we want the region of space-time traversed by the light from the quasar to the earth to be globally hyperbolic. Geroch and Horowitz[28] discuss whether the universe possesses a Cauchy surface. There is no definitive answer but they consider that there is the potential for some solid evidence.

The Gödel universe, a plane wave (gravitational and/or electromagnetic), and the anti-de Sitter universe (hereafter referred to as M_1, M_2, and M_3, respectively) are all homeomorphic to R^4 and are not globally hyperbolic. I have not

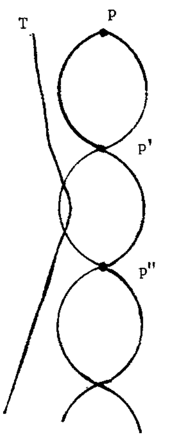


FIG. 2. The past null cone at an event p in the Gödel space-time is refocused at p',p'',etc. The timelike curve T intersects the past null cone of p at only two events.

been able to find p and T in M_2 or M_3 such that there are an even number of null geodesics from T to p. The Gödel universe has numerous pathological properties (the best known being that it possesses closed timelike curves) and so perhaps it is not surprising that there are p and T in M_1 joined by only an even number of null geodesics. See Fig. 2. A more complete diagram showing the global structure of M_1 is given in Ref. 12. The past null cone at any event p is refocused at p', p'', etc. The curve T shown is timelike (but not a geodesic) yet it intersects the past null cone of p at only two events and so there are only two null geodesics from T to p. Although the Gödel universe is not physically realistic this example shows that some assumptions on the causal structure of the space-time are necessary for there to be an odd number of images. It is probably necessary to assume at least that the space-time is stably causal[29] (i.e., even when the null cones "opened out" a small amount there were no closed timelike curves in the space-time). It is interesting to note that M_2 and M_3 are both weakly asymptotically simple[30] with a past null infinity $\mathscr{I}^-$ which is topologically $R \times S^2$. This could be significant because if a space-time M is asymptotically empty and simple then $\mathscr{I}^+$ and $\mathscr{I}^-$ are null and topologically $R \times S^2$, and M is globally hyperbolic and homeomorphic to R^4 (Ref. 30). Thus, it could be that there are an odd number of images when $\mathscr{I}^-$ is topologically $R \times S^2$, and the timelike curve T has its past endpoint on i^-, past timelike infinity.

ACKNOWLEDGMENTS

The author would like to thank Dr. P. Price and Professor A. Wightman for their constructive criticism. Some of his work was done while the author was a student in the Department of Theoretical Physics, Faculty of Science, Australian National University.

This research was supported in part by National Science Foundation Grant No. DMR8117463.

[1] P. Young, J. E. Gunn, J. Kristian, J. B. Oke, and J. A. Westphal, Astrophys. J. **241**, 507 (1980).
[2] P. Young, R. S. Deverill, J. E. Gunn, J. A. Westphal, and J. Kristian, Astrophys. J. **224**, 723 (1981).
[3] D. W. Weedman, R. J. Weymann, R. F. Green, and T. M. Heckman, Astophys. J. Lett. **255**, L5 (1982).
[4] C. R. Lawrence, D. P. Schneider, M. Schmidt, C. L. Bennett, J. N. Hewitt, B. F. Burke, E. L. Turner, and J. E. Gunn, Science **223**, 46 (1984).
[5] S. G. Djorgovski and H. Spinrad (to appear).
[6] C. C. Dyer and R. C. Roeder, Astrophys. J. Lett. **238**, L67 (1980).
[7] W. L. Burke, Astrophys. J. Lett. **244**, L1 (1981).
[8] V. Guillemin and A. Pollack, *Differential Topology* (Prentice–Hall, Englewood Cliffs, NJ, 1974).
[9] D. E. Thomsen, Science News **125**, 154 (1984).
[10] K. Uhlenbeck, Topology **14**, 69 (1975).
[11] J. Milnor, *Morse Theory* (Princeton U. P., Princeton, NJ, 1963).
[12] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., London, 1973).
[13] The formal definition of a Cauchy surface is given in Ref. 12, p. 205.
[14] Ref. 10, Theorem 2, p. 71.
[15] R. H. McKenzie, B. Sc. (Honors) thesis, Australian National University, 1982, p. 30.
[16] Ref. 10, Theorem 4.5, p. 85.
[17] Ref. 10, Theorem 3, p. 73.
[18] J. F. Adams, *Infinite Loop Spaces* (Princeton U. P., Princeton, NJ, 1978), p. 8.
[19] Reference 15, p. 146.
[20] J. P. Serre, Ann. Math. **54** 425 (1951), see p. 484.
[21] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic, New York, 1978), p. 59.
[22] Reference 11, p. 96.
[23] An intuitive explanation is given in Ref. 15, p. 153.
[24] R. S. Palais, Topology **2**, 299 (1963), see p. 338.
[25] R. R. Bourassa and R. Kantowski, Astrophys. J. **195**, 13 (1975).
[26] P. Young, J. E. Gunn, J. Kristian, J. B. Oke, and J. A. Westphal, Astrophys. J. **244**, 736 (1981).
[27] J. E. Gunn, "Gravitational Lenses in Astrophysics," in *Tenth Texas Symposium on Relativistic Astrophysics*, edited by R. Ramaty and F. C. Jones (N. Y. Acad. Sci., New York, 1981), p. 287.
[28] R. Geroch and G. T. Horowitz, "Global Structure of Spacetimes," in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U. P., London, 1979), p. 253.
[29] The formal definition is given in Ref. 28, p. 241.
[30] A proof is given in Ref. 12, p. 223.

Topological closure as the necessary condition for frustration or phase transitions

H. Nencka-Ficek[a)]
International Centre for Theoretical Physics, Trieste, Italy

(Received 17 August 1984; accepted for publication 4 January 1985)

It is shown that the same conditions are necessary for both frustration in Ising models with negative multispin interactions present and phase transition in analogous ferromagnetically coupled models.

I. INTRODUCTION

Since 1851 when Riemann outlined a program of "the study of the general meaning of an entity obtained by a successive increase of dimensionality"[1] topology has not only become a very important branch of mathematics but also a very useful tool to investigate a variety of properties of nature. As our paper concerns some problems of statistical mechanics of spin systems which are very closely connected with gauge theory,[2] let us present a short list of the applications of topology in physics, starting from the gauge theory. The best example for hidden topological concepts in gauge theory is the phenomenon of flux trapping in a superconductor. The quantization of flux can be expressed as $\oint \mathbf{A}\cdot d\mathbf{x} = \alpha N$, where α is a constant dependent on the Cooper-pair charge, and N is an integer which is a physically measurable quantum number. One can reinterpret this number as a topological property of a geometrical space (called a winding number[2]). Another application of topological methods in physics is the classification of the topological excitations[3] and defects, for example, the disclination lines.[4] By using the homotopy theory[5] it is easy to detect which topological defects or excitations are stable. Polymer problems[6] can be treated by an application of the knot theory.[7] The evolution in time of the motion of the polymers behaves differently for knotted and unknotted polymers.

In this paper we would like to present the application of algebraic topology to statistical mechanics. We consider two problems, the appearance of frustration in the models with Z_2 symmetry with negative couplings present and the phase transitions in other ferromagnetically coupled models of the same symmetry. This means that we consider the Ising models with multispin interactions with or without frustration. The model with frustration is a version of the model introduced by Wegner[8] and by Balian, Drouffe, and Itzykson.[9] The frustration considered here is the generalization of the frustration for two-spin interactions introduced by Toulouse.[10,11]

The second example considered is the Griffiths–Wood ferromagnetic Ising model with multispin interactions.[12]

Considering the mechanism of appearance of frustration[13] or the phase transitions in different models of Z_2 symmetry, we find the necessary conditions for each of them separately. These conditions are topological. Subsequently, we can perceive that both of them, although applying to physically different systems and different phenomena, are mathematically the same. In the next section we present some useful definitions and in Sec. III the necessary conditions.

II. DEFINITIONS

If we investigate the general mechanism of the arbitrary effect we can fix the way out by generalizing the model describing it. As we examine the mechanism of frustration or phase transition we may consider somewhat more general models but with the assumption that its symmetry does not change. Therefore, we consider the Ising model with multispin interactions in D-dimensional arbitrary lattice. The Hamiltonians are as follows:

$$-\beta\mathscr{H} = K_1 \sum_{\{b_1\}} A(b_1) s_{i_1} s_{i_2} + K_2 \sum_{\{b_2\}} A(b_2) s_{i_1} s_{i_2} s_{i_3} + K_2' \sum_{\{b_2'\}} A(b_2') s_{i_1} \cdots s_{i_4} + \cdots + mH \sum_{\{b_0\}} s_i, \tag{2.1}$$

where K is the reduced coupling constant $(K = \beta|J|)$, $\beta(=1/k_B T)$ is a constant, the random variables $A(b_k)$ (± 1) characterize the distribution of the spin interactions, $s(\pm)$ are the Ising spins, the sum runs over the elementary k-dimensional cells (or links) b_k of the lattice. We look for frustration in models with -1 or $+1$ values at $A(b_k)$ and for phase transitions in models with $A(b_k) = +1$. We assume that the interactions are "placed" on the elementary k-dimensional cells b_k of the lattice. This means that the three-spin interactions are placed on the triangles (two-dimensional cells or two-dimensional links) of the triangular lattice T, the four-spin interactions on the squares b_2' of the hypercubic lattice Z^d, etc.

If every $A(b_k) = 1$ $(k = 1,2,\ldots,d)$ the model (2.1) is equivalent to the Griffiths–Wood Ising model, however if $m = 0$ and $A(b_k) = \pm 1$ this is another Ising model but now with frustration. In the case when $k = 1$, and $A(b_1) = +1$ or $A(b_1) = \pm 1$, we have the usual Ising model or the Ising model for spin glasses.

The study of such a model requires the knowledge of the properties of k-dimensional surface in Euclidean space. Therefore we take into account the following definition.

Definition 1: We call surface a connected polyhedron whose covering complex $\overline{K}$ has the following two properties: (a) every branch b_1 of $\overline{K}$ is adjacent to exactly two cells b_2, $\overline{b}_2$ of $\overline{K}$ and their boundary loops have only b_1 in common; and

[a)] Permanent address: Institute of Molecular Physics, Polish Academy of Sciences, Smoluchowskiego 17/19, Poznań, Poland.

 0022-2488/85/071597-03$02.50

(b) to each node there is attached a unique umbrella U (see Ref. 14).

Let us consider property (a). There is a cell $b_2^{\,1}$ adjacent to $b_1^{\,1}$ and $b_1^{\,2}$, the other cell $b_2^{\,2}$ is adjacent to the branches $b_1^{\,2}$ and $b_1^{\,3}$, the next one $b_2^{\,3}$ is adjacent to $b_1^{\,3}$ and $b_1^{\,4}$, etc. We obtain a sequence of branches and cells $\{b_1^{\,1}, b_2^{\,1}, b_1^{\,2}, b_2^{\,2},...\}$ such that

$$\begin{aligned} &b_1^{\,2} \cap b_2^{\,2} = b_1^{\,1}, \\ &b_2^{\,2} \cap b_2^{\,3} = b_1^{\,2}, \\ &\quad\vdots \\ &b_2^{\,l} \cap b_2^{\,l+1} = b_1^{\,l}, \end{aligned} \tag{2.2}$$

which can be finite or infinite. We say that the sequence is a circular system P_1 if

$$b_2^{\,l+1} \equiv b_2^{\,1}. \tag{2.3}$$

Generalizing, we can talk about the k-dimensional circular system P_k if

$$b_k^{\,l+1} \equiv b_k^{\,1}$$

for the sequence $\{b_{k-1}^{\,1}, b_k^{\,1}, b_{k-1}^{\,2},...\}$. We obtain a umbrella U (or generalizing a k-dimensional umbrella U_k) when the circular system P (a k-dimensional circular system P_k) has one common node.

Let us consider the triangular, the square, the cubic, and d-dimensional hypercubic lattices. In the first case the cell b_2 is the triangles. In such a lattice we can obtain both finite or infinite sequences $\{b_1^{\,1}, b_2^{\,1},...\}$. However, the umbrella U can be obtained only if the number of b_2 cells is equal to 6. So, we have a six-cell umbrella such that every one of its sites belongs to two of its cells.

Definition 2: If there is an umbrella U such that every one of its sites belongs to two of its cells, we say that U is compact and write

$$U \equiv U^C.$$

Let us now consider the square lattice. The smallest umbrella in the square lattice consists of four cells but it is impossible to obtain the compact umbrella U^C in this lattice. There are at least four points b_0 which belong exactly to one single cell of the umbrella. We can easily see that the compact umbrella can never appear in the d-dimensional hypercubic lattices, for an arbitrary dimension. In spite of the nonexistence of the U^C umbrellas in the hypercubic lattice, one can find the circular system P_k having a property similar to U^C.

Definition 3: The smallest circular system in the hypercubic lattice constructed from k-dimensional cells b_k ($k = 2,...,d$) such that every one of its sites belongs to two of its cells will be called a k-dimensional plaquette or simply k plaquette L_k (see Ref. 14). This means that the k plaquette L_k exists if there is the sequence

$$\{b_{k-1}^{\,1}, b_k^{\,1}, b_{k-1}^{\,2}, b_k^{\,2},...,b_k^{\,l+1}\}$$

such that

$$b_k^{\,l+1} \equiv b_k^{\,1},$$

and for every point $b_0^{\,P}$ and L_k ($b_0^{\,P} \in L_k$)

$$b_0^{\,P} \in b_k^{\,i} \cap b_k^{\,i+1},$$

we have

$$b_k^{\,i} \cap b_k^{\,i+1} = b_{k-1}^{\,i}, \quad b_k^{\,1} \cap b_k^{\,l} = b_{k-1}^{\,1},$$

with $l = 1,2,...$. The smallest circular system L_k is build of four k-dimensional cells b_k ($l = 1,...,4$). In general, we give the following definition.

Definition 4: The circular system is called a compact circular system P^C when each of its points belong to two cells of this system.

III. TWO MODELS

Let us consider two models. We look for the mechanism of the appearance of the frustration or of the phase transitions.

Example I: Let us take the model described by Eq. (2.1). We see that this Hamiltonian is local gauge invariant, that is, invariant under the following transformation:

$$s_i \rightarrow -s_i, \quad A(b_k) \mapsto -A(b_k), \tag{3.1}$$

with s_i sitting at the site $b_0^{\,i} \in b_k$. As is well known the frustration effect in the Ising model for spin glasses appears when there is an odd number of antiferromagnetic interactions in the plaquette. The definition of the frustration is then that the frustration arises when the spins of the plaquette cannot find a fully satisfactory ground state.[10,11,15]

Let us consider the condition required for the existence of the frustration in general. The distribution of spins of the interactions in the ground state is the most important while dealing with the frustration effect. There are only two configurations: ferromagnetic and antiferromagnetic in the ground state in the Ising model. However, the problem becomes more complicated when the multispin interactions appear. The spins are in the ground state when the energy is $E = E_{\min} = -1$. The problem of finding the frustration of multispin interactions is equivalent to studying a system of n equations:

$$\begin{aligned} &A(b_k^{\,1}) s_1 \cdots s_w = 1, \\ &A(b_k^{\,2}) s_1' \cdots s_w' = 1, \\ &\quad\vdots \\ &A(b_k^{\,n}) s_1 s_2^{(n-1)} \cdots s_w^{(n-1)} = 1, \end{aligned} \tag{3.2}$$

where $A(b_k) s_1 \cdots s_w = E$ is the energy of the ground state for the k link b_k with w-spin interactions. It turns out that there is no frustration in the system described by Eqs. (3.2) if this system of energetical equations is consistent. In this case there is more than one solution and each of the variables s (spins) can take two values $(+1)$ or (-1). This fact means that every spin s can take a position which minimizes the energy of its link b_k. However, in the case when the system is inconsistent, there is no solution and there is a spin which could not minimize the energy of its link b_k. In such a case there is frustration in our system with multispin interactions. Now we are able to give the general necessary condition for the appearance of frustration.

Theorem 1: The necessary condition for the appearance of the frustration in the arbitrary lattice with multispin interactions is the existence of the compact circular system P^C. The proof is based on the arguments given above.

In particular the triangular lattice $P^C \equiv U^C$, the smallest surface in T lattice on which the frustration can arise is a honeycomb built of six triangles. The necessary condition for frustration in the hypercubic lattice Z^d is given by[13] the next theorem.

Theorem 2: In Z^d a frustration can arise if

$$w \leqslant 2^{d-1}, \tag{3.3}$$

where w is the number of spins taking part in a single interaction of b_k.

Proof: If $k = d$, then $b_d{}^1 \cap b_d{}^l = \varnothing$, so there are points (sites of the lattice), which belong only to the single $b_d{}^1$ or $b_d{}^l$ cells (links). The spins are sitting at every site of the cell of Z^d. Therefore, at least one spin of $b_d{}^1$ or $b_d{}^l$ can minimize the energy of their links. In the case when the system of b_k is finite the system of the n energetical equations (3.2) is consistent, but it has more than one solution. Hence there is no frustration. If $k < d - 1$, then

$$b_k{}^1 \cap b_k{}^1 = b_{k-1},$$

we obtain the circular system, and, what is more, it is a compact one. The suitable distribution of the signs of the multispin interactions $\{A(b_k)\}$ gives rise to the inconsistent systems of the energetical equations. This means that there is a frustration effect. As $k \leqslant d - 1$, then every b_k has exactly 2^k corners. Because the spins s are sitting at the corners of b_k, and since b_k represents a single interaction of w spins then

$$w \leqslant 2^k. \qquad \text{QED}$$

We see that these necessary conditions express the topological property of closure. Therefore, one requires the existence of closed k-dimensional surface in the spin system in order that frustration can appear.

Example II: Let us consider the problem of phase transition in the Ising model with positive $w = 2^k$-spin interactions in the hypercubic lattice. In particular, the interactions can be described as follows:

$$-\beta\mathscr{H} = K_k \sum_{\{b_k\}} s_{i_1} \cdots s_{i_w}. \tag{3.4}$$

The partition function is

$$\begin{aligned} Z &= \sum_{\{s\}} \exp\left\{K_k \sum_{\{b_k\}} s_{i_1} \cdots s_{i_w}\right\} \\ &= \sum_{\{s\}} \prod_{\{b_k\}} \cosh K \{1 + s_{i_1} \cdots s_{i_w} \tanh K\}, \end{aligned} \tag{3.5}$$

and we must notice that

$$\sum_{\{s\}} s = 0, \quad \sum_{\{s\}} s^2 = 2. \tag{3.6}$$

Considering the product $s_{i_1} \cdots s_{i_1} \tanh K$ we see that it depends on different graphs in the lattice. However, in the case when $k = d$, this product does not contribute to the partition function because of (3.6) and a lack of closed graphs made of b_d cells. So, we obtain the specific heat in the following form:

$$c_v = a k_B (J_W / k_B T)^2 \operatorname{sech}^2(J_W / k_B T), \tag{3.7}$$

where a is a number of d-dimensional cells b_d in the lattice. We see that in this case there is no phase transition in the lattice. Griffiths and Wood[12] have formulated the necessary condition for the appearance of a phase transition in the Ising model with multispin interactions as a necessity of the existence of closed graphs in the lattices. Our generalization is that the condition for the phase transitions is the existence of the arbitrary compact circular systems P^C or U^C in the lattices. In particular, for a Z^d lattice the phase transition can appear only if a number of spins taking part in a single interaction of b_k cell is

$$w \leqslant 2^{d-1}. \tag{3.8}$$

Taking into account these conditions we can easily show the type of the interactions and/or the sort of lattices in which the phase transition can appear (in the model with Z_2 internal symmetry).

IV. CONCLUSIONS

In this paper we have presented the application of the algebraic topology to statistical mechanics of spin systems. The striking feature is that the necessary condition for the frustration and the phase transition is the same mathematical form (3.3) and (3.8). This means that the topological closure plays a fundamental role in the problems of statistical mechanics. It seems that the knowledge of these conditions might help in finding the solutions of the unsolved problems of statistical mechanics presented by Wu,[16] i.e., in the determination of the critical points for such lattices as a Kagomé lattice or a checkerboard one.

ACKNOWLEDGMENTS

The author would like to thank Professor J. Morkowski and Dr. A. R. Ferchmin for useful discussions. Special thanks are due to Professor M. P. Tosi for his critical comments. She would also like to thank Professor Abdus Salam, the International Atomic Energy Agency, and The United Nations Educational, Scientific, and Cultural Organization for hospitality at the International Centre for Theoretical Physics, Trieste.

This work was supported partially by the Polish Academy of Sciences under Project No. MR-I.9.

[1]G. Choquet, *Topology* (Academic, New York, 1966).

[2]K. Moriyasu, *An Elementary Primer for Gauge Theories* (World Scientific, Singapore, 1983); P. Orland, "Disorder, frustration and semiclassical calcualtions in lattice gauge theories," Imperial/TP/83-84/49; F. A. Bais, Nucl. Phys. B **170**, 32 (1980).

[3]F. Englert and P. Windey, Phys. Rev. D **14**, 2728 (1976).

[4]M. Kléman, *Points. Lignes. Parois* (Les Editors de Physique, Orsay, 1977); V. Poenaru and G. Toulouse, J. Phys. **38**, 887 (1977); J. Bricmont and J. Fröhlich, Nucl. Phys. B **23**, 407 (1974).

[5]G. W. Whitehead, *Homotopy Theory* (M.I.T. Press, Cambridge, 1966).

[6]S. F. Edwards, "Path integrals and polymer problems," in *Path Integrals and Their Applications in Quantum Statistical and Solid State Physics*, edited by G. J. Papodopoulos and J. T. Devreese (Plenum, New York, 1977); R. Ball and M. L. Mehta, J. Phys. **42**, 1193 (1981).

[7]D. Rolfsen, *Knots and Links* (University of British Columbia, Vancouver, 1976).

[8]F. Wegner, J. Math. Phys. **12**, 2259 (1971).

[9]R. Balian, J. M. Drouffe, and C. Itzykson, Phys. Rev. D **11**, 2098 (1975).

[10]G. Toulouse, Comm. Phys. **2**, 115 (1977).

[11]G. Toulouse, Phys. Rep. **49**, 267 (1979).

[12]H. P. Griffiths and D. W. Wood, J. Phys. C **6**, 2533 (1973).

[13]H. Nencka-Ficek, Physica A **114**, 254 (1982).

[14]S. Lefschetz, *Application of Algebraic Topology* (Springer, New York, 1975).

[15]G. Toulouse, in *Proceedings of the VIII International Colloquium on Group Theoretical Methods in Physics*, Kiriat Anarim, Israel, 1979, edited by L. P. Horowitz and Y. Ne'eman (Tel-Aviv University, Tel-Aviv, 1979).

[16]F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).

Random walk representations for spinor and vector propagators

Ted Jacobson[a)]
Department of Physics, University of California, Santa Barbara, California 93106

(Received 12 December 1984; accepted for publication 15 March 1985)

A Euclid-invariant random walk representation for spin-$\frac{1}{2}$ and vector propagators is developed in analogy to random walks with internal states.

I. INTRODUCTION

The aim of this paper is to develop a Euclid-invariant random walk representation for spin-$\frac{1}{2}$ and vector propagators. We shall consider random walks in which rotational symmetry is maintained, and all steps have the *same length*. Such a discretization is of interest first because it permits a Euclid-invariant elimination of singularities of Green functions and hence of divergences in Feynman diagrams. Of course, invariant regularizations are not hard to come by, so this alone does not particularly recommend the awkward abandoning of continuity. Looking beyond regularization, however, to replace quantum fields by Feynman diagrams, which are in turn replaced by discrete networks of elementary steps, provides at least a framework in which one can imagine doing without the background continuum space-time altogether.[1,2] This idea has been one of the motivations behind the investigations reported here.

That the following construction treats a Euclidean rather than a Minkowskian position space surely weakens its prospects as a conceptual substitute for continuum space-time. Euclideanization seems unavoidable in this context for at least three reasons. First, the random walk representation fails to exist for Green functions of second-order hyperbolic operators. Second, going to Euclidean space seems to be the only local way to arrive at the Feynman propagators to be used in amplitudes, as opposed to the retarded and advanced propagators. Third, the basic idea being to build diagrams out of small steps, how does one specify "smallness"? One way is to choose a timelike vector t and take steps whose spatial part relative to t is small. Of course this smallness depends on t, and a theory built upon this notion would not be Lorentz invariant.[3] Another possibility is to take steps small in the Minkowski metric, however, this set of "small' steps forms a noncompact manifold (a hyperboloid) with infinite Lorentz invariant measure. In Euclidean space on the other hand, the steps of a small length span the (finite measure) surface of a sphere. For these reasons the analysis has been carried out in a Euclidean context.

Spinor and vector propagators have matrix indices that label polarization states. A random walk construction for these propagators might thus plausibly resemble a random walk with internal states.[4] Such a stochastic process is specified by giving a distribution matrix $\rho_{ab}(x,x')$ indicating the probability of a transition from internal state b to state a, and position x' to x. For a probability interpretation to hold, $\rho_{ab}(x,x')$ should be a non-negative real number for all a, b, x, and x', and should satisfy the identity

$$\sum_a \int dx\, \rho_{ab}(x,x') = 1$$

(for every x' and b) which asserts that the total probability for some transition to occur is unity.

In our constructions, $\rho_{ab}(x,x')$ will depend only on $x - x'$ and will be nonzero only when $x - x'$ has a fixed length τ, so it can be given as a function of the unit vectors $n = (x - x')/\tau$. *Only for the scalar will the probability interpretation remain intact.* For the spinor, $\rho_{ab}(n)$ will contain imaginary numbers, while for the vector it will be real but not, in general, non-negative.

II. SPIN-$\frac{1}{2}$ GREEN FUNCTION

Let us consider the Dirac operator $\partial\!\!\!/ + m$ in d Euclidean dimensions, where $\partial\!\!\!/ := \gamma^i \partial_i$ and $\gamma^i (i = 1,\ldots,d)$ are Hermitian matrices generating the Euclidean Clifford algebra

$$\gamma^i\gamma^j + \gamma^j\gamma^i = 2\delta^{ij}.$$

We seek an approximation to the Green function of $(\partial\!\!\!/ + m)$ in the form of a path integral over polygonal paths built with steps of length τ, and which in the limit $\tau \to 0$ agrees with $(\partial\!\!\!/ + m)^{-1}$. The key is the identity

$$(\partial\!\!\!/ + m)^{-1} = \int_0^\infty ds\, e^{-s(\partial\!\!\!/ + m)} = \lim_{\tau\to 0} \tau \sum_{N=1}^{\infty} e^{-N\tau(\partial\!\!\!/ + m)}, \tag{1}$$

which is valid since the Hermitian part of $(\partial\!\!\!/ + m)$ (i.e., m) is positive definite.

The operator $\exp[-s(\partial\!\!\!/ + m)]$ can be regarded as the evolution operator corresponding to the equation

$$\partial_s \psi = -(\partial\!\!\!/ + m)\psi. \tag{2}$$

Our method is to derive a path integral for the kernel of this evolution operator by finite differencing (2), and then to sum as in the last expression of (1) over the total number $N = s/\tau$ of steps.

Equation (2) can be finite differenced in a rotationally symmetric fashion by first rewriting it in the form

$$\left\{\int d\hat{n}(1 + \gamma^i n_i)(d^{-1}\partial_s + n^i\partial_i)\right\}\psi = -d^{-1}m\psi. \tag{3}$$

Here the integral is over the vectors n^i on the unit $(d-1)$ sphere, with respect to the normalized rotationally invariant measure $d\hat{n}$. [To verify (3) note that $\int d\hat{n} = 1$ and that, by symmetry, $\int d\hat{n}\, n^i = 0$ and $\int d\hat{n}\, n_i n^j \propto \delta^j_i$; that the (annoying but) correct factor is d^{-1} is seen by taking the trace of both sides.]

[a)] Research completed while the author was a Chercheur Associé au Centre National de la Recherche Scientifique at l'Observatoire de Meudon and l'Institut Henri Poincaré, Paris France.

 0022-2488/85/071600-05$02.50

Now with small τ we approximate the directional derivative as a finite difference

$$\tau(d^{-1}\partial_s + n^j\partial_j)\psi(s,x)$$
$$= \psi(s,x) - \psi(s - d^{-1}\tau, x - n\tau) + O(\tau^2), \tag{4}$$

and substitute (4) in (3) to obtain after rearranging

$$\psi(s,x) = \int d\hat{n}\,\rho(n)\psi(s - d^{-1}\tau, x - n\tau) + O(\tau^2), \tag{5}$$

where

$$\rho(\mathrm{n}) := e^{-m\tau d^{-1}}(1 + \not{n}). \tag{6}$$

By iterating (5) it follows that, neglecting the $O(\tau^2)$ contribution, the kernel $(x|\exp[-Nd^{-1}\tau(\not{\partial} + m)]|0)$ of the evolution operator can be approximately written in the form

$$K_N := \int \prod_{a=1}^{N} d\hat{n}_a\,\rho(n_N)\cdots\rho(n_1)\delta\Big(x - \tau\sum_{a=1}^{N} n_a\Big), \tag{7}$$

which is in fact an integral over polygonal paths with N steps of length τ. It is this approximation of the exact kernel that will be substituted in (1) to arrive at an expression for $(\not{\partial} + m)^{-1}$.

It will be convenient to rewrite K_N as a single d-dimensional integral. This is done by substituting the Fourier representation of the delta function[5] and interchanging the integrations over $\Pi_a d\omega_a$ and $d^d k$. This yields

$$K_n = (2\pi)^{-d}\int d^d k\, e^{ikx} A^N(k),$$

where

$$A(k) := \int d\hat{n}\,\rho(n)e^{-ikn\tau}$$
$$= e^{-m\tau d^{-1}}\Gamma(\nu + 1)(2/\kappa)^\nu[J_\nu(\kappa) - i\hat{\not{k}}J_{\nu+1}(\kappa)]$$
$$= 1 - (i\tau d^{-1})(\not{k} - im) + O(\tau^2), \tag{8}$$

with $\nu := d/2 - 1$, $\kappa := |k|\tau$, $|k| := (k^2)^{1/2}$, $\hat{k} := k/|k|$, and J_ν is a Bessel function having the expansion[6]

$$J_\nu(z) = \Big(\frac{z}{2}\Big)^\nu \sum_{m=0}^{\infty}\frac{(-z^2/4)^m}{m!\Gamma(\nu + 1 + m)}.$$

Now the sum over N is easily carried out:

$$\sum_{N=p}^{\infty} K_N = \sum_{N=p}^{\infty}(2\pi)^{-d}\int d^d k\, e^{ikx}A^N(k) \tag{9}$$
$$= (2\pi)^{-d}\int d^d k\, e^{ikx}A^p(k)[1 - A(k)]^{-1}. \tag{10}$$

In the limit $\tau \to 0$, the exact Green function is recovered when the sum is multiplied by τd^{-1} [cf. (8)]:

$$\lim_{\tau\to 0}(\tau d^{-1})\sum_{N=p}^{\infty} K_N = (2\pi)^{-d}\int d^d k\, e^{ikx}[i\not{k} + m]^{-1}$$
$$= (\mathrm{x}|(\not{\partial} + m)^{-1}|0). \tag{11}$$

Note that in the limit $\tau \to 0$, the value of p (the minimum number of steps) is inconsequential. This occurs because the entire sum is multiplied by τ, so that any finite collection of (finite) terms contributes nothing in the limit. The choice of p will be discussed further in Sec. V.

The step from (9) to (10) involves interchanging an infinite series and integral, as well as summing the series, while in (11) the limit $\tau \to 0$ has been interchanged with the integral over $d^d k$. It must be checked that these manipulations are justified. Taking into account[5] the (implicit) factor $\exp(-\eta k^2)$, the interchanges can be justified by examining properties of uniform convergence and large k behavior of the integrand.[7] As for the series, it is of the form $\Sigma_{N=p}^{\infty} A^N$ with A a matrix given by (8). This series converges to $A^p[1 - A]^{-1}$ provided $A^N \to 0$ as $\mathrm{N} \to \infty$. Since $A(k)$ is diagonalizable, we will have $A^N(\mathrm{k}) \to 0$ provided its eigenvalues have modulus less than unity. The eigenvalues of $\hat{\not{k}}$ are ± 1, so the moduli are given by

$$|\lambda_\pm(k)|^2$$
$$= e^{-2m\tau d^{-1}}\{[\Gamma(\nu + 1)(2/\kappa)^\nu]^2(J_\nu^2(\kappa) + J_{\nu+1}^2(\kappa))\}$$
$$= e^{-2m\tau d^{-1}}f(\kappa).$$

Now $f(0) = 1$ and $f(\kappa)$ is monotonic decreasing ($f'(\kappa) = -[\Gamma(\nu + 1)2^\nu]^2\, 2(2\nu + 1)\kappa^{-2\nu-1}J_{\nu+1}^2(\kappa) \leqslant 0$) so we have $|\lambda_\pm| \leqslant \exp(-m\tau d^{-1})$, with equality only when $k = 0$. Hence the series converges.[8]

Let us discuss the results of this section. The Green function of $(\not{\partial} + m)$ has been represented as an integral over paths with an arbitrary number of steps all of length τ. The path integral is a good approximation to the continuum Green function when $|x| \gg \tau$, and converges to it in the limit $\tau \to 0$. The "weight" or "amplitude" for a given path is the ordered product of the matrices $\rho(n_a) = \exp(-m\tau d^{-1})(1 + \not{n}_a)$; because permutation of the sequence $n_1, \ldots, n_N$ does not change the sum $\Sigma_{a=1}^{N} n_a$, in fact only the symmetrized product enters. When $m > 0$, the weight of paths with more steps is exponentially damped.

Since $\int d\hat{n}(1 + \not{n}) = 1$, $1 + \not{n}$ plays a role analogous to that of a distribution matrix for a random walk with internal states (cf. Sec. I). The polarization state is correlated with the direction vector n in the following sense: A step in direction n puts the particle in a polarization state that has zero amplitude to travel in the opposite direction $-n$, since $(1 - \not{n})(1 + \not{n}) = 0$. The Fourier transform (with respect to x) of the path integral for finite τ can, somewhat surprisingly, be evaluated *exactly* and is given by $A^p(k)[1 - A(k)]^{-1}$, with $A(k)$ as defined in Eq. (8).

III. SCALAR GREEN FUNCTION

We consider now the second-order, scalar operator $-\partial^2 + m^2$, where $\partial^2 = \partial^a\partial_a$ is the d-dimensional Laplacian. Following the development of the preceding section we make use of the identity

$$(-\partial^2 + m^2)^{-1} = \int_0^\infty ds\, e^{-s(-\partial^2 + m^2)}. \tag{12}$$

The operator $\exp[s(\partial^2 - m^2)]$ can be regarded as the evolution operator corresponding to the equation

$$\partial_s\varphi = (\partial^2 - m^2)\varphi; \tag{13}$$

to proceed as before one should find a finite differencing of this equation that will yield a path integral of the desired form. This could be done using the fact that $\partial^2\varphi(x)$ is proportional to the difference between $\varphi(x)$ and the average of φ over an infinitesimal sphere surrounding x. Proceeding along these lines, one would demonstrate that the evolution operator of the diffusionlike equation (13) can be obtained from a random walk, a well-known result.[9]

Our path integral for the spin-$\frac{1}{2}$ case [(7) and (9)] is much like a random walk, and in fact it is easiest simply to adapt it by inspection to the case of the scalar. In a free, scalar random walk all steps are equally weighted, so in place of the matrix $1 + \not{n}$ appearing in $\rho(n)$ we expect to have the number unity. Furthermore, as indicated by (12), $\exp(-m\tau d^{-1})$ should be replaced by $\exp[-m^2\tau^2(2d)^{-1}]$ [the choice of $(2d)^{-1}$ ensures that m^2 will be the mass term in the corresponding differential operator; cf. (14) and (15)]. In place of K_N of Eq. (7) one now has the N-step scalar evolution operator

$$\begin{aligned} G_N &= (e^{-m^2\tau^2(2d)^{-1}})^N \int \prod_{a=1}^{N} d\hat{n}_a\, \delta\Big(x - \tau \sum_{a=1}^{N} n_a\Big) \\ &= (2\pi)^{-d} \int d^dk\, e^{ikx} a^N(k), \end{aligned}$$

with

$$\begin{aligned} a(k) &= e^{-m^2\tau^2(2d)^{-1}} \int d\hat{n}\, e^{-ikn\tau} \\ &= e^{-m^2\tau^2(2d)^{-1}} \Gamma(\nu+1)(2/\kappa)^\nu J_\nu(\kappa) \\ &= 1-(k^2+m^2)\tau^2(2d)^{-1} + O(\tau^4), \end{aligned} \tag{14}$$

where ν,κ are as defined after (8).

Summing now over N,

$$\sum_{N=p}^{\infty} G_N = (2\pi)^{-d} \int d^dk\, e^{ikx} a^p(k)[1-a(k)]^{-1} \tag{15}$$

so that

$$\begin{aligned} \lim_{\tau\to 0} \tau^2(2d)^{-1} \sum_{N=p}^{\infty} G_N &= (2\pi)^{-d} \int d^dk\, e^{ikx}(k^2+m^2)^{-1} \\ &= (-\partial^2+m^2)^{-1}. \end{aligned} \tag{16}$$

The same comments[5,7] as in the spin-$\frac{1}{2}$ case apply to justify the steps in (15) and (16). The series $\Sigma\, a^N$ converges since $|a(k)| < \exp(-m^2\tau^2(2d)^{-1})$, thanks to a known inequality on Bessel functions[6] $|\Gamma(\nu+1)(2/\kappa)^\nu J_\nu(\kappa)| \leqslant 1$ for κ real.

IV. VECTOR GREEN FUNCTION

We turn now to the vector operators

$$\delta_{ij}(-\partial^2+m^2) + (1-\xi^{-1})\partial_i\, \partial_j.$$

The case $\xi = 1$ can be taken over directly from the scalar if one merely multiplies everything by δ_{ij}. In this case the propagation is independent of the polarization, and in the limit $m \to 0$ one obtains the (Euclidean) photon propagator in the Feynman gauge. In the general case the propagator has the form

$$(k^2+m^2)^{-1}[\delta_{ij} - (1-\xi)k_ik_j/(k^2+\xi m^2)]. \tag{17}$$

Can this propagator be obtained in the limit $\tau \to 0$ from some random walk-type path integral? To answer this we shall proceed by guesswork.

The step distribution matrix appearing in the sought-after path integral must have two vector indices, so that, assuming it must be built with unit vectors, the only possibilities are $\rho_{ij}(n) = \alpha\delta_{ij} + \beta n_i n_j$, with α and β some constants. From the previous analyses it is clear that to zeroth order in τ we must have $\int d\hat{n}\, \rho_{ij}(n) = \delta_{ij}$. Treating first the *massless case* for simplicity, ρ_{ij} is independent of τ, so this must hold identically and ρ_{ij} must have the form

$$\rho_{ij}(n) = \alpha\delta_{ij} + d(1-\alpha)n_in_j. \tag{18}$$

Our strategy is simply to try all of these.

Once again, the N-step propagator is easily computed to be

$$K_{ijN} = (2\pi)^{-d} \int d^dk\, e^{ikx} A_{ij}^N(k), \tag{19}$$

where now

$$\begin{aligned} A_{ij}(k) &= \int d\hat{n}(\alpha + d(1-\alpha)n_in_j)e^{-ikn\tau} \\ &= \Gamma(\nu+1)(2/\kappa)^\nu[\mathrm{J}_\nu + (1-\alpha)\mathrm{J}_{\nu+2} \\ &\quad - d(1-\alpha)\mathrm{J}_{\nu+2}\hat{k}_i\hat{k}_j] \\ &= 1 - \Big(\frac{2\alpha+d}{2d(d+2)}\Big)\kappa^2 - \Big(\frac{1-\alpha}{d+2}\Big)\kappa_i\kappa_j + O(\kappa^4), \end{aligned} \tag{20}$$

with $\nu := d/2 - 1$ and $\kappa_i = k_i\tau$. The implicit argument of the Bessel functions is κ, and terms without indices are understood to be multiplied by δ_{ij}.

If the sum over N converges, the Fourier transform of the path integral/sum will be given by $A_{ij}^p(k)[1-A_{ij}(k)]^{-1}$, as in (10) and (15), and the limit $\tau \to 0$ will be determined by

$$\begin{aligned} &\lim_{\tau\to 0} \tau^2[1-A_{ij}(k)]^{-1} \\ &= \Big(\frac{2d(d+2)}{2\alpha+d}\Big) k^{-2}\Big[\delta_{ij} - \Big(\frac{2d(1-\alpha)}{3d-2\alpha(d-1)}\Big)\hat{k}_i\hat{k}_j\Big]. \end{aligned} \tag{21}$$

Thus, modulo convergence, we can obtain the propagators (17) with $m = 0$ and $1-\xi = 2d(1-\alpha)/(3d-2\alpha(d-1))$.

The sum over N will converge provided the moduli of the eigenvalues of $A_{ij}(\mathrm{k})$ are less than unity. Since the eigenvalues of $\hat{k}_i\hat{k}_j$ are 0 (for $d \geqslant 2$) and 1, the eigenvalues of A_{ij} [Eq. (20)] are given by

$$\begin{aligned} \lambda_\perp &= \Gamma(\nu+1)(2/\kappa)^\nu[J_\nu + (1-\alpha)J_{\nu+2}] \\ &= 1 - [(2\alpha+d)/2d(d+2)]\kappa^2 + \cdots, \\ \lambda_\parallel &= \Gamma(\nu+1)(2/\kappa)^\nu[J_\nu - (d-1)(1-\alpha)J_{\nu+2}] \\ &= 1 - [(3d-2\alpha(d-1))/2d(d+2)]\kappa^2 + \cdots. \end{aligned} \tag{22}$$

These eigenvalues *do not* have moduli less than unity for arbitrary of α, and it is difficult to determine in general the allowed range of α. A necessary but not sufficient restriction is easily obtained by looking at the eigenvalues near $\kappa = 0$: $\lambda(0) = 1$, so the coefficients of κ^2 in the expansions (22) must be < 0 which yields the constraint

$$-(d/2) < \alpha < 3d/2(d-1). \tag{23}$$

This ensures that $|\lambda|$ starts out diminishing as κ grows from zero. The problem is that at some finite value of κ, $|\lambda|$ may nevertheless exceed unity. For example, if α is too large and negative, $\lambda_\parallel$ will go below -1 as $J_{\nu+2}$ reaches its first maximum. As one can only proceed in particular cases, let us investigate the choices $\alpha = 0$ and $\alpha = (1 - d^{-1})^{-1}$.

With $\alpha = 0$ the step distribution matrix is $\rho_{ij}(n) = d n_i n_j$, corresponding to "no transverse propagation." The eigenvalues of A_{ij} are [using (22)]

$$\lambda_\perp = \Gamma(\nu+2)(2/\kappa)^{\nu+1} J_{\nu+1},$$
$$\lambda_\parallel = \Gamma(\nu+1)(2/\kappa)^{\nu}[J_\nu - (d-1)J_{\nu+2}].$$

It follows from the known inequality $|\Gamma(\nu+1)(2/\kappa)^\nu J_\nu(\kappa)| \leqslant 1$ that $|\lambda_\perp| \leqslant 1$, and equality holds only for $\kappa = 0$. I have determined that $|\lambda_\parallel| \leqslant 1$ for $d \leqslant 4$, and it seems very likely that it is true for all d, although it has not been demonstrated. At least for $d \leqslant 4$ then, the scheme with $\alpha = 0$ is convergent, and from (19) and (21) we have

$$[2(d+2)]^{-1} \lim_{\tau \to 0} \tau^2 \sum_{N=p}^{\infty} K_{ijN}$$
$$= (2\pi)^{-d} \int d^d k \, e^{ikx} k^{-2}\Big(\delta_{ij} - \frac{2}{3}\hat{k}_i \hat{k}_j\Big),$$

corresponding to the propagator (17) with $m = 0$ and $\xi = \frac{1}{3}$. It is curious that ξ is independent of d. Note also that whereas step by step there is no transverse propagation ($\rho_{ij} \sim n_i n_j$), in terms of wave vectors transverse polarizations are propagated with greater weight than longitudinal polarizations! The reason for this remains hidden in the mysteries of the sum over paths of arbitrary length.

With $\alpha = d(d-1)^{-1}$ ($d \geqslant 2$), ρ_{ij} takes the form $\rho_{ij}(n) = d(d-1)^{-1}(\delta_{ij} - n_i n_j)$, corresponding to "no longitudinal propagation." This α satisfies the constraint (23), and the eigenvalues (22) of the corresponding $A_{ij}(k)$ are

$$\lambda_\perp = \Gamma(\nu+1)(2/\kappa)^\nu[J_\nu - (d-1)^{-1}J_{\nu+2}],$$
$$\lambda_\parallel = \Gamma(\nu+2)(2/\kappa)^{\nu+1}J_{\nu+1}.$$

Now $|\lambda_\parallel| \leqslant 1$, and I have checked that $|\lambda_\perp| \leqslant 1$ for $2 \leqslant d < 4$; once again it seems likely that $|\lambda_\perp| \leqslant 1$ for all $d \geqslant 2$ but is has not been demonstrated. At least for $2 \leqslant d \leqslant 4$ then, the scheme with $\alpha = d(d-1)^{-1}$ is convergent and from (19) and (21) we have

$$\left[\frac{d+1}{2(d-1)(d+2)}\right] \lim_{\tau \to 0} \sum_{N=p}^{\infty} K_{ijN}$$
$$= (2\pi)^{-d} \int d^d k \, e^{ikx} k^{-2}[\delta_{ij} + 2(d-1)^{-1}\hat{k}_i \hat{k}_j],$$

corresponding to the propagator (17) with $m = 0$ and $1 - \xi = 2(d-1)^{-1}$. In this case ξ does depend on d and indeed $\xi \to 1$ as $d \to \infty$. While it is perhaps plausible that as $d \to \infty$ the effect of excluding one (longitudinal) step direction becomes negligible, it is difficult to see for finite d why, in terms of wave vectors, the longitudinal polarization should propagate with greater weight than the transverse.

Finally, the restriction to zero mass in the above discussion is removed by multiplying $\rho_{ij}(n)$ (18) with $\exp(-m^2\tau^2\gamma)$. To obtain the propagator of the form (17), γ must be set equal to the coefficient of k^2 in $A_{ij}(k)$ (20), i.e., $\gamma = (2\alpha + d)/2d(d+2)$. Any choice for α that is convergent with $m = 0$ will converge with $m \neq 0$ (since the eigenvalues of A_{ij} will only be smaller) and, depending on the magnitude of $m^2\tau^2\gamma$, some otherwise divergent α will converge.

V. REGULARITY OF THE DISCRETE PROPAGATORS

The exact continuum scalar, spinor, and vector Green functions are singular at the origin. This can be seen either directly from the form of the equation $\mathscr{D}G(x,x') = \delta(x,x')$ that defines them ($\mathscr{D}$ is a differential operator) or from their Fourier representations. The spinor and vector cases are obtained by applying a differential operator to the scalar Green function $\int d^d k \, e^{ikx}(k^2 + m^2)^{-1}$, which with $x = 0$ reads $\sim \int_0^\infty dk \, k^{d-1}(k^2 + m^2)^{-1}$. This integral is divergent except for the case $d = 1$ (in which case the Green function is not differentiable at $x = 0$).

Having introduced the discrete step length τ, our approximations to the continuum Green functions can be made regular everywhere by a suitable choice of p, the minimum number of steps in the path integrals. We now indicate how this comes about.

The Fourier transform of the discrete propagators $\Sigma_{N=p}^{\infty} K_N$ is of the form $A^p(k)[1 - A(k)]^{-1}$, with $A(k)$ given in (8), (14), and (20) for the spinor, scalar, and vector cases, respectively. Convergence of the integral over k depends in general on the oscillations of e^{ikx} and $A(k)$. For simplicity, let us consider only the conditions for *absolute* convergence. From the asymptotic behavior $J_\nu(k) \sim O(k^{-1/2})$ of the Bessel functions, it follows that $A(k) \sim O(k^{-\nu-1/2}) = O(k^{-(d-1)/2})$ as $k \to \infty$, so that $A^p(k) \times [1 - A(k)]^{-1} \sim O(k^{-p(d-1)/2})$. The integration measure being $\sim dk \, k^{d-1}$, the integral will converge absolutely provided $d - 1 - p(d-1)/2 < -1$, i.e., $p > 2d/(d-1)$. As long as p satisfies this inequality the discrete propagators will necessarily be everywhere regular. Due to oscillations of the integrand, they may in certain cases be regular for lesser values of p.

In one dimension $2d/(d-1)$ is divergent, so for no choice of p are the propagators regular. This is because the set of points that can be reached from the origin with steps of length τ span a lattice in one dimension, so that our propagators are singular at every lattice point. (This case could be regularized by employing a Kronecker delta rather than a Dirac delta function to enforce the constraint $x = \tau\Sigma_{a=1}^{N} n_a$ in the path integral.) For $d \geqslant 4$ the choice $p \geqslant 3$ yields everywhere regular propagators, whereas $d = 2$ or 3 requires $p \geqslant 4$ or 5, respectively.

VI. DISCUSSION

With the random walk representations obtained here, one can consider carrying out the program mentioned in the introduction. That is, given the expansion in Feyman diagrams of a quantum field theory, one can replace each line by an integral over the discrete paths connecting its endpoints, thus obtaining a Euclid-invariant discretization of the quantum field theory.

The usual perturbation expansion is recovered if one performs the sum over each line separately, going to the limit $\tau \to 0$ *before* applying Feynman rules. Other summations can now be considered, however. For example, rather than expanding in the coupling constant or the topology of the diagrams, one might sum first over all diagrams (networks) with a fixed total number N of linking steps and then sum over N, as we have done for single lines. I have not determined whether such a resummation is even well-defined, much less what it yields, but merely note that it is an interesting possibility arising only because the diagrams are broken down into pieces smaller than usual.

As for application to quantum field theories of physical interest, at least two potential problems arise. One problem is that the relation to Minkowski field theory is not evident. Consider, for example, our Euclidean scalar propagator given in Eq. (15), $a^p(k)[1 - a(k)]^{-1}$. This expression has a countable infinity of poles in the complex k^2 plane, in contrast to $(k^2 + m^2)^{-1}$, which has only 2. Furthermore, the numerator $a^p(k)$ grows exponentially as $\exp(p|\mathrm{Im}\, k|)$. It remains to be seen whether our propagator is nevertheless related by analytic continuation to something that converges in the limit $\tau \to 0$ to the Minkowski propagator.

Probably a much more serious problem is that our discretization destroys gauge invariance if applied, for example, to Euclidean quantum electrodynamics. The finite τ amplitudes do not satisfy Ward identities, even at zeroth order in the coupling constant, and in particular there is nothing to prevent the photon from acquiring a mass. A lattice gauge theory preserves gauge invariance by attaching the gauge potential [or rather $\exp(ie\int A_\mu\, dx^\mu)$] to lattice *links* rather than points. Lacking a background lattice we cannot adopt this approach.

ACKNOWLEDGMENTS

This research was begun at the University of Texas at Austin, supported by National Science Foundation Grant No. PHY 84-04931, and supported in part by National Science Foundation Grant No. PHY 81-07384.

[1]Something like this idea was the subject of the "spacetime code" papers of D. Finkelstein. See D. Finkelstein, Phys. Rev. **184**, 1261 (1969); **D5**, 320 (1972); **D5**, 2422 (1972); **D9**, 2219 (1979); and D. Finkelstein, G. Frye, and L. Susskind, Phys. Rev. **D9**, 2231 (1974).

[2]K. Symanzik developed an interpretation of Euclidean quantum field theory as the statistical mechanics of Brownian particles, replacing fields with particles executing random walks but maintaining continuity and a background (Euclidean) space. See K. Symanzik, in *Local Quantum Theory*, edited by R. Jost (Academic New York, 1969), pp. 152–226.

[3]Steps in this direction were taken in T. Jacobson, J. Phys. A **17**, 2433 (1984).

[4]E. W. Montroll and B. J. West, in *Studies in Statistical Mechanics, Vol.* 7, *Fluctuation Phenomena*, edited by E. W. Montroll and J. L. Lebowitz (North-Holland, Amsterdam, 1979), Chap. 2, Sec. 4.4, p. 109.

[5]To put the following calculations on a precise mathematical basis we interpret the δ function in (7) to be the highly peaked Gaussian of width $\sim \eta^{1/2}$ whose Fourier transform is $\exp(ikx - \eta k^2)$. At the end of all other computations the limit $\eta \to 0$ is taken.

[6]M. Abramowitz and A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).

[7]The definitions of uniform convergence and relevant theorems are given, for example, in T. M. Apostol, *Calculus* (Xerox, Lexington, MA, 1967), Vol. 1. The theorems concerning interchange of series or limits with integrals apply to integrals with finite range of integration, and must be supplemented with arguments invoking the large k behavior of $\exp(-\eta k^2)$.

[8]The case $m = 0$ can be defined by the limit $m \to 0$.

[9]S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).

Abelian integrals and the reduction method for an integrable Hamiltonian system

L. Gagnon, J. Harnad, and P. Winternitz
Centre de Recherche de Mathématiques Appliquées, Université de Montréal, Montréal, Québec, Canada H3C 3J7

J. Hurtubise
Département de Mathématiques, Université du Québec à Montréal, Montréal, Québec, Canada H3C 3P8

(Received 20 September 1984; accepted for publication 25 January 1985)

A classical finite-dimensional integrable Hamiltonian system, corresponding to the motion of a particle constrained to an n-dimensional sphere $\Sigma_{\mu=0}^{n} x_\mu^2 = 1$, with the Hamiltonian $H = \Sigma_\mu(\frac{1}{2}y_\mu^2 + u_\mu^2/x_\mu^2 + \epsilon\alpha_\mu x_\mu^2)$ (where u_μ, α_μ, and ϵ are constants and y_μ are the momenta conjugate to x_μ), is integrated using several different methods. These are the following: (1) The projection of geodesic (free) flow on a larger space, namely the sphere S^{2n+1} (for $\epsilon = 0$). The flow is obtained in terms of elementary functions. (2) Separation of variables in the Hamilton–Jacobi equation in elliptic coordinates or, alternatively, the use of a complete set of integrals of motion in involution to reduce Hamilton's equations to quadratures. The flow is obtained in terms of Abelian integrals which are then inverted in terms of generalized θ functions. The relation between the different methods and results is clarified using methods of algebraic geometry, in particular the geometry of quadrics.

I. INTRODUCTION

A certain unifying geometrical and algebraic structure has been found at the base of a wide variety of integrable Hamiltonian systems. These include finite-dimensional systems such as the Toda lattice, the Calogero–Marchiori model, and numerous similar systems[1–4] as well as several families of field theories such as the Korteweg–de Vries (KdV) equation, sine-Gordon system, etc.

This structure manifests itself in two distinct ways: first, at the level of the phase space, the invariants of motion, and the structure of Hamilton's equations; second, in the actual determination of the flow. The first may be described as follows: there exists some primordial system in a larger space which is either free (i.e., geodesic for some Riemannian metric) or has very simple dynamics such as harmonic oscillator motion and which possesses a very large invariance group. Reduction of the system through the Marsden–Weinstein reduction theorem[5] (i.e., by restriction to the level sets of certain of the invariants of motion) frequently leaves a residue of this original system, providing, for example, a complete set of commuting integrals which render the reduced system completely integrable. In the case where the original system involves biinvariant flow on a Lie group, the reduced system lives in the dual to a part of the Lie algebra and the Kirillov–Poisson structure gives rise to reduced Hamiltonian equations of the Lax type through the Adler–Kostant–Symes theorem.[6–9] For finite-dimensional systems, with finite-dimensional symmetry groups, the flow can be directly determined by a suitable projection of the original flow.[1,4,8–10]

However, for many interesting systems, whether finite (e.g., periodic Toda lattice) or infinite dimensional (e.g., KdV), the underlying symmetry is of the Kac–Moody variety,[7–10] and thus involves infinite-dimensional reduction. Since the notion of exponentiation and projection in the corresponding groups is not yet well understood, this approach does not lead in the same way to a determination of the flow.

On the other hand, geometric structure appears in these systems in another form, namely, via the algebraic geometry of curves and their Jacobi varieties.[7,11–14] In the cases involving the Kirillov–Poisson structure, one obtains a spectral curve from a Lax pair associated to the system, and the flow can be linearized in terms of the Jacobi variety of the curve. Although the flow is thus explicitly determined, it is difficult to see the link between the curve and the symmetries of the primordial system.

In a more classical vein, the presence of algebraic curves is implicit in the Abelian integrals that appear in the solution of the Hamilton–Jacobi equations for many classical Hamiltonian systems, through separation of variables in elliptic coordinates.[14–16] Moreover, there is a well-developed theory relating separation of variables to isometry groups.[17–24] This suggests that a more direct link could be established between the geometric notion of reduction by symmetries and the algebraic geometric linearization of flows on Jacobi varieties.

The purpose of the present work is to establish this link as clearly as possible for a rather simple finite-dimensional system that has been known and studied since the nineteenth century: the so-called Rosochatius system.[16,25] Actually, most of our attention will be confined to a particular special case of this system which lends itself to the most complete geometrical interpretation in terms of projections of geodesic motion, although a part of the analysis is applicable to the general system. For another approach which relates to the Lie algebraic structure underlying the Rosochatius system, see Ref. 26.

Let $\{x^\mu\}$, $\mu = 0,\ldots,n$ denote standard coordinates in $\mathbb{R}^{n+1}$ and $\{x^\mu, y_\mu\}$ the associated canonical coordinates on $T^*\mathbb{R}^{n+1}$. The configuration space of the system is the unit sphere $S^n \in \mathbb{R}^{n+1}$ and the phase space $T^*S^n \subset T^*\mathbb{R}^{n+1}$, defined by

 0022-2488/85/071605-08$02.50

$$\sum_{\mu=0}^{n} x^{\mu^2} = 1, \quad \sum_{\mu=0}^{n} x^{\mu} y_\mu = 0. \tag{1.1}$$

The Hamiltonian for the Rosochatius system is

$$H = \frac{1}{2} \sum_{\mu=0}^{n} y_\mu^2 + \sum_{\mu=0}^{n} \frac{u_\mu^2}{x_\mu^2} + \epsilon \sum_{\mu=0}^{n} \alpha_\mu x_\mu^2, \tag{1.2}$$

where the constants $\{u_\mu, \alpha_\mu\}$ are arbitrary and for the generic case distinct. The parameter ϵ is introduced because we shall be concerned with the limit $\epsilon \to 0$ (where the harmonic terms in the potential vanish).

Another specialization is the Neumann oscillator problem,[15,16] where only the harmonic terms are retained and all $u_\mu = 0$. As pointed out in Ref. 23, the $\epsilon \to 0$ case (with $\Sigma u_\mu = 0$) corresponds to geodesic motion on $\mathbb{C}P^n$ under the standard Fubini–Study metric reduced by the maximal torus $[\mathrm{U}(1)]^n$ in the isometry group $\mathrm{SU}(n+1)$. More generally, as mentioned in Ref. 16 the system (1.2) may be obtained by reducing the Neumann oscillator problem on S^{2n+1} with all oscillator frequencies occurring in degenerate pairs by the product of $\mathrm{O}(2)$ rotations in the planes of the degenerate frequencies.

In the following section, these reductions will be obtained more explicitly and the flow for the restricted ($\epsilon = 0$) case of system (1.2) will be obtained from the projection of geodesic flow on S^{2n+1} (or $\mathbb{R}P^{2n+1}$). In Sec. III, the flow will be reobtained in elliptic coordinates in two related ways, the first based upon the use of a known complete set of integrals, the second by separation of variables in the Hamilton–Jacobi equation. Although the flow is obtained thus in terms of hyperelliptic integrals, the geometrical significance of the underlying hyperelliptic curve is not yet apparent from this computation. The particular case $\epsilon = 0$ turns out to lower the genus of the curve from n to $n-1$, giving rise to a singular Abelian integral together with $n-1$ regular ones. For this case the geometrical significance of the curve is given in Sec. IV in terms of the intersection of geodesics in $\mathbb{R}P^{2n+1}$ with confocal families of quadrics, and the dual description in terms of pencils of quadrics in $\mathbb{R}P^{2n+1*}$. The fact that the flow involves linearization on an extended Jacobi variety such that the image under the regular part of the Abel map is constant while the singular part gives a linear time dependence is shown to follow from the Abel theorem and a reciprocity formula for singular Abelian integrals of the third type. Finally, in Sec. V, the flow for the $\epsilon = 0$ case is given explicitly in terms of Clebsch's generalized θ functions, thereby establishing another relation with the elementary form of the projection of geodesic flow in terms of the ambient $T^*\mathbb{R}^{n+1}$ coordinates.

II. GEODESICS ON S^{2n+1} AND SYMMETRY REDUCTION

It will be convenient to identify S^{2n+1} as the unit sphere in $\mathbb{C}^{n+1}$ with complex Cartesian coordinates $\{W^\mu\}_{\mu=0,\ldots,n}$

$$W^\mu = s_\mu + it_\mu, \tag{2.1}$$

$$S^{2n+1}: \quad \sum_{\mu=0}^{n} W^\mu \overline{W}^\mu = 1. \tag{2.2}$$

Introducing canonical coordinates $\{W^\mu, p_\mu\}$ on $T^*\mathbb{C}^{n+1}$ with $p_\mu = y_\mu + iz_\mu, y_\mu, z_\mu \in \mathbb{R}$, the phase space T^*S^{2n+1} may be identified as the submanifold determined by the relation (2.2) together with

$$\sum_{\mu=0}^{n} [W^\mu p_\mu + \overline{W}^\mu \bar{p}_\mu] = 0.$$

Consider now the Neumann oscillator with Hamiltonian

$$H = \tfrac{1}{2} \sum_{\mu=0}^{n} p_\mu \bar{p}_\mu + \epsilon \sum_{\mu=0}^{n} \alpha_\mu |W_\mu|^2 \tag{2.3}$$

(i.e., such that each frequency is doubly degenerate between the real and imaginary parts of each complex axis).

In general, the oscillator strengths α_μ are otherwise distinct, but we shall be particularly concerned with the case of free motion, i.e., where ϵ vanishes. For this case, the flow is geodesic and given by

$$\begin{aligned} W^\mu(t) &= w^\mu \cos \omega t + v^\mu \sin \omega t, \\ p_\mu(t) &= -\omega \overline{w}^\mu \sin \omega t + \omega \bar{v}^\mu \cos \omega t, \end{aligned} \tag{2.4}$$

where $w = \{w^\mu\}$ and $v = \{v^\mu\}$ define an orthonormal two-frame

$$w^+ w = v^+ v = 1, \quad w^+ v + v^+ w = 0. \tag{2.5}$$

Note that such geodesic flow may already be interpreted in terms of Weinstein–Marsden reduction in the following sense. The sphere may be identified as a symmetric space $\mathrm{SO}(2n+2)/\mathrm{SO}(2n+1)$ with isometry group $\mathrm{SO}(2n+2)$, such that the projection

$$\pi: \mathrm{SO}(2n+2) \to S^{2n+1}$$

is a principal $\mathrm{SO}(2n+1)$ fibration. Geodesic flow on $\mathrm{SO}(2n+2)$ which is horizontal with respect to the canonical left-invariant connection projects to geodesic flow on S^{2n+1}. The natural lift of such horizontal flow to $T^*\mathrm{SO}(2n+2) \sim T\mathrm{SO}(2n+2)$ (identified via the metric) represents free Hamiltonian flow with zero-momentum map with respect to the right $\mathrm{SO}(2n+1)$ action. The reduced phase space is just T^*S^{2n+1} and the reduced Hamiltonian the corresponding free Hamiltonian with projected flow (2.4).

Returning to the general case with Hamiltonian (2.3), note that the maximal torus $T^{n+1} \subset \mathrm{SO}(2n+2)$, represented in complex coordinates by the diagonal $\mathrm{SU}(n+1)$ matrices with action

$$\mathrm{diag}\{e^{i\phi_\mu}\}: \begin{cases} \{W^\mu\} \to \{e^{i\phi_\mu} W^\mu\}, \\ \{p_\mu\} \to \{e^{-i\phi_\mu} p_\mu\}, \end{cases} \tag{2.6}$$

is a symmetry group, with momentum map

$$\begin{gathered} \Phi: T^*S^{2n+1} \to t_{n+1}^*, \\ \Phi: \{W^\mu, p_\mu\} \to \mathrm{diag}\{W^\mu p_\mu - \overline{W}^\mu \bar{p}_\mu\}_{\mu=0,\ldots,n}, \end{gathered} \tag{2.7}$$

where t_{n+1}^*, the dual to the Lie algebra $t_{n+1} \sim R^{n+1}$, is represented by imaginary diagonal matrices. Applying the Marsden–Weinstein reduction procedure again, we pick a point

$$U = \mathrm{diag}\{2\sqrt{2} i u_\mu\} \in t_{n+1}^*, \tag{2.8}$$

with all u_μ distinct and nonzero. The reduced phase space $\Phi^{-1}(U)/T^{n+1}$ may be identified locally with T^*S^n embedded in $T^*\mathbb{R}^{n+1}$ with canonical coordinates $\{x^\mu, y_\mu\}$ by the relations

$$\sum_{\mu=0}^{n} x^{\mu^2}=1, \quad \sum_{\mu=0}^{n} x^{\mu} y_\mu =0, \tag{2.9}$$

where

$$x^\mu = \pm |W^\mu|, \quad y_\mu = \operatorname{Re} p_\mu . \tag{2.10}$$

More precisely, we must remove all the S^{n-1} defined by intersecting the hyperplanes. The resulting manifold is diffeomorphic to an open disk. The reduced Hamiltonian is precisely (1.2). Since H is invariant under these reflections, the system may equivalently be studied on the sphere S^n (minus the n intersections with coordinate hyperplanes) without quotienting by the reflections. Moreover, since the potentials repel away from the coordinate hyperplanes, the motion is confined to the sectors of the sphere which these bound.

For the case $\epsilon=0$, the invariants associated with the geodesic flow (2.4) have values

$$u_\mu = (1/2i\sqrt{2})(W^\mu p_\mu - \overline{W}^\mu \overline{p}_\mu) = (\omega/\sqrt{2}) \operatorname{Im}\{ w^\mu \overline{v}_\mu \}. \tag{2.11}$$

The corresponding projected flow for the special case of the Rosochatius system with $\epsilon=0$ is thus given by (2.10) in terms of the flow (2.4), where in view of the T^{n+1} quotient, the vector $w^\mu = W^\mu(0)$ may be taken as real. The ambiguity of the sign in (2.10) is resolved by continuity. The flow for the general system (2.3) may be explicitly given in terms of hyperelliptic integrals, which we do in the following section, but in what follows we shall be particularly concerned with the geodesic case.

III. ABELIAN INTEGRALS

The flow for the Hamiltonian (1.2) will now be obtained by two related classical methods (a) through a complete set of commuting integrals, and (b) through separation of variables in the Hamilton–Jacobi equation. To treat the case where the harmonic oscillator forces vanish uniformly with the generic one we have introduced the parameter ϵ. For the case $\epsilon=0$, we shall understand $\{\alpha_\mu\}$ to be any distinct set of $n+1$ real constants.

A. Commuting integrals

On T^*S^n, we introduce the following set of $n+1$ Poisson commuting functions (cf. Refs. 15 and 16):

$$F_\mu^\epsilon \equiv \sum_{\substack{\nu=0\\ \nu\neq\mu}}^{n} \frac{\tilde{I}_{\mu\nu}^2}{\alpha_\mu - \alpha_\nu} + 2\epsilon x_\mu^2, \quad \mu=0,1,\ldots,n, \tag{3.1}$$

where

$$\tilde{I}_{\mu\nu}^2 \equiv I_{\mu\nu}^2 + 2[(x_\mu^2/x_\nu^2)u_\nu^2 + (x_\nu^2/x_\mu^2)u_\mu^2], \tag{3.2}$$

$$I_{\mu\nu} = x_\mu y_\nu - x_\nu y_\mu . \tag{3.3}$$

There are really only n independent functions, in view of the linear relation

$$\sum_{\mu=0}^{n} F_\mu^\epsilon = 2\epsilon. \tag{3.4}$$

The Hamiltonian (1.2) may be expressed in terms of these quantities as

$$H=\frac{1}{2}\sum_{\mu=0}^{n} \alpha_\mu F_\mu^\epsilon + \sum_{\mu=0}^{n} u_\mu^2 \tag{3.5}$$

and hence the quantities F_μ^ϵ Poisson commute with H, forming a complete set of invariants. These invariants have a simple group theoretical significance in the $\epsilon=0$ limit.[23] Namely, they are the reductions under the T^{n+1} action discussed above of a maximal algebraically independent set of commuting quadratic invariants in the enveloping algebra of the isometry algebra so$(2n+2)$. Such sets form strata under the action of SO$(2n+2)$ and the set $\{F_\mu^0\}$ is the generic (open dense) stratum whose only symmetries consist of the finite group of reflections in all coordinate planes.

To relate the invariants F_μ^0 with the basis of Refs. 23 and 24 we introduce the rational function of complex parameter λ

$$\begin{aligned} F^\epsilon(\lambda) &\equiv \frac{1}{2}\sum_{\mu=0}^{n} \frac{F_\mu^\epsilon}{\lambda-\alpha_\mu} \\ &= \sum_{k=1}^{n} \frac{(-1)^{k+1} I_k \lambda^{n-k}}{2A(\lambda)} + \epsilon \frac{\Sigma x_\mu^2}{\lambda-\alpha_\mu}, \end{aligned} \tag{3.6}$$

where

$$A(\lambda) \equiv \prod_{\mu=0}^{n} (\lambda-\alpha_\mu), \tag{3.7}$$

$$I_k = \sum_{\alpha,\beta} S_{k-1}^{\alpha\beta} \tilde{I}_{\alpha\beta}^2, \tag{3.8}$$

$$S_l^{\alpha\beta} = \frac{1}{l!} \sum_{\substack{\mu_1\neq\cdots\neq\mu_l\\ \neq\alpha,\beta}} \alpha_{\mu_1}\cdots\alpha_{\mu_l}, \quad l=0,\ldots,n. \tag{3.9}$$

For $\epsilon=0$, the F_μ^0 are thus linear combinations of I_k, which are the invariants of Refs. 23 and 24.

Now, introduce ellipsoidal coordinates on S^n as the n zeros a_i of the rational function

$$\sum_{\mu=0}^{n} \frac{x_\mu^2}{\lambda-\alpha_\mu} = \frac{P(\lambda)}{A(\lambda)}, \tag{3.10}$$

where

$$P(\lambda) = \prod_{i=1}^{n} (\lambda - a_i). \tag{3.11}$$

Similarly, express the right-hand side of (3.6) as the quotient of two polynomials

$$F^\epsilon(\lambda) \equiv Q^\epsilon(\lambda)/A(\lambda), \tag{3.12}$$

where $Q^\epsilon(\lambda)$ in general is an nth-order invariant polynomial with leading term $\epsilon\lambda^n$, reducing therefore to an $(n-1)$-order polynomial in the $\epsilon=0$ limit. On T^*S^n, we have canonical coordinates $\{a_i, b_i\}$ where

$$b_i = \frac{1}{2}\sum_\mu \frac{x^\mu y_\mu}{a_i - \alpha_\mu}. \tag{3.13}$$

To express the flow in ellipsoidal coordinates, introduce another invariant polynomial $S^\epsilon(\lambda)$ of order $(2n+1)$ with leading term $-8\epsilon\lambda^{2n+1}$ defined by

$$-\frac{Q^\epsilon(\lambda)}{A(\lambda)} - \sum_{\mu=0}^{n} \frac{u_\mu^2}{(\lambda-\alpha_\mu)^2} \equiv \frac{S^\epsilon(\lambda)}{8A^2(\lambda)}. \tag{3.14}$$

It follows from (3.7) and (3.14) that

$$\frac{S(a_i)}{16A^2(a_i)} = b_i^2 = \frac{1}{4}\left[\sum_{\mu=0}^{n} \frac{x^\mu y_\mu}{a_i-\alpha_\mu}\right]^2. \tag{3.15}$$

In order to verify (3.15) one evaluates (3.14) at $\lambda = a_i$, using the definitions (3.1) and (3.6) and the fact that the second term in (3.6) vanishes for $\lambda = a_i$.

To obtain the flow differentiate the identity

$$\sum_{\mu=0}^{n} \frac{x_\mu^2}{a_i - \alpha^\mu} = 0$$

along the flow lines

$$2 \sum_{\mu=0}^{n} \frac{x^\mu y_\mu}{a_i - \alpha_\mu} - \sum_{\mu=0}^{n} \frac{x^{\mu^2}}{(a_i - \alpha_\mu)^2} \dot{a}_i = 0$$

(the dot denotes differentiation with respect to time). Thus, defining the polynomial $W(\lambda)$ of degree $2n$ by

$$\sum_{\mu=0}^{n} \frac{x^{\mu^2}}{(\lambda - \alpha_\mu)^2} \equiv \frac{W(\lambda)}{A^2(\lambda)}, \tag{3.16}$$

we have, using (3.15),

$$\dot{a}_i = \sqrt{S(a_i)}\,[A(a_i)/W(a_i)]. \tag{3.17}$$

From the definitions (3.10) and (3.16), we have

$$\frac{W(\lambda)}{A^2(\lambda)} = -\frac{d}{d\lambda}\left(\frac{P(\lambda)}{A(\lambda)}\right)$$

and hence

$$W(a_i)/A(a_i) = -P'(a_i).$$

Therefore (3.17) reduces to the Abelian differential system

$$\dot{a}_i = -\frac{\sqrt{S(a_i)}}{\Pi_{j\neq i}(a_i - a_j)}. \tag{3.18}$$

Using a standard transformation, this may be expressed in the equivalent form

$$\sum_{i=1}^{n} \frac{a_i^{\,n-j}\dot{a}_i}{\sqrt{S^\epsilon(a_i)}} = \begin{cases} -1, & \text{if } j=1, \\ 0, & \text{if } 2 \leqslant j \leqslant n, \end{cases} \tag{3.19}$$

or, in terms of Abelian integrals

$$\sum_{i=1}^{n} \int_{a_i(0)}^{a_i(t)} \Omega_j^\epsilon \equiv \sum_{i=1}^{n} \int_{a_i(0)}^{a_i(t)} \frac{\lambda^{n-j}\,d\lambda}{\sqrt{S^\epsilon(\lambda)}} = \begin{cases} -t, & \text{if } j=1, \\ 0, & \text{if } 2 \leqslant j \leqslant n, \end{cases} \tag{3.20}$$

on the hyperelliptic curve X^ϵ defined by

$$z^2 = S^\epsilon(\lambda). \tag{3.21}$$

The above procedure amounts to an application of the Liouville theorem, reducing a Hamiltonian system to quadratures, given a complete set of integrals of motion. Namely, (3.15) expresses $\{b_i\}$ in terms of $\{a_i\}$ and the set of conserved quantities defining the coefficients of the invariant polynomial $Q^\epsilon(\lambda)$. The latter provide the action variables and the flow becomes linear in terms of the corresponding angle variables defined by differentiating the generating function G obtained from integrating the canonical form

$$G = \sum_i \int b_i\, da^i = \frac{1}{4} \sum_i \int^{a^i} \frac{\sqrt{S(\lambda)}}{A(\lambda)}\, d\lambda$$

with respect to these coefficients.

Note that if $\epsilon \neq 0$, the genus of the curve X^ϵ is $g = n$ and hence the Abelian integrals in (3.20) are all regular, while in the limit $\epsilon = 0$, the polynomial $S^\epsilon(\lambda)$ drops to degree $2n$, and hence the genus is $g = n - 1$ and the $j = 0$ case becomes a singular Abelian integral of third type with poles of first order at $\lambda = \infty$, on each sheet of X. This is the case corresponding to the projection of geodesic flow on S^{2n+1} and will be examined from a more geometrical viewpoint in Sec. IV.

B. Hamilton–Jacobi equation

We shall now briefly summarize how the flow given above in terms of Abelian integrals may be deduced by separation of variables in the Hamilton–Jacobi equation. The main interest is to relate the separation constants in the present procedure to the integrals of motion given above.

To express the Hamiltonian (1.2) in elliptic coordinates, we use the relation

$$x_\mu^2 = P(\alpha_\mu)/A'(\alpha_\mu), \tag{3.22}$$

following from (3.10) by evaluation of residues. The potential energy terms in H may then be expressed in ellipsoidal coordinates as

$$\sum_{\mu=0}^{n} \epsilon \alpha_\mu x_\mu^2 = \epsilon \sum_{\mu=0}^{n} \alpha_\mu - \epsilon \sum_{i=1}^{n} \frac{a_i^{\,n}}{P'(a_i)},$$

$$\sum_{\mu=0}^{n} \frac{u_\mu^2}{x_\mu^2} = \sum_{\mu=0}^{n} u_\mu^2 - \sum_{i=1}^{n} \frac{R(a_i)}{A(a_i)P'(a_i)}, \tag{3.23}$$

where $R(\lambda)$ is a polynomial of degree $2n$ defined by

$$\sum_{\mu=0}^{n} \frac{u_\mu^2}{(\lambda - \alpha_\mu)^2} \equiv \frac{R(\lambda)}{A^2(\lambda)}. \tag{3.24}$$

The coordinate change involved in (3.23) is most easily computed using the standard trick[15] of converting sums over poles to contour integrals evaluated at ∞.

Making the appropriate transformation for the kinetic energy term, the Hamiltonian (1.2) expressed in ellipsoidal coordinates is

$$H = -2\sum_{i=1}^{n} \frac{A(a_i)}{P'(a_i)} b_i^{\,2} - \sum_{i=1}^{n} \frac{R(a_i)}{A(a_i)P'(a_i)} - \epsilon \sum_{i=1}^{n} \frac{a_i^{\,n}}{P'(a_i)} + \sum_{\mu=0}^{n} [u_\mu^2 + \epsilon \alpha_\mu]. \tag{3.25}$$

Using the Jacobi method described in Ref. 15 we introduce a polynomial

$$\tilde{Q}^\epsilon(\lambda) \equiv \epsilon \lambda^n + P_1 \lambda^{n-1} + \cdots + P_n, \tag{3.26}$$

where, in terms of the energy E, P_1 is defined to be

$$P_1 = E - \sum_{\mu=0}^{n} [u_\mu^2 + \epsilon \alpha_\mu] \tag{3.27}$$

and the other $\{P_i\}_{i=2,\ldots,n}$ are arbitrary separation constants. Putting

$$E - \sum_{\mu=0}^{n} (u_\mu^2 + \epsilon\alpha_\mu) + \epsilon \sum_{i=1}^{n} \frac{a_i^{\,n}}{P'(a_i)} = \sum_{i=1}^{n} \frac{\tilde{Q}^\epsilon(a_i)}{P'(a_i)},$$

we obtain the Hamilton–Jacobi equation

$$H\left(a_i, \frac{\partial G}{\partial a_i}\right) = E \tag{3.28}$$

in the form

$$\sum_{i=1}^{n} \left[\frac{2A(a_i)}{P'(a_i)} \left(\frac{\partial G}{\partial a_i}\right)^2 + \frac{R(a_i)}{A(a_i)P'(a_i)} + \frac{\tilde{Q}^\epsilon(a_i)}{P'(a_i)}\right] = 0. \tag{3.29}$$

This admits the separated complete solution

$$G(a_1,...,a_n)=\sum_{i=1}^{n} s_i(a_i), \tag{3.30}$$

where

$$s_i(a_i)=\int_{a_i(0)}^{a_i(t)} \frac{\sqrt{\widetilde{S}^\epsilon(\lambda)}}{4A(\lambda)}\,d\lambda \tag{3.31}$$

and

$$\widetilde{S}^\epsilon(\lambda)/8\equiv-[A(\lambda)\widetilde{Q}^\epsilon(\lambda)+R(\lambda)]. \tag{3.32}$$

Treating P_i as the new momenta (action variables), the flow in terms of the canonically conjugate Q_i (angle variables) is

$$Q_j=\frac{\partial G}{\partial P_j}=-\sum_i\int_{a_i(0)}^{a_i(t)}\frac{\lambda^{n-j}\,d\lambda}{\sqrt{\widetilde{S}^\epsilon(\lambda)}}$$

$$=\begin{cases} t, & \text{if } j=1,\\ 0, & \text{if } n\geqslant j>1. \end{cases} \tag{3.33}$$

To complete the correspondence with the flow deduced above in terms of the commuting invariants F_μ^ϵ, we must identify the separation constants as functions on phase space, and hence relate the polynomials $\widetilde{S}^\epsilon(\lambda)$ and $S^\epsilon(\lambda)$.

Lemma: The polynomials $\widetilde{S}^\epsilon(\lambda)$ and $S^\epsilon(\lambda)$ are identical and hence, for the particular case $\epsilon=0$ the separation constants P_i are related to the integrals of motion (3.8) by

$$P_k=\tfrac{1}{2}(-1)^{k+1}I_k. \tag{3.34}$$

Proof: From (3.31), it follows that along the flow,

$$b_i^2=\left(\frac{\partial S}{\partial a_i}\right)^2=\frac{\widetilde{S}(a_i)}{16A^2(a_i)}. \tag{3.35}$$

But, by Eq. (3.15), we have

$$\widetilde{S}^\epsilon(a_i)=S^\epsilon(a_i). \tag{3.36}$$

However, in view of Eq. (3.32), and the corresponding relation

$$S^\epsilon(\lambda)/8=-A(\lambda)Q^\epsilon(\lambda)-R(\lambda) \tag{3.37}$$

following from (3.14) and (3.24), we conclude that the two nth-degree polynomials $Q^\epsilon(\lambda)$ and $\widetilde{Q}^\epsilon(\lambda)$ are equal at the n distinct points $\lambda=a_i$. Since they also have the same leading term $\epsilon\lambda^n$, they are identically equal, and hence so are $S^\epsilon(\lambda)$ and $\widetilde{S}^\epsilon(\lambda)$.

In the following sections, we shall only be concerned with the case $\epsilon=0$ and denote the polynomial $S^0(\lambda)$ henceforth $S(\lambda)$.

IV. GEODESICS AND THE GEOMETRY OF QUADRICS

We have obtained a linearization of the system (1.2) on the reduced phase space in terms of Abelian integrals, i.e., integrals of holomorphic differentials over the algebraic curve $X^\epsilon=\{z^2=S^\epsilon(\lambda)\}$, where $S^\epsilon(\lambda)$ is of degree $2n+1$ if $\epsilon\neq0$. As noted above, for the case $\epsilon=0$, which we concentrate on from here on, the polynomial $S^0(\lambda)\equiv S(\lambda)$ is of degree $2n$ and the genus of the curve $X\equiv X^0$ drops to $g=n-1$. For this case, the differential $\Omega_0\equiv\Omega_0^0$ in (3.20) becomes singular, having simple poles at the points (∞_1,∞_2) on X over $\lambda=\infty$. It follows from the definition (3.14), or equivalently (3.32), that $S(\lambda)$ has leading term $-4\omega^2\lambda^{2n}$, where ω is the frequency of the geodesic motion (2.4), and hence the residues of Ω_0 at (∞_1,∞_2) are $\pm i/2\omega$.

We now turn to showing how the curve X arises naturally in relation to geodesic flow on $\mathbb{R}P^{2n+1}=S^{2n+1}/\mathbb{Z}_2$, i.e., on the unreduced system. (The distinction between $\mathbb{R}P^{2n+1}$ and S^{2n+1} is irrelevant, in view of the T^{n+1} quotienting.)

Invariantly, the map

$$(A_0,\mathbf{A}):\{a_i\}\rightarrow\sum_{i=1}^{n}\int_{a_i(0)}^{a_i(t)}(\Omega_0,\mathbf{\Omega}), \tag{4.1}$$

$\mathbf{\Omega}=(\Omega_1,...,\Omega_g)$, $\mathbf{A}=(A_1,...,A_g)$ is a map from $S^{g+1}X$, the $(g+1)$th symmetric power of X, into an extended Jacobian $\widetilde{J}=C^{g+1}/\mathscr{L}$, where $\mathscr{L}$ is a lattice isomorphic to $\mathbb{Z}^{2g+1}$; the integrals are defined modulo the $2g$ cycles of the curve and the residues of Ω_0. The extended Jacobian $\widetilde{J}$ is thus related to the standard Jacobian by projection onto the last g coordinates

$$0\rightarrow\mathbb{C}/(i/2\omega)\mathbb{Z}\rightarrow\widetilde{J}\rightarrow J\rightarrow0.$$

Under this projection, we see from (3.20) that the flow is mapped to a single point: $\mathbf{A}(a_i(t))=\mathbf{A}(a_i(0))$. By Abel's theorem,[27] this implies that there exists a meromorphic function on X with zeros $\{a_i(t)\}$ and poles $\{a_i(0)\}$. The flow is thus a flow of meromorphic functions. Furthermore, we shall see that it is of the form $\hat{k}-\cot(\omega t)$, where $\hat{k}$ is a fixed function on X and therefore the flow takes place in a one-dimensional linear system. It is the pair $(X,\hat{k})$ that we must obtain in terms of geodesics.

In view of the T^{n+1} quotienting, we may limit ourselves to the geodesic flow lines in (2.4) with w^μ real. Interpreting (s^μ,t^μ) as homogeneous coordinates, one can rewrite (2.4) as

$$s^\mu=Kb^\mu+c^\mu,\quad t^\mu=d^\mu, \tag{4.2}$$

where

$$K=\cot(\omega t), \tag{4.3}$$

$$w^\mu\equiv b^\mu,\quad v^\mu\equiv c^\mu+id^\mu. \tag{4.4}$$

The geodesic is the projective line spanned by $\{(b_\mu,0),(c_\mu,d_\mu)\}$. The elliptic coordinates of the image under reduction by T^{n+1} of (s^μ,t^μ) are the values of λ such that

$$\sum_{\mu=0}^{n}\frac{s^{\mu^2}+t^{\mu^2}}{\lambda-\alpha_\mu}=0. \tag{4.5}$$

The family of confocal quadrics defined by (4.5) will be denoted $\{Q_\lambda\}$, $\lambda\in\mathbb{C}$. (From now on, we complexify our problem by letting all parameters take values in $\mathbb{C}$.) Now choose a specific geodesic. This is equivalent to choosing initial values and u^μ in the reduced problem. One has two families: the family of points p_k in the geodesic, and the family Q_λ of quadrics. Consider the set

$$\{(k,\lambda)\in\mathbb{C}P^1\times\mathbb{C}P^1|p_k\in Q_\lambda\}. \tag{4.6}$$

Putting (4.2) and (4.5) together, one has

$$0=k^2\left(\sum_{\mu=0}^{n}\frac{b^{\mu^2}}{\lambda-\alpha_\mu}\right)+2k\left(\sum_{\mu=0}^{n}\frac{b^\mu c^\mu}{\lambda-\alpha_\mu}\right)+\left(\sum_{\mu=0}^{n}\frac{c^{\mu^2}+d^{\mu^2}}{\lambda-\alpha_\mu}\right)$$

$$\equiv k^2\frac{R_2(\lambda)}{A(\lambda)}+2k\frac{R_1(\lambda)}{A(\lambda)}+\frac{R_0(\lambda)}{A(\lambda)}. \tag{4.7}$$

Comparison with (3.14), using $u_\mu = -(\omega/\sqrt{2})b^\mu d^\mu$, $x^\mu(0) = b^\mu$, $y_\mu(0) = \omega c^\mu$, shows that this is precisely the curve X, $(z^2 = S(\lambda))$ with the identification

$$S(\lambda) = 4\omega^2[R_1^2 - R_0R_2] \tag{4.8}$$

and the change of variable

$$z \equiv 2\omega[kR_2(\lambda) + R_1(\lambda)]. \tag{4.9}$$

Now k is a function on the curve X and (4.9) gives it in terms of z and λ.

A particular value K of k corresponds to a particular point p on the geodesic. The λ coordinates of the point on X where k takes the value K are just the elliptic coordinates of p by (4.5). Thus, the values $\{a_i(t)\}$ of λ at the zeros of k cot(ωt) give the flow in elliptic coordinates. Also, from (4.9), the poles of k are at the zeros of R_2, and so the values of λ at these poles are $a_i(0)$. Thus k is precisely the $\hat{k}$ we were looking for, and the flow of functions is $f_t = k - \cot(\omega t)$.

We now obtain the linearization in terms of Abelian integrals, using k. As noted before, the constancy of the last g integrals in (3.20) follows from Abel's theorem. There remains the evaluation of the first integral. This may be done using the following reciprocity formula.

Proposition: Let a_i, b_i be a set of representative cycles for a basis of $H_1(X, \mathbb{Z})$ with $a_i \cdot a_j = 0$, $a_i \cdot b_j = \delta_{ij}$, $b_i \cdot b_j = 0$, $i, j = 1,...,g$.

Let f be a function on X with poles at q_i and zeros at p_i, $i = 1,...,d$, and Ω be an Abelian differential with simple poles at ∞_1, ∞_2 with residues γ, $-\gamma$. Setting $\int_{a_j} d(\log f) = 2\pi i n_j$, $\int_{b_j} d(\log f) = 2\pi i m_j$ $(n_j, m_j \in \mathbb{Z})$, $\int_{a_j}\Omega = \Omega_j^a$, $\int_{b_j}\Omega = \Omega_j^b$, one has

$$\sum_{i=1}^{d} \int_{a_i}^{p_i} \Omega = \sum_{j=1}^{g} \Omega_j^a m_j - \Omega_j^b m_j + \gamma(\log f(\infty_1) - \log f(\infty_2)). \tag{4.10}$$

Proof: The proof is the same as that of the Weil relation for meromorphic functions (see, e.g., Ref. 28), with Ω substituted for an exact differential $d(\log g)$.

We now apply this formula to $f_t = k - \cot(\omega t)$, and $\Omega = \Omega_0$. One notes that, by continuity, n_i and m_j are constants, and so

$$\sum_{i=1}^{d} \int_{a_i(0)}^{a_i(t)} \Omega_0 = \frac{i}{2\omega} \log\left(\frac{k(\infty_1) - \cot(\omega t)}{k(\infty_2) - \cot(\omega t)}\right) = -t, \tag{4.11}$$

since $k(\infty_1)$, $k(\infty_2) = \pm i$. This completes the correspondence between geodesic flow (2.4) and the solution (3.20) in terms of Abelian integrals.

The geometry of the situation is best exhibited by going to the dual space $(\mathbb{C}P^{2n+1})^*$, as in Ref. 29. Denoting points in $(\mathbb{C}P^{2n+1})^*$ by dual coordinates (s_μ, t_μ) the object which is dual to the confocal family of quadrics $\{Q_\lambda\}$ is the pencil of quadrics $\{Q_\lambda^*\}$ defined by the relations

$$\sum_{\mu=0}^{n} (\lambda - \alpha_\mu)(s_\mu^2 + t_\mu^2) = 0. \tag{4.12}$$

The duality correspondence may be summarized as in Table I.

We have thus obtained two pencils, the pencil p_k^* of hyperplanes through l^* and the pencil Q_λ^* of quadrics. The curve X, after dualizing, is just the set

$$\{(k, \lambda) \in \mathbb{C}P^1 \times \mathbb{C}P^1 | p_k^* \text{ is tangent to } Q_\lambda^*\}.$$

TABLE I. The duality correspondence.

$\mathbb{C}P^{2n+1}$		$(\mathbb{C}P^{2n+1})^*$
The point $p = (s^\mu, t^\mu)$	$\leftrightarrow$	The hyperplanes $\{(s_\mu, t_\mu) \| s^\mu s_\mu + t^\mu t_\mu = 0\}$
The line l spanned by $(b^\mu, 0)$ (c^μ, d^μ)	$\leftrightarrow$	The $(2n-1)$ plane l^* $\{(s_\mu, t_\mu) \| b^\mu s_\mu = 0, c^\mu s_\mu + d^\mu t_\mu = 0\}$
$p \in l$	$\leftrightarrow$	$l^* \subset p^*$
$p = (kb^\mu + c^\mu, d^\mu)$		p^* is the hyperplane $p_k^* = \{(s_\mu, t_\mu) \| (kb^\mu + c^\mu)s_\mu + d^\mu t_\mu = 0\}$ in the pencil of hyperplanes through l^*
The confocal family Q_λ	$\leftrightarrow$	The pencil Q_λ^*
$p \in Q_\lambda$	$\leftrightarrow$	p^* tangent to Q_λ^*

The above exhibits X as an invariant of the flow; this can also be seen in a way that is more natural to an algebraic geometer. Nondegenerate pencils of quadrics Q_λ^* on $\mathbb{P}^{2n-1}$ are classified by an associated hyperelliptic curve, of genus $(n-1)$, whose branch points λ_i over $P_1(a)$ are given by the condition that $Q_{\lambda_i}^*$ be singular.[29,30] In our case, if we restrict Q_λ^* to our $(2n-1)$ space l^* that gives the flow, one obtains the curve X.

Remark: The phase space of the projected system is foliated locally by the real parts of the extended Jacobian $\tilde{J}$; equivalently, the space of flows (lines in $\mathbb{C}P^{2n+1}$, after complexification) is foliated by standard Jacobians J. There remains the question of determining how $J(X)$ parametrizes the set $I(x)$ of lines that correspond to a given curve X, up to a finite multiplicity.

For each l^* corresponding to X, choose an isomorphism $\psi_l^*: l^* \to \mathbb{C}P^{2n-1}$ so that Q_λ^* is mapped to some standard pencil $\tilde{Q}_\lambda^*$; this is well defined modulo a finite number of choices and, locally, can be done continuously.

One has on $(\mathbb{C}P^{2n+1})^*$ the $(n+1)$ quadrics $s_\mu^2 + t_\mu^2 = 0$, giving pairs of hyperplanes H_μ^0, H_μ^1. The pencil Q_λ^* is built up from these quadrics. Their complete intersection is a union of 2^{n+1} n-dimensional planes $P_J = \cap_{\mu=0}^n H_i^{J(i)}$, where $J: \{0,...,n\} \to \{0, 1\}$. These planes, restricted to l^*, give $(n-2)$-dimensional planes on the intersection of the Q_λ^*, restricted to l^*. One then uses the ψ_l^* to get an $(n-2)$ plane on the intersection of the $\tilde{Q}_\lambda^*$. However, the space of these, by results in Refs. 29–31, is precisely $J(X)$. Thus choosing one P_J, for example, $P = \cap_{\mu=0}^n H_\mu^0$, one obtains a map $\varphi: I(X) \to J(X)$.

Conversely, given a $\mathbb{C}P^{2n-1}$ with a pencil $\tilde{Q}_\lambda^*$ on it whose invariant is X, choosing an $(n-2)$-dimensional plane on the intersection of the Q_λ^*'s and calling it P, and using again results in Refs. 29–31, one can recreate the other P_J's, and so the H_μ^0, H_μ^1. One can then embed $\mathbb{P}^{2n-1} \to \mathbb{P}^{2n+1}$ in a unique way such that $\tilde{Q}_\lambda^* = Q_\lambda^*|_{\mathbb{P}^{2n-1}}$ and the P_J's correspond to those given above. This means that the map φ is invertible.

V. θ FUNCTIONS

The inversion of the extended Abel map was studied by Clebsch,[32] using a generalized θ function $\tilde{\theta}$. We sketch this theory and its application to geodesic flow and see how $\widetilde{\theta}$ reduces to exponentials.

First, renormalize the differentials (Ω_0, Ω_i) so that the a_j periods of (Ω_0, Ω_i) are $(0, \delta_{ij})$ and the residues of Ω_0 at ∞_1, ∞_2 are ± 1. Also, redefine the Abel map so that if $\mathbf{p} = (p_1,...,p_{g+1})$,

$$(A_0(\mathbf{p}), \mathbf{A}(\mathbf{p})) = \sum_{i=1}^{g+1}\left(\int_{p_0}^{p_i}\Omega_0, \int_{p_0}^{p_i}\mathbf{\Omega}\right),$$

where $\mathbf{\Omega} = (\Omega_1,...,\Omega_g)$; that is, we are taking the integrals from a fixed point instead of $a_i(0)$. With these conventions, the flow becomes

$$(A_0(\mathbf{p}(t)), \mathbf{A}(\mathbf{p}(t))) = (c_0 + i\omega t, c), \tag{5.1}$$

where c_0, $c = (c_1,...,c_g)$ are constants.

We also define the Abel map for individual points $q \in X$

$$(\widetilde{A}_0(q), \widetilde{\mathbf{A}}(q)) = \left(\int_{p_0}^{q}\Omega_0, \int_{p_0}^{q}\mathbf{\Omega}\right). \tag{5.2}$$

Now define the function $\widetilde{\theta}$ for $\widetilde{J}$ in terms of the normal θ function for J,

$$\widetilde{\theta}(z_0, z) = \theta(z - \nu)e^{-z_0/2} + \theta(z + \nu)e^{z_0/2}, \tag{5.3}$$

where $z = (z_1,...,z_g)$, $\nu = (\nu_1,...,\nu_g)$, $\nu_i = \frac{1}{2}\int_{\infty_1}^{\infty_2}\Omega_i$. Just as θ can be used to invert the standard Abel maps A, $\widetilde{\theta}$ can be used to invert our extended map $(A_0, \mathbf{A})$.

More specifically, if $q \in X, k_0 \in \mathbb{C}, k \in \mathbb{C}^g$, set $F_{k_0, k}(q) = \widetilde{\theta}(\widehat{A}_0(q) - k_0, \widehat{A}(q) - k)$; then one has, in complete analogy with standard θ functions, the following.

Lemma 5.1 (Refs. 32 and 33): If $F_{k_0, k}(p) \neq 0$, then $F_{k_0, k}(p)$ has $(g+1)$ zeros on X.

Theorem 5.1 (Refs. 32 and 33): There exist $k_0 \in \mathbb{C}, k \in \mathbb{C}^g$ such that for $\lambda_0 \in \mathbb{C}, \lambda \in \mathbb{C}^g$ generic, if $p_i, i = 1,...,g+1$ are the zeros of $F_{\lambda_0 + k_0, \lambda + k}(p)$ then $\Sigma_{i=1}^{g+1}(\widehat{A}_0(p_i), \widehat{\mathbf{A}}(p_i)) = (\lambda_0, \lambda)$.

Thus, one can invert $(A_0, \mathbf{A})$: the inverse image of λ_0, λ is the zero set of $F_{\lambda_0 + k_0, \lambda + k}$ on X.

We now solve for the flow (5.1) in terms of the coordinates x_μ, and show how one can reobtain (2.4). First, the inversion of the Abel map gives points $p_i(t)$ on X and under the two-sheeted projection $\pi: X \to \mathbb{P}_1(\mathbb{C})$, one has $\pi(p_i(t)) = a_i(t)$. However, the $a_i(t)$ are the zeros of the function

$$g_t(\lambda) = (\lambda - \alpha_0)\sum_{\mu=0}^{g+1}\frac{x_\mu^2(t)}{(\lambda - \alpha_\mu)} \tag{5.4}$$

on $\mathbb{P}_1(\mathbb{C})$. Pulling back g_t to function h_t on X, h_t has zeros $\{p_i(t), \tau(p_i(t))\}$, where τ is the hyperelliptic involution, and poles at the $(2g+2)$ points $\{\beta_\mu, \tau(\beta_\mu)\}$ that are the inverse image under π of $\{\alpha_\mu, \mu = 1,...,g+1\}$.

Note now that if $\mathbf{p} = (p_1,...,p_{g+1})$ are points on X, then $[A_0(\tau(\mathbf{p})), A(\tau(\mathbf{p}))] = (T_0, \mathbf{T}) - (A_0(\mathbf{p}), \mathbf{A}(\mathbf{p}))$, where $(T_0, \mathbf{T})$ are constants. Using this, one can recreate h_t from $\widetilde{\theta}$, up to constants

$$h_t(q) = \frac{F_{c_0 + i\omega t + k_0, c + k(q)}F_{-c_0 - i\omega t + T_0 + k_0, -c + T + k(q)}}{F_{\gamma_0 + k_0, \gamma + k(q)}F_{-\gamma_0 + T_0 + k_0, -\gamma + T + k(q)}}, \tag{5.5}$$

where $(\gamma_0, \gamma) = \Sigma_{\mu=1}^{g+1}(\widehat{A}_0(\beta_\mu), \widehat{\mathbf{A}}(\beta_\mu))$.

To see this, one verifies that Theorem 5.1 implies that the right-hand side of (5.5) has the same zeros and poles as h_t. The fact that it is a well-defined function on X, independent of the paths of integration in the definition of $(\widehat{A}_0, \widehat{\mathbf{A}})$ follows from the periodicity relations for $\widetilde{\theta}$.

By evaluating residues at the β_μ, and separating out the t dependence, which enters only through the exponentials in (5.3), one has

$$x_\mu^2(t) = e_\mu e^{-2i\omega t} + f_\mu + g_\mu e^{2i\omega t}, \tag{5.6}$$

where e_μ, f_μ, g_μ are constants. The reality of the coefficients of the function g implies $g_t(\bar{\lambda}) = \overline{g_t(\lambda)}$; and so $e_\mu = \bar{g}_\mu, f_\mu = \bar{f}_\mu$. The resulting flow is identical to that obtained in (2.4), with suitable identification of the constants.

VI. CONCLUSIONS

The main result of this paper is a clarification of the relation between different methods of integrating a specific Hamiltonian system. The simplest method, when applicable, is that of the projection of geodesic flow from an appropriate larger space. The Rosochatius system, studied in this paper, is obtained by such a projection only if $\epsilon = 0$ in (1.2). The alternative methods, making explicit use of the known system of integrals of motion, are more general and have been applied for ϵ arbitrary.

The "free" Rosochatius system ($\epsilon = 0$) could of course be solved by separation of variables in any system of coordinates which allows separation of the free Hamilton–Jacobi equation on S_n (see Ref. 23), e.g., in spherical coordinates.

The methods of this article can be directly applied to other systems. The system (1.2) with $\epsilon = 0$ (and $\Sigma u_\mu = 0$) was obtained from the free system on $\mathbb{C}P^n \sim \mathrm{SU}(n+1)/\mathrm{U}(n)$ by quotienting by the maximal torus.[23] Spaces with noncompact groups of isometries yield much richer results. Thus, starting from the Hermitian hyperbolic space $\mathrm{HH}(n) \sim \mathrm{SU}(n, 1)/\mathrm{U}(n)$ we can quotient by any one of the many different mutually inequivalent maximal Abelian subgroups of $\mathrm{SU}(n, 1)$. Different subgroups will give different integrable systems,[22] all of them integrable by the group projection method (as well as the other methods of this paper).

As we have seen, the relationship in the $\epsilon = 0$ case between reduction of geodesic flow and linearization on a Jacobi variety is somewhat trivial because all the time dependence is contained in the singular component of the extended Abel map. It would be of great interest to obtain such a relationship for the case of regular maps. Of perhaps even greater interest would be the extension of these considerations to the case of infinite-dimensional symmetry algebras of the Kac–Moody variety.

ACKNOWLEDGMENTS

The authors thank P. van Moerbeke for helpful discussions.

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada and the Fonds FCAC pour l'aide et le soutien à la rechereche du Gouvernement du Québec.

[1] M. A. Olshanetsky and A. M. Perelomov, Phys. Rev. 71, 313 (1981).
[2] F. Calogero, "Integrable dynamical systems and related mathematical results, in *Lecture Notes in Physics*, Vol. 189, edited by K. B. Wolf (Springer, New York, 1982), pp. 47–96.

[3]M. Adler, J. Math. Phys. **20**, 60 (1979).
[4]D. Kazhdan, B. Kostant, and S. Sternberg, Commun. Pure Appl. Math. **31**, 481 (1978).
[5]J. Marsden and A. Wienstein, Rep. Math. Phys. **5**, 121 (1974).
[6]M. Adler, Invent. Math. **50**, 219 (1979).
[7]M. Adler and P. van Moerbeke, Adv. Math. **38**, 267 (1980); **38**, 318 (1980).
[8]B. Kostant, Adv. Math. **34**, 105 (1979).
[9]W. W. Symes, Physica D **1**, 339 (1980).
[10]A. G. Reyman and M. Semenov-Tian-Shansky, Invent. Math. **54**, 81 (1979); **63**, 423 (1981).
[11]I. M. Krichever, Usp. Mat. Nauk **32**, 183 (1977) [Russ. Math. Surv. **32**, 185 (1977)].
[12]I. M. Krichever and S. P. Novikov, Usp. Mat. Nauk **35**, 47 (1980) [Russ. Math. Surv. **35**, 53 (1980)].
[13]B. A. Dubrovin, Usp. Mat. Nauk **36**, 11 (1981) [Russ. Math. Surv. **36**, 11 (1981)].
[14]H. P. McKean, "Integrable systems and algebraic curves," in *Global Analysis, Lecture Notes in Mathematics*, Vol. 755, edited by M. Grmela and J. E. Marsden (Springer, New York, 1979), pp. 83–200.
[15]J. Moser, "Various Aspects of Integrable Hamiltonian Systems," in *Dynamical Systems*, CIME Lectures, Bressanone, Italy, June 1978; *Progress in Mathematics*, Vol. 8 (Birkhauser, Boston, 1980).
[16]J. Moser, "Geometry of quadrics and spectral theory," in *The Chern Symposium, 1979*, edited by W. Y. Hsiang, S. Kobayashi, I. M. Singer, A. Weinstein, J. Wolf, and H. H. Wu (Springer, New York, 1980).
[17]P. Winternitz and I. Friš, Yad. Fiz. **1**, 889 (1965) [Sov. J. Nucl. Phys. **1**, 636 (1965)].
[18]W. Miller, Jr., *Symmetry and the Separation of Variables* (Addison–Wesley, Reading, MA, 1977).
[19]W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. **22**, 251 (1981).
[20]E. G. Kalnins and W. Miller, Jr., SIAM J. Math. Anal. **11**, 1011 (1980).
[21]E. G. Kalnins and W. Miller, Jr., SIAM J. Math. Anal. **12**, 617 (1981).
[22]C. P. Boyer, E. G. Kalnins, and P. Winternitz, J. Math. Phys. **24**, 2022 (1982).
[23]C. P. Boyer, E. G. Kalnins, and P. Winternitz, SIAM J. Math. Anal. **16**, 93 (1985).
[24]E. G. Kalnins and W. Miller, Jr., "Separation of variables on n-dimensional Riemannian manifolds: 1. The n-sphere S_n and Euclidean n-space R_n," research report No. 102, Waikato, New Zealand, 1982.
[25]E. Rosochatius, Dissertation, Göttingen, Gebr. Unger, Berlin, 1877.
[26]T. Ratiu, "The Lie algebraic interpretation of the complete integrability of the Rosochatius system," in *Mathematical Methods in Hydrodynamics and Integrability in Dynamical Systems, AIP Conference Proceedings*, Vol. 88 (AIP, New York, 1982).
[27]C. L. Siegel, *Topics in Complex Function Theory* (Wiley, New York, 1971), Vol. 2.
[28]P. Griffiths and J. Harris, *Principles of Algebraic Geometry* (Wiley, New York, 1978).
[29]H. Knörrer, Invent. Math. **59**, 119 (1980).
[30]M. Reid, Ph. D.thesis, Cambridge University, Great Britain, 1972.
[31]R. Donagi, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **7**, 217 (1980).
[32]A. Clebsch, *Leçons sur la Géométrie* (Gauthier-Villars, Paris, 1883), Vol. 3.
[33]L. Gagnon, M.Sc. thesis, Université de Montréal, 1984.

Classes of potentials of time-dependent central force fields which possess first integrals quadratic in the momenta

P. G. L. Leach
Department of Applied Mathematics, The University of the Witwatersrand, 1 Jan Smuts Avenue, Johannesburg, South Africa

(Received 16 February 1984; accepted for publication 16 November 1984)

The general form for the potential of a time-dependent central force which possesses an energylike first integral is determined. In two special cases additional first integrals (apart from the angular momentum) are found. These enable the orbit equation to be determined without further quadrature. The Schrödinger equation is solved fully in these cases.

I. INTRODUCTION

Interest in dynamical systems for which a first integral exists is a continuing fact. One source of impetus for the study of first integrals is the study of fusion processes. The particle distribution function is a solution of the Vlasov–Poisson equations in one dimension and the Vlasov–Maxwell equations in three dimensions. Since the Vlasov equation has the same form as the Liouville equation in Hamiltonian mechanics it follows that the distribution function is a function of the first integrals of the dynamical system. There is an aesthetic pleasure in solving a problem such as determining the first integrals of a dynamical system. The practical problem of developing a theoretical fusion model to provide a guide for experiment adds utility to the pleasure.

The attention given to one-dimensional systems in recent years[1–7] has established that the Hamiltonian

$$H=\frac{1}{2}p^2-\frac{1}{2}\frac{\ddot\rho}{\rho}(q-\alpha)^2-\ddot\alpha q+\frac{1}{\rho^2}U\left(\frac{q-\alpha}{\rho}\right) \qquad (1.1)$$

has the first integral

$$I=\tfrac{1}{2}\{\rho(p-\alpha)-\dot\rho(q-\alpha)\}^2+U((q-\alpha)/\rho). \qquad (1.2)$$

The functions $\rho(t)$, $\alpha(t)$, and $U(\xi)$ are arbitrary. The first integral above is quadratic in the momentum. The question has been asked[2(c),3,5] whether an invariant which is not quadratic in the momentum will exist for an interesting potential as the potential above has not been of much use in fusion applications.[8] A first integral linear in the momentum occurs only when the potential is quadratic.[2(b)] For a posited polynomial integral of degree greater than 2 it is not possible to solve the problem in closed form,[2(b),5] and there has been speculation that the permissible potential will be contained in that above.[5] The situation is different in the case of one posited nonpolynomial integral. For

$$\begin{aligned}H=\frac{1}{2}p^2&-\frac{1}{2}\frac{\ddot\rho}{\rho}(q-\alpha)^2\\&-\ddot\alpha(q-\alpha)-\dot\sigma\log(q-\alpha)-\frac{1}{2}\frac{\sigma^2}{(q-\alpha)^2},\end{aligned} \qquad (1.3)$$

there exists the first integral

$$I=T-[(q-\alpha)/\rho]/[\rho(p-\dot\alpha)-\dot\rho(q-\alpha)-\sigma\rho/(q-\alpha)], \qquad (1.4)$$

where $\rho(t)$, $\alpha(t)$, and $\sigma(t)$ are arbitrary functions of time and $T=\int\rho^{-2}\,dt$ [see Refs. 2(c), 3, and 9].

While the problem of finding the most general potential for one-dimensional particle motion for which an explicit first integral can be written down cannot yet be said to be solved completely, the realities of three-dimensional space have turned attention to determining first integrals and the corresponding class(es) of permissible potentials for the Hamiltonian[3,10–12]

$$H=\tfrac{1}{2}\mathbf{p}^2+V(\mathbf{q},t). \qquad (1.5)$$

One of the interesting features of many-dimensional systems is the opportunity for a proliferation of first integrals. The two great classical systems, the harmonic oscillator and the Kepler problem, possess a sufficient number of first integrals to enable the orbit equation to be derived without painful integration. In the case of the oscillator, be it time-independent[13] or time-dependent,[14] the angular momentum and a matrix with elements

$$A_{ij}=q_iq_j+p_ip_j \qquad (1.6)$$

(in the case $H=\frac{1}{2}\mathbf{p}^2+\frac{1}{2}\mathbf{q}^2$) are conserved quantities. For the Kepler problem the conserved quantities are the energy, angular momentum, and the Laplace–Runge–Lenz vector. In two recent papers Katzin and Levine[11,12] have extended the existence of the same number of first integrals to the time-dependent Kepler problem with Hamiltonian

$$H=\tfrac{1}{2}\mathbf{p}^2-\mu_0/(\alpha t+\beta)r \qquad (1.7)$$

and also to the as yet nameless system

$$H=\tfrac{1}{2}\mathbf{p}^2-\tfrac{1}{2}\ddot U r^2/U-\mu_0/Ur, \qquad (1.8)$$

where μ_0,α,β are constants, U is an arbitrary function of time, and $r=|\mathbf{q}|$.

In the various papers cited above there are gaps. In Refs. 3 and 10 the generalization of (1.1) and (1.2) to three (or many) dimensions is obtained, but there is no indication of the special cases which give more first integrals as are found in Refs. 11 and 12. However, Ref. 11 fails to obtain all of the first integrals in one process. The purpose of this note is to fill in these gaps and to provide a unified treatment of the central force system described by the Hamiltonian

$$H=\tfrac{1}{2}\mathbf{p}^2+V(r,t) \qquad (1.9)$$

as far as determining all first integrals which are quadratic in the momentum. In addition we shall provide a simple explanation of the results and provide the solution of the corresponding quantum mechanical problem. The orbit equation will be obtained for one class of potentials.

II. THE CHOICE OF METHOD

The first question which arises in this. What method will be used to determine the first integrals? The direct method has been used in Refs. 10 and 11 for multidimensional systems and Noether's theorem with point transformations in Ref. 11.

The method which we choose to use here is the version of Noether's theorem which guarantees completeness, i.e., there is a one-to-one correspondence betwen first integrals and equivalence classes of symmetries for a given Lagrangian. For a full discussion we refer the reader to the masterly review of Noether's theorem by Sarlet and Cantrijn.[15] We summarize the relevant results. If an infinitesimal transformation

$$\bar{t} = t + \epsilon\xi(t, \mathbf{q}, \dot{\mathbf{q}}), \quad \bar{q}^i = q^i + \epsilon\eta^i(t, \mathbf{q}, \dot{\mathbf{q}}) \tag{2.1}$$

leaves the action integral

$$A = \int_{t_1}^{t_2} L(t, \mathbf{q}(t), \dot{\mathbf{q}}(t))dt, \tag{2.2}$$

in which the Lagrangian L is assumed regular, invariant up to gauge terms, then there exists a constant of the motion given by

$$I(t, \mathbf{q}, \dot{\mathbf{q}}) = f(t, \mathbf{q}, \dot{\mathbf{q}}) - \left[L\xi + \frac{\partial L}{\partial \dot{q}^i}(\eta^i - \dot{q}^i\xi)\right]. \tag{2.3}$$

(The method of calculating ξ, η^i, and f is found at the beginning of the Appendix.) The way the theorem has been stated, the existence of a Noether symmetry implies the existence of the first integral I. However, Sarlet and Cantrijn (See Ref. 15, Lemma 6.2) have shown that, if I is an arbitrary constant of the motion of the Euler–Lagrange equations derived from (2.2), then all symmetries corresponding to I are determined by

$$\eta^i - \dot{q}^i\xi = g^{ij}\frac{\partial I}{\partial \dot{q}^j}, \tag{2.4}$$

where g^{ij} is defined by

$$\frac{\partial^2 L}{\partial \dot{q}^i\, \partial \dot{q}^j} g^{jk} = \delta_i^k \tag{2.5}$$

(hence the necessity for L to be regular). Furthermore (see Ref. 15, p. 487) the choice of the time component ξ is completely free and we are at liberty to set $\xi = 0$ so that (2.4) becomes

$$\eta^i = -g^{ij}\frac{\partial I}{\partial \dot{q}^j}. \tag{2.6}$$

In summary, provided we permit velocity-dependent infinitesimal transformations, Noether's theorem provides a one-to-one correspondence between the existence of a Noether symmetry and the existence of a first integral.

We recall that we are looking for first integrals quadratic in the momenta [which for H (1.9) are equivalent to the velocities]. The Lagrangian corresponding to H (1.9) is

$$L = \tfrac{1}{2}\dot{\mathbf{r}}^2 - V(r, t), \tag{2.7}$$

which is regular, and g^{jk} is independent of the velocity components. From (2.6) the choice of the nature of the velocity dependence of the first integral, viz., quadratic, means that η^i is linear in the velocity components. (This is an example of the systematic approach proposed by Kobussen.[16]) Since L (2.7) satisfies the requirements of Noether's theorem as stated above, if we find all η^i linear in the velocities and the corresponding f's which satisfy Eqs. (A1) and (A2), then (2.3) will provide all first integrals of L which are quadratic in the velocities.

Before commencing the calculations we may simplify them by making use of the radial symmetry of the potential $V(r, t)$. *A priori* we know that the angular momentum $\mathbf{l}$ is a first integral. In particular the vector $\mathbf{l}$ is constant and so the motion is in a plane. We choose coordinates (r, θ) and origin such that the origin lies in this plane and θ is the angular displacement in the plane. This reduces the number of functions η^i to two, thereby making (A1) and (A2) more manageable. With this choice of coordinates we define

$$\eta^1 = a\dot{r} + b\dot{\theta} + c, \quad \eta^2 = u\dot{r} + v\dot{\theta} + w, \tag{2.8}$$

where the functions a,b,c,u,v,w are functions for r,θ, and t only, η^1 corresponds to $r \to \bar{r}$, and η^2 corresponds to $\theta \to \bar{\theta}$.

III. CLASSES OF FIRST INTEGRALS AND THE CORRESPONDING POTENTIALS

The determination of the functions η^1, η^2, and f is, as usual in multidimensional problems, a lengthy and complicated task (cf. Ref. 10). There are 12 first-order linear partial differential equations to be solved. The details are found in the Appendix. We distinguish three classes of potential.

Class I: $V(r, t) = \tfrac{1}{2}\lambda(t)r^2$. Substituting for η^1, η^2, and f into (2.3) we obtain the first integral

$$\begin{aligned} -I = {} & \tfrac{1}{2}(\phi\dot{r} - \dot{\phi}r)^2 + \tfrac{1}{2}\phi^2 r^2\dot{\theta}^2 + \tfrac{1}{2}(K_5 r^2/\phi^2) + \sin 2\theta\,\{\tfrac{1}{2}(\tau\dot{r} - \dot{\tau}r)^2 - \tfrac{1}{2}\tau^2 r^2\dot{\theta}^2 + \tfrac{1}{2}(K_3 r^2/\tau^2)\} + \cos 2\theta\,(\tau\dot{r} - \dot{\tau}r)\tau r\dot{\theta} \\ & - \sin 2\theta\,(\zeta\dot{r} - \zeta r)\zeta r\dot{\theta} + \cos 2\theta\,\{\tfrac{1}{2}(\zeta\dot{r} - \dot{\zeta}r)^2 - \tfrac{1}{2}\zeta^2 r^2\dot{\theta}^2 + \tfrac{1}{2}(K_4 r^2/\zeta^2)\} + \{\sin\theta\,(\delta\dot{r} - \dot{\delta}r) + \cos\theta\,\delta r\dot{\theta}\}r^2\dot{\theta} \\ & + \{-\sin\theta\,\epsilon r\dot{\theta} + \cos\theta\,(\epsilon\dot{r} - \dot{\epsilon}r)\}r^2\dot{\theta} + \sin\theta\,(\dot{\sigma}r - \sigma\dot{r}) - \sigma r\dot{\theta}\cos\theta + \sin\theta\,\rho r\dot{\theta} + \cos\theta\,(\dot{\rho}r - \rho\dot{r}) \\ & + K_1(r^2\dot{\theta})^2 - K_2 r^2\dot{\theta}, \end{aligned} \tag{3.1}$$

where, if $\chi(t)$ is a solution of the so-called auxiliary equation[17]

$$\ddot{\chi} + \lambda\chi = \chi^{-3} \tag{3.2}$$

and

$$T = \int^t \chi^{-2}(t')dt', \tag{3.3}$$

then

$$\delta = A_1\chi \sin T + B_1\chi \cos T, \tag{3.4}$$

$$\epsilon = A_2\chi \sin T + B_2\chi \cos T, \tag{3.5}$$

$$\sigma = A_3\chi \sin T + B_3\chi \cos T, \tag{3.6}$$

$$\rho = A_4\chi \sin T + B_4\chi \cos T, \tag{3.7}$$

$$\tau^2 = K_3^{1/2}\chi^2(A_5 \sin^2 T + B_5 \cos^2 T + 2C_5 \sin T \cos T), \tag{3.8}$$

$$\zeta^2 = K_4^{1/2}\chi^2(A_6 \sin^2 T + B_6 \cos^2 T + 2C_6 \sin T \cos T), \tag{3.9}$$

and the constants are arbitrary subject to the constraints

$$A_5B_5 - C_5^2 = 1, \quad A_6B_6 - C_6^2 = 1. \tag{3.10}$$

The expressions associated with $\phi,\tau,\zeta,\delta,\epsilon,\sigma,\rho,K_1$, and K_2 are each first integrals. The conserved angular momentum is readily identified. The integral containing ϕ is an energylike expression. The integrals involving τ and ζ resemble those found for the oscillator in Cartesian coordinates.[18] The linear expressions containing δ,ϵ,σ, and ρ are essentially the initial conditions and are independent (or, more strictly, are a combination of four linearly independent first integrals). The fact that any two of (3.4)–(3.7) and any two of the corresponding integrals in (3.1) are independent means that we do in fact have four independent integrals. Furthermore, neither τ nor ζ need be the same as χ (up to a constant multiplier). This is a feature which has not been so evident in other studies of this nature.

Case II: $V(r,t) = \frac{1}{2}\ddot{\phi}r^2/\phi - \mu_0/\phi r$. The first integral corresponding to this potential is given by

$$\begin{aligned} -I = {} & \frac{1}{2}(\phi\dot{r} - \dot{\phi}r)^2 + \frac{1}{2}\phi^2 r^2\dot{\theta}^2 - \frac{\mu_0\phi}{r} + K_7\{r^3\dot{\theta}^2\phi\cos\theta \\ & + r^2\dot{r}\dot{\theta}\phi\sin\theta - r^3\dot{\theta}\dot{\phi}\sin\theta - \mu_0\cos\theta\} \\ & + K_8\{-r^3\dot{\theta}^2\phi\sin\theta + r^2\dot{r}\dot{\theta}\phi\cos\theta - r^3\dot{\theta}\dot{\phi}\cos\theta \\ & - \mu_0\sin\theta\} - K_2r^2\dot{\theta} + \tfrac{1}{2}K_1(r^2\dot{\theta})^2. \end{aligned} \tag{3.11}$$

Four independent first integrals occur. The conserved angular momentum is obvious. The expression without a constant multiplier is an energylike first integral. If we denote the angular momentum by $\mathbf{l}$, the coefficients of K_7 and K_8 may be combined to give the conserved vector

$$\boldsymbol{\sigma} = \phi\,\mathbf{l}\times\dot{\mathbf{r}} - \dot{\phi}\,\mathbf{l}\times\mathbf{r} + \mu_0\hat{\mathbf{r}}, \tag{3.12}$$

which accords with the result of Katzin and Levine.[11]

Case III: $V(r,t) = -\frac{1}{2}(\ddot{\phi}/\phi)r^2 + \phi^{-2}U(r/\phi)$. The first integral is given by

$$-I = \tfrac{1}{2}(\phi\dot{r} - \dot{\phi}r)^2 + \tfrac{1}{2}r^2\dot{\theta}^2\phi^2 + U(r/\phi) + \tfrac{1}{2}K_1(r^2\dot{\theta})^2 - K_2r^2\dot{\theta}, \tag{3.13}$$

i.e., an energylike first integral and the conserved angular momentum.

IV. A SIMPLE EXPLANATION FOR THE EXISTENCE OF THE FIRST INTEGRALS

The existence of the various first integrals associated with the three types of potential in Sec. III may, now that we know what they are, be explained in terms of a simple generalized canonical transformation. Consider the Hamiltonian

$$\bar{H}(\mathbf{R},\mathbf{P},T) = \tfrac{1}{2}(P_R^2 + P_\Theta^2 R^{-2}) + U(R). \tag{4.1}$$

Under the generalized canonical transformation

$$\begin{aligned} (R,\Theta,P_R,P_\Theta,T) \mapsto \Big((r,\theta,p_r,p_\theta,t:\ & R = r\chi^{-1},\ \Theta = \theta, \\ & P_R = \chi p_r - \dot{\chi}r,\ P_\Theta = p_\theta, \\ & T = \int^t \chi^{-2}(t')dt' \Big), \end{aligned} \tag{4.2}$$

we obtain

$$H = \tfrac{1}{2}(p_r^2 + p_\theta^2 r^{-2}) - \tfrac{1}{2}(\ddot{\chi}/\chi)r^2 + \chi^{-2}U(r/\chi). \tag{4.3}$$

The system (4.1) has its Hamiltonian and the angular momentum as first integrals for general U. In terms of the transformed coordinates these first integrals are

$$I_1 = P_\Theta = R^2\frac{d\Theta}{dT} = \frac{r^2}{\chi^2}\dot{\theta}\chi^2 = r^2\dot{\theta} \tag{4.4}$$

and

$$\begin{aligned} I_2 = h &= \tfrac{1}{2}(P_R^2 + P_\Theta^2 R^{-2}) + U(R) \\ &= \tfrac{1}{2}\{(\chi p_r - \dot{\chi}r)^2 + p_\theta^2\chi^2 r^{-2}\} + U(r\chi^{-1}). \end{aligned} \tag{4.5}$$

This immediately gives the result for case III.

For case I, the potential is specified and we require the potential in (4.3) to match it, i.e., we seek a function U such that

$$-\tfrac{1}{2}(\ddot{\chi}/\chi)r^2 + \chi^{-2}U(r/\chi) = \tfrac{1}{2}\lambda r^2. \tag{4.6}$$

It is evident that

$$U(r/\chi) = \tfrac{1}{2}K(r/\chi)^2, \tag{4.7}$$

so that χ is a solution of

$$\ddot{\chi} + \lambda\chi = K\chi^{-3}, \tag{4.8}$$

which is an equation of the type in (A31)–(A33). To simplify the discussion we set $K = 1$. Then $\bar{H}$ is

$$\bar{H} = \tfrac{1}{2}(P_R^2 + P^2R_\Theta^{-2} + R^2), \tag{4.9}$$

which is just the Hamiltonian of the two-dimensional harmonic oscillator. Apart from the integrals already given in (4.4) and (4.5), $\bar{H}$ has as integrals quadratic in the momenta, the elements of the Jauch–Hill–Fradkin tensor.[13,19] Expressed in polar coordinates, the two integrals apart from the energylike integral are

$$I_3 = \tfrac{1}{2}\sin 2\Theta(R^2 + P_R^2 - P_\Theta^2 R^{-2}) + P_R P_\Theta R^{-1}\cos 2\Theta, \tag{4.10}$$

$$I_4 = \tfrac{1}{2}\cos 2\Theta(R^2 + P_R^2 - P_\Theta^2 R^{-2}) - P_R P_\Theta R^{-1}\sin 2\Theta. \tag{4.11}$$

Under the transformation given in (4.2), these become

$$\begin{aligned} I_3 = {} & \tfrac{1}{2}\sin 2\theta\{(\chi p_r - \dot{\chi}r)^2 - \chi^2 P_\theta^2 r^{-2} + r^2\chi^{-2}\} \\ & + (\chi p_r - \dot{\chi}r)p_\theta\chi r^{-1}\cos 2\theta \end{aligned} \tag{4.12}$$

and

$$\begin{aligned} I_4 = {} & \tfrac{1}{2}\cos 2\theta\{(\chi p_r - \dot{\chi}r)^2 - \chi^2 p_\theta^2 r^{-2} + r^2\chi^{-2}\} \\ & - (\chi p_r - \dot{\chi}r)p_\theta\chi r^{-1}\cos 2\theta. \end{aligned} \tag{4.13}$$

These two integrals have the same structure as the corresponding first integrals in (3.1). Further, $\bar{H}$ also has four first integrals which are linear in the momentum and which represent the initial conditions. In terms of polar coordinates they are

$$I_5 = R \cos \Theta \cos T - (P_R \cos \Theta - P_\Theta R^{-1} \sin \Theta) \sin T, \tag{4.14}$$

$$I_6 = R \cos \Theta \sin T + (P_R \cos \Theta - P_\Theta R^{-1} \sin \Theta) \cos T, \tag{4.15}$$

$$I_7 = R \sin \Theta \cos T - (P_R \sin \Theta + P_\Theta R^{-1} \cos \Theta) \sin T, \tag{4.16}$$

$$I_8 = R \sin \Theta \sin T + (P_R \sin \Theta + P_\Theta R^{-1} \cos \Theta) \cos T. \tag{4.17}$$

From the transformation (4.2) we see that these become

$$I_5 = \sin \theta p_\theta r^{-1}(\chi \sin T) + \cos \theta \{ r(\chi^{-1} \cos T + \dot{\chi} \sin T) - p_r \chi \sin T \}, \tag{4.18}$$

$$I_6 = -\sin \theta p_\theta r^{-1}(\chi \cos T) + \cos \theta \{ p_r \chi \cos T - r(\dot{\chi} \cos T - \chi^{-1} \sin T) \}, \tag{4.19}$$

$$I_7 = -\cos \theta p_\theta r^{-1}(\chi \sin T) + \sin \theta \{ r(\chi^{-1} \cos T + \dot{\chi} \sin T) - p_r \chi \sin T \}, \tag{4.20}$$

$$I_8 = \cos \theta p_\theta r^{-1}(\chi \cos T) + \sin \theta \{ p_r \chi \cos T - r(\dot{\chi} \cos T - \chi^{-1} \sin T) \}. \tag{4.21}$$

Since χ is a solution of (4.8) it follows that $\chi \sin T$ and $\chi \cos T$ are solutions of the differential equation

$$\ddot{\psi} + \lambda \psi = 0, \tag{4.22}$$

which is of the same type as (A27), (A29), and (A30) and so the integrals I_5–I_8 have the form of the integrals in (3.1) which are linear in the momenta.

In case II the potential also is specified and we require the potential in (4.3) to be such that

$$-\frac{1}{2}\frac{\ddot{\chi}}{\chi} r^2 + \frac{1}{\chi^2} U\left(\frac{r}{\chi}\right) = -\frac{1}{2}\frac{\ddot{\phi}}{\phi} r^2 - \frac{\mu_0}{\phi r}. \tag{4.23}$$

clearly this occurs only if

$$\chi = \phi, \quad U(r/\phi) = -\mu_0(\phi/r). \tag{4.24}$$

Now the Hamiltonian

$$\overline{H} = \tfrac{1}{2}(P_R^2 + P_\Theta^2 R^{-2}) - \mu_0 R^{-1} \tag{4.25}$$

has, in addition to the energy and angular first integrals, the conserved Laplace–Runge–Lenz vector

$$\mathbf{I} = \mathbf{P}_\Theta \times \mathbf{P} + \mu_0 \hat{\mathbf{R}}. \tag{4.26}$$

Under the transformation (4.2)

$$\hat{\mathbf{R}} = \hat{\mathbf{r}}, \quad \mathbf{P} = \phi \mathbf{p} - \dot{\phi} \mathbf{r}, \tag{4.27}$$

and so the conserved vector is

$$\mathbf{I} = \phi \mathbf{p}_\theta \times \mathbf{p} - \dot{\phi} \mathbf{p}_\theta \times \mathbf{r} + \mu_0 \hat{\mathbf{r}}, \tag{4.28}$$

which is the conserved vector given in (3.56).

In summary, we have seen that all of the first integrals quadratic (or up to quadratic) in the momenta for a time-dependent potential of permitted form for a central force field are obtained by the action of the simple point transformation (4.2) on the first integral of the corresponding time-independent problem.

V. ORBIT EQUATIONS

The orbit equation for case III cannot usually be found by simple quadrature, a situation which already occurs in time-independent problems. The use of the quadratic first integrals in Cartesian coordinates to determine the orbit equation in case I was discussed some time ago in Günther and Leach[14] and need not be reproduced here. However, it does seem to be appropriate to derive the orbit equation for case II by a method which is simpler than that given in Katzin and Levine.[11] We recall that we have the conserved vector

$$\mathbf{I} = \phi \mathbf{l} \times \mathbf{p} - \dot{\phi} \mathbf{l} \times \mathbf{r} + \mu_0 \hat{\mathbf{r}}. \tag{5.1}$$

Taking the scalar product of $\mathbf{I}$ with $\mathbf{r}$ (see Ref. 20) and taking $\hat{\mathbf{I}}$ as the direction of the reference line in the plane of the orbit, we have

$$\begin{aligned} Ir \cos \theta &= \mathbf{I} \cdot \mathbf{r} \\ &= \phi (\mathbf{l} \times \mathbf{p}) \cdot \mathbf{r} - \dot{\phi} (\mathbf{l} \times \mathbf{r}) \cdot \mathbf{r} + \mu_0 \hat{\mathbf{r}} \cdot \mathbf{r} \\ &= -l^2 \phi + \mu_0 r. \end{aligned} \tag{5.2}$$

The orbit is given by

$$r^{-1} = (\mu_0 - I \cos \theta)/(l^2 \phi). \tag{5.3}$$

We may in principle determine ϕ from the angular momentum since

$$l = r^2 \dot{\theta} \tag{5.4}$$

may be rewritten as

$$\frac{d\theta}{(\mu_0 - I \cos \theta)^2} = \frac{dt}{l^3 \phi^2}. \tag{5.5}$$

The left-hand side of (5.5) is integrable although the expression is not awfully tidy. Let us suppose that the integrated form of (5.5) is

$$M(\theta) = N(t). \tag{5.6}$$

Assuming that we can invert (5.6) to obtain t as a function of θ, something which usually will be locally possible, we have

$$\phi(t) = \phi \circ N^{-1} \circ M(\theta) \tag{5.7}$$

and the orbit equation

$$1/r = (\mu_0 - I \cos \theta)/[l^2 \phi \circ N^{-1} \circ M(\theta)]. \tag{5.8}$$

The only problems in a practical implementation of this scheme are the integration of $\phi^{-2}(t)$ and the inversion of $N(t)$.

For an example let us consider the case for which $\mu_0 > I$. then

$$\mathrm{M}(\theta) = \frac{I}{\mu_0^2 - I^2} \frac{\sin \theta}{\mu_0 - I \cos \theta} + \frac{2\mu_0}{(\mu_0^2 - I^2)^2} \arctan\left\{ \left(\frac{\mu_0 + \mathrm{I}}{\mu_0 - \mathrm{I}} \right)^{1/2} \tan \frac{\theta}{2} \right\}. \tag{5.9}$$

Let

$$\phi(t) = l^{-3/2}(a + b \cos t)^{1/2}, \quad a > |b| > 0. \tag{5.10}$$

Then

$$N(t) = \int_0^t \frac{dt'}{l^3 \phi^2(t')} = \frac{2}{(a^2 - b^2)^{1/2}} \arctan\left(\frac{a - b}{a + b} \right)^{1/2} \tan \frac{t}{2} \tag{5.11}$$

and so

$$\phi^2(t) = l^{-3}\{a + b\cos^2 \arctan(a+b)/(a-b)^{1/2}$$
$$\times \tan[(a^2-b^2)^{1/2}]M(\theta)\}, \qquad (5.12)$$

where $M(\theta)$ is the right-hand side of (5.9). To simplify the discussion we fix $a^2 - b^2$ by writing

$$a^2 - b^2 = (\mu_0^2 - I^2)^3/\mu_0^2 \qquad (5.13)$$

and for ease of notation define

$$w = \frac{1}{2}(a^2-b^2)^{1/2}\frac{I}{\mu_0^2 - I^2}\frac{\sin\theta}{\mu_0 - I\cos\theta}. \qquad (5.14)$$

Then (5.12) simplifies to

$$\phi^2(t) = l^{-3}\{(\mu_0^2 - I^2)^3/\mu_0^2\}(\mu_0 - I\cos\theta)$$
$$\times [a(\mu_0 - I\cos\theta) + b\{(I - \mu_0\cos\theta)\cos 2w$$
$$+ (\mu_0^2 - I^2)^{1/2}\sin\theta\sin 2w\}]^{-1}. \qquad (5.15)$$

It is evident from (5.15) that $\phi^2(t)$ is a periodic function of θ of period 2π. Hence from (5.8) it is obvious that $r(\theta)$ is a periodic function of θ and so the orbit is closed. An example of such an orbit is given in the Fig. 1.

VI. SOLUTION OF THE CORRESPONDING SCHRÖDINGER EQUATION

Solutions of the Schrödinger equation for time-dependent potentials have been obtained in some instances.[21,22] For the potentials of the types discussed here, the procedure for solving the Schrödinger equation is particularly simple. Wolf[23] has shown that for time-independent systems, solutions of Schrödinger equations of systems whose classical Hamiltonians are related by time-independent linear transformations are related by integral transforms except when the transformation is a point of transformation. In this case the transform collapses to a geometric transform, i.e., the wave functions are related by a phase factor with possibly a scaling factor. Leach[21] extended this result to time-dependent systems and time-dependent linear transformations.

For the three cases obtained in Sec. III we showed in Sec. IV that they were related to time-independent systems by means of a generalized canonical transformation which is pointlike in the canonically conjugate variables. We consider case III first as this illustrates the general method. For cases I and II we can obtain the actual wave functions as the related time-independent systems are fully integrable.

For case III we have

$$H(\mathbf{r},\mathbf{p},t) = \tfrac{1}{2}(p_r^2 + p_\theta^2 r^{-2}) - \tfrac{1}{2}(\ddot{\chi}/\chi)r^2 + \chi^{2-1}U(r/\chi), \qquad (6.1)$$

$$\overline{H}(\mathbf{R},\mathbf{P},T) = \tfrac{1}{2}(P_R^2 + P_\Theta^2 r^{-2}) + U(R), \qquad (6.2)$$

which are related by the generalized canonical transformation

$$(R,\Theta,P_R,P_\Theta,T) \to \Big(r,\theta,p_r,p_\theta,t\colon\ R = r\chi^{-1},\ \Theta = \theta,$$
$$P_R = \chi p_r - \dot{\chi} r,\ P_\Theta = p_\theta,$$
$$T = \int^t \chi^{-2}(t')dt'\Big) \qquad (6.3)$$

The Schrödinger equation corresponding to $\overline{H}$ is

$$\hat{\overline{H}}\,\overline{\psi} = i\hbar\frac{\partial\overline{\psi}}{\partial T}. \qquad (6.4)$$

For two spatial dimensions, we may separate variables by writing

$$\overline{\psi}(R,\Theta,T) = e^{-i\lambda T/\hbar}e^{im\Theta}\phi(R)R^{-1/2} \qquad (6.5)$$

so that $\phi(R)$ is a solution of the ordinary differential equation

$$\phi'' + \left(\frac{2\lambda}{\hbar^2} - \frac{2}{\hbar^2}U - \frac{m^2 - \frac{1}{4}}{R^2}\right)\phi = 0. \qquad (6.6)$$

[For three dimensions, $Y_{lm}(\theta,\phi)$ and the appropriate Laplacian would be used. There is no essential difference in the treatment.] The problem is to find the eigenvalues and eigenfunctions of (6.6). For general U this is not possible in closed form. As two potentials for which an exact solution is available are the harmonic oscillator and the Coulomb problem, we shall obtain exact solutions for cases I and II. As an aside we remark that for certain classes of anharmonic oscillators an exact solution is also available.[24] Let us suppose that we have found $\phi(R)$. The solution (6.5) and the solution of the Schrödinger equation for H are related by[21]

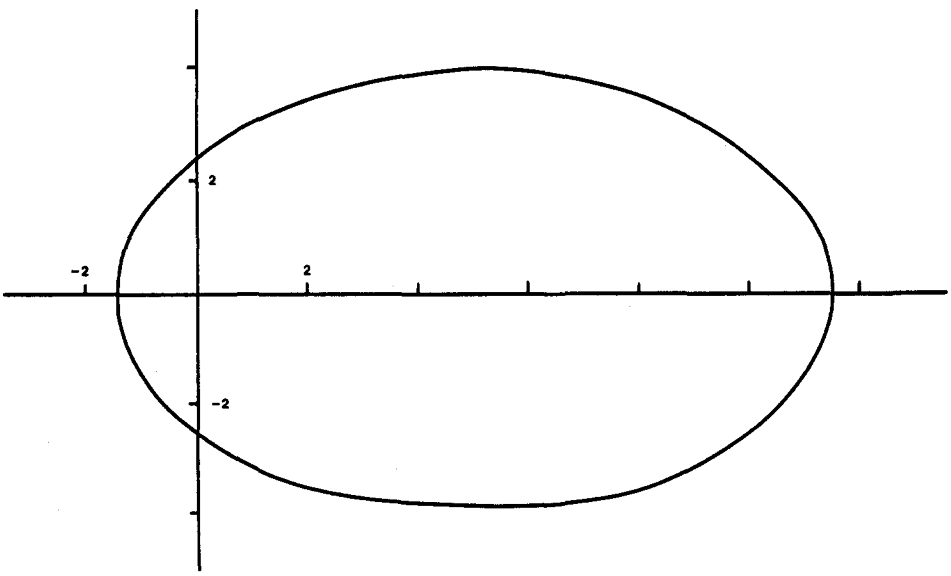


FIG. 1. Sketch of the closed orbit for the potential $V(r,t) = -\frac{1}{2}(\ddot{\phi}/\phi)r^2 - \mu_0/\phi r$ when $\phi(t) = (a + b\cos t)^{1/2}$, with $a = 20$, $b = 1$, and $\mu_0 = 9$.

$$\psi(r,\theta,t) = |\chi|^{-1/2}\exp\left[\frac{i}{2\hbar}\frac{\dot{\chi}}{\chi}r^2\right]\overline{\psi}\left(\frac{r}{\chi},\theta,T(t)\right) \quad (6.7)$$

so that provided we can find the solution of (6.6) we know $\psi(r,\theta,t)$.

For case I,

$$U(R) = \tfrac{1}{2}R^2 \quad (6.8)$$

(we set the constant K at unity for simplicity) and (6.6) is now

$$\phi'' + \left(\frac{2\lambda}{\hbar^2} - \frac{1}{\hbar^2}R^2 - \frac{m^2 - \frac{1}{4}}{R^2}\right)\phi = 0. \quad (6.9)$$

The eigenfunctions of (6.9) (see Ref. 25, p. 781, 22.6.18) (to within a normalizing factor) and the eigenvalues are

$$\phi_{n,m}(R) = \exp(-R^2/2\hbar)R^{m+1/2}L_n^{(m)}(R^2/\hbar), \quad (6.10)$$

$$\lambda_{n,m} = \hbar(2n + m + 1), \quad (6.11)$$

where $L_n^{(m)}$ is the generalized Laguerre polynomial. Hence the solution for the time-dependent potential of case I is, to within a normalizing factor,

$$\psi_{n,m}(r,\theta,t) = |\chi|^{-1/2}\exp\left[\frac{i}{2\hbar}\frac{\dot{\chi}}{\chi}r^2 - \frac{r^2}{2\hbar\chi^2}\right]\left[\frac{r}{\chi}\right]^m L_n^{(m)}\left(\frac{r^2}{\hbar\chi^2}\right)$$
$$\times e^{im\theta}\exp[-i(2n + m + 1)T(t)/\hbar]. \quad (6.12)$$

For case II,

$$U(R) = -\mu_0/R, \quad (6.13)$$

and (6.6) is

$$\phi'' + \left(\frac{2\lambda}{\hbar^2} + \frac{\mu_0}{\hbar^2 R} - \frac{m^2 - \frac{1}{4}}{R^2}\right)\phi = 0. \quad (6.14)$$

The eigenfunctions of (6.14) (see Ref. 25, p. 781, 22.6.17) (to within a normalizing factor) and the eigenvalues are

$$\phi_{n,m}(R) = \exp[-2kR]R^{m+1/2}L_n^{(2m)}(kR), \quad (6.15)$$

$$\lambda_{n,m} = -\mu_0^2/2\hbar^2(2n + 2m + 1)^2, \quad (6.16)$$

where again $L_n^{(2m)}$ is a generalized Laguerre polynomial and the constant k is given by

$$k = \mu_0/\hbar^2(2n + 2m + 1). \quad (6.17)$$

Hence the solution of the Schrödinger equation for the time-dependent potential of case II is, to within a normalizing factor,

$$\psi_{n,m}(r,\theta,t) = |\chi|^{-1/2}\exp\left[\frac{i}{2\hbar}\frac{\dot{\chi}}{\chi}r - \frac{2kr}{\chi}\right]\left(\frac{r}{\chi}\right)^m L_n^{(2m)}\left(\frac{kr}{\chi}\right)$$
$$\times e^{im\theta}\exp[-i\mu_0^2 T(r)/2\hbar^3(2n + 2m + 1)^2]. \quad (6.18)$$

As a final remark on the solution of the Schrödinger equations with the particular time-dependent potentials given above, we point out that it is not necessary to calculate matrix elements and expectation values in terms of the wave functions $\psi_{n,m}(r,\theta,t)$ (see Ref. 21). Writing the states $\psi_{\mathbf{n}}$ and $\overline{\psi}_{\mathbf{n}}$ in the Dirac notation as $|\mathbf{n}\rangle$ and $|\overline{\mathbf{n}}\rangle$, respectively, we have

$$\langle\mathbf{m}|f(\mathbf{r},\mathbf{p})|\mathbf{n}\rangle = \langle\overline{\mathbf{m}}|F(\mathbf{R},\mathbf{p})|\overline{\mathbf{n}}\rangle, \quad (6.19)$$

where

$$F[\mathbf{R}(\mathbf{r},t),\mathbf{P}(\mathbf{r},\mathbf{p},t)] = f(\mathbf{r},\mathbf{p}). \quad (6.20)$$

So, for example,

$$\langle\mathbf{m}|r|\mathbf{n}\rangle = \langle\overline{\mathbf{m}}|\chi R|\overline{\mathbf{n}}\rangle = \chi\langle\overline{\mathbf{m}}|R|\overline{\mathbf{n}}\rangle. \quad (6.21)$$

This simplifies the calculations to an extent since the r-dependent phase term is removed from the calculation.

VII. CONCLUSION

In this paper we have obtained all first integrals quadratic (and linear) in the momenta for the potential of a central force field and also the potentials for which such integrals exist. This was done using the form of Noether's theorem discussed by Sarlet and Cantrijn[15] which guarantees a one-to-one correspondence between Noether symmetries and first integrals. The existence of such potentials and first integrals has been explained in terms of a simple time-dependent linear generalized canonical transformation. In view of the contents of the Appendix, it may well be asked why the latter method was not used rather than the former. We have already discussed the completeness property of the appropriate formulation of Noether's theorem. There is not a corresponding result for canonical transformations. From Noether's theorem we know that we will obtain integrals quadratic in the momenta if the infinitesimal transformations are linear in the momenta (in the case of the Lagrangian used here). However, in the context of canonical transformations (again for the type of Hamiltonian used here), the *only if* fails as can be seen by a simple counterexample. For

$$H = \tfrac{1}{2}p^2 + \tfrac{1}{2}q^2, \quad \overline{H} = \tfrac{1}{2}P^2,$$

the transformation from H to $\overline{H}$ can be achieved by

$$Q = \tfrac{1}{2}\arcsin\{q/\sqrt{q^2 + p^2}\}, \quad P^2 = q^2 + p^2, \quad (7.1)$$

which is not linear.

For case I it has already been shown[14] how the third set of integrals (the first being the energylike integral and the second the angular momentum) may be used to obtain the orbit equation. In this paper we have shown that the orbit equation for case II may be obtained also using the third set of integrals. Finally, we demonstrated how easy it was to solve the corresponding Schrödinger equation.

ACKNOWLEDGMENT

It is a pleasure to thank the referee for useful and constructive suggestions.

APPENDIX: DETERMINATION OF ξ, η^i, AND F

The functions ξ, η^i, and f in (2.3) are determined from the partial differential equations

$$L\frac{\partial\xi}{\partial\dot{q}^j} + \frac{\partial L}{\partial\dot{q}^i}\left(\frac{\partial\eta^i}{\partial\dot{q}^j} - \dot{q}^i\frac{\partial\xi}{\partial\dot{q}^j}\right) = \frac{\partial f}{\partial\dot{q}^j}, \quad (A1)$$

$$\xi\frac{\partial L}{\partial t} + \eta^i\frac{\partial L}{\partial q^i} + L\left(\frac{\partial\xi}{\partial t} + \dot{q}^i\frac{\partial\xi}{\partial q^i}\right) + \frac{\partial L}{\partial\dot{q}^i}\left\{\frac{\partial\eta^i}{\partial t} + \frac{\partial\eta^i}{\partial q^j}\dot{q}^j - \dot{q}^i\left(\frac{\partial\xi}{\partial t} + \dot{q}^j\frac{\partial\xi}{\partial q^j}\right)\right\} = \frac{\partial f}{\partial t} + \dot{q}^j\frac{\partial f}{\partial q^j}. \quad (A2)$$

Substituting L (2.7) and the η^i (2.8) into (A1) and integrating, we find that

$$f = \tfrac{1}{2}\dot{r}^2 a + b\dot{r}\dot{\theta} + \tfrac{1}{2}r^2 v\dot{\theta}^2 + g(r,\theta,t), \tag{A3}$$

$$b = r^2 u, \tag{A4}$$

where g is an arbitrary function. The integration of (A2) is achieved by equating the coefficients of powers of $\dot{r}$ and $\dot{\theta}$ to zero after substitution with (2.7), (2.8), (A3), and (A4). The cubic and quadratic terms yield

$$a = A(\theta,t), \tag{A5}$$

$$b = \frac{1}{2} r \frac{\partial A}{\partial \theta} + r^2 B(\theta,t), \tag{A6}$$

$$c = -\frac{1}{2} r \frac{\partial A}{\partial t} - D(\theta,t), \tag{A7}$$

$$u = \frac{1}{2r}\frac{\partial A}{\partial \theta} + B, \tag{A8}$$

$$v = A + \frac{1}{2}\frac{\partial^2 A}{\partial \theta^2} + 2r\frac{\partial B}{\partial \theta} + r^2 C(\theta,t), \tag{A9}$$

$$w = -r\frac{\partial B}{\partial t} - \frac{1}{r}\frac{\partial D}{\partial \theta} - E(\theta,t), \tag{A10}$$

plus six differential equations for the functions $A(\theta,t)$ through $E(\theta,t)$ which may be integrated to give

$$A = \alpha + \beta \sin 2\theta + \gamma \cos 2\theta, \tag{A11}$$

$$B = \delta \sin \theta + \epsilon \cos \theta, \tag{A12}$$

$$C = K_1, \tag{A13}$$

$$D = \sigma \sin \theta + \rho \cos \theta, \tag{A14}$$

$$E = \nu + \tfrac{1}{2}\dot{\beta}\cos 2\theta - \tfrac{1}{2}\dot{\gamma}\sin 2\theta. \tag{A15}$$

In the right-hand side of (A11) through (A15), lowercase Greek letters represent functions of time and the uppercase Latin letter is a constant.

The remaining coefficients provide the three equations

$$-a\frac{\partial V}{\partial r} + \frac{\partial c}{\partial t} = \frac{\partial g}{\partial r}, \tag{A16}$$

$$-b\frac{\partial V}{\partial r} + r^2\frac{\partial w}{\partial r} = \frac{\partial g}{\partial \theta}, \tag{A17}$$

$$-c\frac{\partial V}{\partial r} = \frac{\partial g}{\partial t}. \tag{A18}$$

We may integrate (A16) to obtain

$$g = AV - \frac{1}{4}r^2\frac{\partial^2 A}{\partial t^2} - r\frac{\partial D}{\partial t} - F(\theta,t). \tag{A19}$$

Substituting this into (A17)

$$\frac{\partial F}{\partial \theta} = \left(\frac{1}{2} r\frac{\partial A}{\partial \theta} + r^2 B\right)\frac{\partial V}{\partial r} - \frac{\partial A}{\partial \theta}V + r^3\frac{\partial^2 B}{\partial t^2} + r^2\dot{\nu}. \tag{A20}$$

Since the left-hand side of (A20) is free of r, the derivative of the right-hand side with respect to r is zero. As A and B contain sine and cosine terms, the coefficient of each is separately zero and we find that for the coefficient of

$$\cos 2\theta:\quad \beta\left(r\frac{\partial^2 V}{\partial r^2} - \frac{\partial V}{\partial r}\right) = 0, \tag{A21}$$

$$\sin 2\theta:\quad \gamma\left(r\frac{\partial^2 V}{\partial r^2} - \frac{\partial V}{\partial r}\right) = 0, \tag{A22}$$

$$\cos \theta:\quad \epsilon\left(2r\frac{\partial V}{\partial r} + r^2\frac{\partial^2 V}{\partial r^2}\right) + 3\ddot{\epsilon}r^2 = 0, \tag{A23}$$

$$\sin \theta:\quad \delta\left(2r\frac{\partial V}{\partial r} + r^2\frac{\partial^2 V}{\partial r^2}\right) + 3\ddot{\delta}r^2 = 0, \tag{A24}$$

$$-:\quad \nu = K_2. \tag{A25}$$

We distinguish three distinct cases in the solutions of (A21)–(A25).

Case I: $\beta,\gamma,\delta,\epsilon \neq 0$,

$$V = \tfrac{1}{2}\lambda(t)r^2, \tag{A26}$$

where

$$\lambda(t) = -\ddot{\delta}/\delta = -\ddot{\epsilon}/\epsilon. \tag{A27}$$

Case II: $\beta,\gamma, = 0,\quad \delta,\epsilon \neq 0$,

$$V = \tfrac{1}{2}\lambda(t)r^2 - \mu(t)/r, \tag{A28}$$

with $\lambda(t)$ defined as in (A27). The possibility that $\lambda(t)$ is zero is contained in the more general case II. The possibility of $\mu(t)$ being identically zero is not contemplated for case II.

Case III:

$\beta,\gamma,\delta,\epsilon = 0.$

Equation (A17) provides no information about the potential.

For case I, $F(\theta,t)$ is an arbitrary function of time, which turns out to be a constant and is ignored. (The actual effect of this constant is to change the value of the first integral by an additive constant.) When we substitute for g, V, and C in (A18) and equate the coefficients of independent functions of r and θ to zero, we find that

$$\ddot{\sigma} + \lambda\sigma = 0, \tag{A29}$$

$$\ddot{\rho} + \lambda\rho = 0, \tag{A30}$$

$$\ddot{\tau} + \lambda\tau = K_3\tau^{-3}, \tag{A31}$$

$$\ddot{\zeta} + \lambda\zeta = K_4\zeta^{-3}, \tag{A32}$$

$$\ddot{\phi} + \lambda\phi = K_5\phi^{-3}, \tag{A33}$$

where β, γ, and α have been replaced by τ^2, ζ^2, and ϕ^2, respectively. In discussions of first integrals for time-dependent systems a function, say $\chi(t)$, is introduced. For example, the time-dependent harmonic oscillator with Hamiltonian

$$H = \tfrac{1}{2}p^2 + \tfrac{1}{2}w^2(t)q^2 \tag{A34}$$

has the first integral

$$I = \tfrac{1}{2}(\chi p - \dot{\chi}q)^2 + \tfrac{1}{2}q^2/\chi^2. \tag{A35}$$

For the problem under discussion here we have seven such differential equations. We can reduce the number of functions to two as follows. Let us take the solution $\phi(t)$ of (A33) to be

$$\phi(t) = K_5^{1/4}\chi(t). \tag{A36}$$

Then the solution of an equation of the form

$$\ddot{y} + \lambda y = 0 \tag{A37}$$

is

$$y = A\chi \sin T + B\chi \cos T, \tag{A38}$$

and of an equation of the form

$$\ddot{y} + \lambda y = Ky^{-3}, \tag{A39}$$

the solution is

$$y^2 = K^{1/2}\chi^2(A \sin^2 T + B \cos^2 T + 2C \sin T \cos T), \tag{A40}$$

where

$$T = \int^t \chi^{-2}(t')dt'. \tag{A41}$$

In (A40) the constants are arbitrary, subject to the constraint

$$AB - C^2 = 1. \tag{A42}$$

For case II we find that

$$F(\theta,t) = \mu\delta \sin\theta + \mu\epsilon \cos\theta, \tag{A43}$$

to within an arbitrary additive function of time, which again turns out to be a constant. We substitute for g, V, and c in (A18) and equate the coefficients of independent functions of r and θ to zero to find that

$$\sigma = 0 = \rho, \tag{A44}$$

$$\mu = \mu_0/\phi, \tag{A45}$$

$$\delta = K_7\phi, \quad \epsilon = K_8\phi, \tag{A46}$$

$$\ddot{\phi} + \lambda\phi = 0, \tag{A47}$$

where again α has been replaced by ϕ^2 and (A27) results from the integration of a third-order differential equation in α. In contrast to (A33) there is no constant of integration because of (A46).

For case III, $F(\theta,t)$ is again an arbitrary function of time which may be ignored as it vanishes from the first integral and has no dynamical effect elsewhere. Substituting for c, V, and g into (A18) we find that for the potential to differ from case I, σ and ρ must be zero. We then have a first-order partial differential equation for V, viz.,

$$\frac{1}{2} r\dot{\alpha}\frac{\partial V}{\partial r} + \alpha\frac{\partial V}{\partial t} = -\dot{\alpha}V - \frac{1}{4}r^2\ddot{\alpha}, \tag{A48}$$

for which the characteristics are

$$w_1 = r^2/\alpha, \quad w_2 = \alpha V + \tfrac{1}{4} w_1(\ddot{\alpha}\alpha - \tfrac{1}{2}\dot{\alpha}^2). \tag{A49}$$

The potential is

$$V(r,t) = -\tfrac{1}{2}(\ddot{\phi}/\phi)r^2 + \phi^{-2}U(r/\phi), \tag{A50}$$

where, as before, α has been replaced by ϕ^2 and U is an arbitrary function of its argument.

[1]H. R. Lewis, J. Math. Phys. **9**, 1976 (1968); H. R. Lewis, Phys. Rev. **172**, 1313 (1968).

[2](a) H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 165 (1982); (b) H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 2371 (1982); (c) H. R. Lewis and P. G. L. Leach, "A resonance formulation for invariants of particle motion in a one-dimensional time-dependent potential," preprint, Los Alamos National Laboratory, LA-UR-82-1417, 1982.

[3]P. G. L. Leach, H. R. Lewis, and W. Sarlet, J. Math. Phys. **25**, 486 (1984).

[4]W. Sarlet and L. Y. Bahar, Int. J. Non-Linear Mech. **15**, 133 (1980); W. Sarlet, Phys. Lett. A **82**, 161 (1981); W. Sarlet and F. Cantrijn, Phys. Lett. A **88**, 383 (1982).

[5]F. G. Gascon, F. B. Ramos, and E. Aguirre-Dabon, J. Math. Phys. **23**, 2281 (1982).

[6]P. G. L. Leach, Phys. Lett. A **84**, 161 (1981); P. G. L. Leach, Phys. Lett. A **98**, 89 (1983).

[7]J. R. Ray and J. L. Reid, J. Math. Phys. **20**, 2054 (1979); J. R. Ray, Phys. Lett. A **78**, 4 (1980); J. L. Reid and J. R. Ray, J. Math. Phys. **21**, 1583 (1980); J. L. Reid and J. R. Ray, J. Phys. A **15**, 2751 (1982).

[8]H. R. Lewis and K. R. Symon, Phys. Fluids **27**, 192 (1984).

[9]Unpublished report of a generalization of the results in W. Sarlet, J. Phys. A **11**, 843 (1978).

[10]H. R. Lewis, J. Math. Phys. **25**, 1139 (1984); B. Grammaticos and B. Dorizzi, "'Two-dimensional time-dependent Hamiltonian systems with an exact invariant," preprint, Department of Applied Mathematics, National Center for Telecommunications Studies, 38-40 ave. du Général Leclerc, 92131 Issy les Moulineaure, France, 1982.

[11]G. H. Katzin and J. Levine, J. Math. Phys. **24**, 1761 (1983).

[12]G. H. Katzin and J. Levine, J. Math. Phys. **23**, 552 (1982).

[13]D. M. Fradkin, Am. J. Phys. **33**, 207 (1965).

[14]N. J. Günther and P. G. L. Leach, J. Math. Phys. **18**, 572 (1977).

[15]W. Sarlet and F. Cantrijn, SIAM Rev. **23**, 467 (1981).

[16]J. A. Kobussen, Helv. Phys. Acta **53**, 183 (1980).

[17]P. G. L. Leach, J. Math. Phys. **21**, 32 (1980).

[18]C. J. Eliezer and A. Gray, SIAM J. Appl. Math. **30**, 463 (1976).

[19]J. M. Jauch and E. L. Hill, Phys. Rev. **57**, 641 (1940).

[20]c.f. C. D. Collinson, Bull. Inst. Math. Appl. **13**, 377 (1973).

[21]P. G. L. Leach, J. Math. Phys. **18**, 1902 (1977); P. G. L. Leach, SIAM J. Appl. Math. **34**, 496 (1978); P. G. L. Leach, J. Phys. A **16**, 3261 (1983).

[22]P. Camiz, A. Gerardi, C. Marchioro, E. Presutti, and E. Scacciatelli, J. Math. Phys. **12**, 2040 (1971); D. C. Khandekar and S. V. Lawande, J. Math. Phys. **16**, 384 (1975); D. R. Truax, J. Math. Phys. **22**, 1959, (1981); D. R. Truax, J. Math. Phys. **23**, 43 (1982); J. R. Ray, Phys. Rev. A **26**, 729 (1982).

[23]K. B. Wolf, *Integral Transforms in Science and Engineering* (Plenum, New York, 1979), Chap. 9.

[24]P. G. L. Leach J. Math. Phys. **25**, 2974 (1984).

[25]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965).

Use of alternative hyperspherical coordinates for three-body systems

H. Klar[a)]
Institute for Theoretical Physics, University of Innsbruck, A-6020 Innsbruck, Austria

(Received 17 July 1984; accepted for publication 16 November 1984)

Relationships between alternative sets of hyperspherical coordinate systems for the treatment of three-body systems are developed. Transformations of hyperspherical harmonics for S states under a change of intrinsic angles are derived, and applied to harmonic expansions of potential energy surfaces.

I. INTRODUCTION

Hyperspherical coordinates are a powerful instrument for accurate treatments of nonrelativistic three-body systems. These coordinates prove particularly convenient if collective modes of motion rather than independent particle motions are dominant. During the last few years they have been applied successfully to problems of atomic,[1–6] chemical,[7–9] and nuclear[10–12] physics.

Except for minor modifications two different sets of hyperspherical coordinates are used in the literature. One set parametrizes Jacobi coordinates,[13–15] the other set—sometimes called "democratic coordinates"—treats all three particles on an equal basis.[16–19] From the practical point of view both sets have advantages as well as disadvantages. Jacobi hyperspherical coordinates are usually more suitable for the formulation of asymptotic boundary conditions. Treatments within the framework of democratic coordinates often show simple harmonic potential expansions. In many situations it appears desirable to use hyperspherical coordinates in a more flexible way, i.e., to use alternative sets of hyperspherical coordinates within the same problem.

The purpose of this paper is to present new relationships useful for applications between different sets of hyperspherical coordinates. Section II introduces sets of Jacobi hyperspherical coordinates and identifies intrinsic angles as parametrization of suitably defined quaternions. Section III develops the connection between Jacobi and democratic hyperspherical coordinates. Transformation properties of scalar hyperspherical harmonics are presented in Sec. IV and applied to harmonic expansions of potentials in Sec. V.

II. JACOBI HYPERSPHERICAL COORDINATES

We consider three mass points described by position vectors $\mathbf{r}_i$ in their center of mass system

$$\sum_{i=1}^{3} m_i \mathbf{r}_i = 0, \tag{1}$$

and introduce suitably scaled Jacobi coordinates in the usual way. With the reduced masses

$$(\mu_{jk})^{-1} = (m_j)^{-1} + (m_k)^{-1} \tag{2}$$

and

$$(\mu_{i(jk)})^{-1} = (m_i)^{-1} + (m_j + m_k)^{-1}, \tag{3}$$

we define two relative vectors given by

$$\mathbf{r}^{(i)} = \sqrt{\mu_{jk}}(\mathbf{r}_j - \mathbf{r}_k),$$
$$\boldsymbol{\rho}^{(i)} = \sqrt{\mu_{i(jk)}}(\mathbf{r}_i - (m_j\mathbf{r}_j + m_k\mathbf{r}_k)/(m_j + m_k)), \tag{4}$$

$i \neq j \neq k = 1,2,3$. Hyperspherical coordinates are now introduced, putting

$$\mathbf{r}^{(i)} = \sqrt{\mu} r \sin \alpha^{(i)} \hat{r}^{(i)},$$
$$\boldsymbol{\rho}^{(i)} = \sqrt{\mu} r \cos \alpha^{(i)} \hat{\rho}^{(i)}, \tag{5}$$

where μ is an arbitrary mass.

The angle $\alpha^{(i)}$ measures the relative length of the two Jacobi vectors, $0 \leqslant \alpha^{(i)} \leqslant \pi/2$. The unit vectors $\hat{r}^{(i)}$ and $\hat{\rho}^{(i)}$ are parametrized by spherical polar angles,

$$\hat{r}^{(i)} = \hat{r}^{(i)}(\vartheta_r^{(i)}, \varphi_r^{(i)}),$$
$$\hat{\rho}^{(i)} = \hat{\rho}^{(i)}(\vartheta_\rho^{(i)}, \varphi_\rho^{(i)}). \tag{6}$$

The hyperradius $r \geqslant 0$ in Eq. (5) is given by

$$\mu r^2 = \sum_{i=1}^{3} m_i \mathbf{r}_i^2. \tag{7}$$

The six hyperspherical coordinates $(r, \Omega^{(i)})$ with $\Omega^{(i)} = (\alpha^{(i)}, \vartheta_r^{(i)}, \varphi_r^{(i)}, \vartheta_\rho^{(i)}, \varphi_\rho^{(i)})$ replace the two Jacobi vectors. The five angles $\Omega^{(i)}$ depend on the choice of Jacobi coordinates whereas the hyperradius r is invariant.

Consider now a cyclic permutation C of the particles

$$C:(i,j,k) \rightarrow (j,k,i). \tag{8}$$

Two Jacobi vectors $\mathbf{r}^{(j)}$ and $\boldsymbol{\rho}^{(j)}$ obtained from Eq. (4) by cyclic permutation are connected with the Jacobi vectors $\mathbf{r}^{(i)}$ and $\boldsymbol{\rho}^{(i)}$ by a 6×6 matrix,

$$\begin{pmatrix} \mathbf{r}^{(i)} \\ \boldsymbol{\rho}^{(i)} \end{pmatrix} = \begin{pmatrix} \mathbb{1} \cos \beta_{ij} & \mathbb{1} \sin \beta_{ij} \\ -\mathbb{1} \sin \beta_{ij} & \mathbb{1} \cos \beta_{ij} \end{pmatrix} \begin{pmatrix} \mathbf{r}^{(j)} \\ \boldsymbol{\rho}^{(j)} \end{pmatrix}. \tag{9}$$

The angle β_{ij} normalized to

$$\pi/2 < \beta_{ij} < \pi \tag{10}$$

is given by

$$\tan \beta_{ij} = -(m_k M / m_i m_j)^{1/2}, \tag{11}$$

where M is total mass,

$$M = m_1 + m_2 + m_3. \tag{12}$$

The six coordinates necessary to describe the three-body system may always be decomposed into three external and three intrinsic coordinates. The external coordinates describe the position of the rigid three-body system in the lab

[a)] Permanent address: Fachbereich Physik, University of Kaiserslautern, D-6750 Kaiserslautern, Federal Republic of Germany.

 0022-2488/85/071621-05$02.50

frame, whereas the intrinsic coordinates describe all possible modes of particle motion in the body-fixed frame. A rotational invariant interaction depends only on intrinsic coordinates.

Often one wants to represent graphically a potential energy surface using hyperspherical coordinates. Intrinsic coordinates within the above-introduced Jacobi hyperspherical coordinates are $r,\alpha^{(i)},\Theta^{(i)}$, $i = 1,2,3$, with

$$\cos \Theta^{(i)} = \hat{r}^{(i)} \cdot \hat{\rho}^{(i)}. \tag{13}$$

In order to see how a potential energy surface mapping depends on the choice of hyperspherical angles we investigate the transformation of the intrinsic angles $(\alpha^{(i)},\Theta^{(i)})$ under a cyclic particle permutation. Straightforward calculation leads to the transformation law

$$\begin{aligned} \cos 2\alpha^{(i)} &= \cos 2\beta_{ij} \cos 2\alpha^{(j)} \\ &\quad - \sin 2\beta_{ij} \sin 2\alpha^{(j)} \cos \Theta^{(j)}, \\ \sin 2\alpha^{(i)} \cos \Theta^{(i)} &= \sin 2\beta_{ij} \cos 2\alpha^{(j)} \\ &\quad + \cos \beta_{ij} \sin 2\alpha^{(j)} \cos \Theta^{(j)}, \\ \sin 2\alpha^{(i)} \sin \Theta^{(i)} &= \sin 2\alpha^{(j)} \sin \Theta^{(j)}. \end{aligned} \tag{14}$$

These equations may be regarded as parametrization of a rotation in a four-dimensional space $\mathbb{R}_4$. This rotational symmetry is established, for instance, identifying Eqs. (14) as multiplication of quaternions.[20] To this end we introduce three sets of quaternions

$$Q^{(i)} = (Q_0^{(i)},\mathbf{Q}^{(i)}), \quad i = 1,2,3, \tag{15a}$$

with scalar parts

$$Q_0^{(i)} = \cos 2\alpha^{(i)} \tag{15b}$$

and vector parts

$$\mathbf{Q}^{(i)} = \hat{e}^{(i)} \sin 2\alpha^{(i)}. \tag{15c}$$

We parametrize the unit vector $\hat{e}^{(i)}$ by spherical polar angles

$$\hat{e}^{(i)} = (\sin \Theta^{(i)} \cos \Phi^{(i)},\sin \Theta^{(i)} \sin \Phi^{(i)},\cos \Theta^{(i)}). \tag{16}$$

The angle $\Phi^{(i)}$ does not appear in Eq. (14) and may be regarded for the moment as an arbitrary constant.

We describe a cyclic permutation by the quaternion

$$P^{(ij)} = (P_0^{(ij)},\mathbf{P}^{(ij)}), \tag{17a}$$

with

$$P_0^{(ij)} = \cos 2\beta_{ij}, \tag{17b}$$

$$\mathbf{P}^{(ij)} = \hat{e}^{(0)} \sin 2\beta_{ij}, \tag{17c}$$

and the unit vector $\hat{e}^{(0)}$ given by

$$\hat{e}^{(0)} = (0,0,1). \tag{18}$$

It is now straightforward to confirm that Eqs. (14) supplemented by

$$\Phi^{(i)} = \Phi^{(j)} + \beta^{(ij)} \tag{19}$$

read in terms of the above-introduced quaternions

$$P^{(ij)}Q^{(j)} = Q^{(i)}, \tag{20}$$

where the product of two quaternions $P = (P_0,\mathbf{P})$ and $Q = (Q_0,\mathbf{Q})$ is defined by

$$PQ = (P_0Q_0 - \mathbf{P}\mathbf{Q},\mathbf{P}\times\mathbf{Q} + P_0\mathbf{Q} + Q_0\mathbf{P}).$$

Quaternions with unit length,

$$Q_0^{(i)2} + \mathbf{Q}^{(i)2} = P_0^{(ij)2} + \mathbf{P}^{(ij)2} = 1 \tag{21}$$

are isomorphic to special unitary matrices $U^{(i)}\in\mathrm{SU}(2)$(see Ref. 20). In our situation the mapping

$$\begin{aligned} Q^{(i)} &\leftrightarrow U^{(i)}, \\ P^{(ij)} &\leftrightarrow U^{(ij)}, \end{aligned} \tag{22}$$

may be written in the forms

$$U^{(i)} = \cos 2\alpha^{(i)}\sigma_0 - i \sin 2\alpha^{(i)}(\hat{e}^{(i)} \cdot \boldsymbol{\sigma}) \tag{23}$$

and

$$U^{(ij)} = \cos 2\beta_{ij}\sigma_0 - i \sin 2\beta_{ij}(\hat{e}^{(0)} \cdot \boldsymbol{\sigma}). \tag{24}$$

The four matrices σ_μ $(\mu = 0,1,2,3)$ form a basis for the vector space of SU(2) matrices,

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$
$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The isomorphism between quaternions and SU(2) matrices implies that Eq. (20) reads in terms of the matrices given by Eq. (23) and Eq. (24)

$$U^{(ij)}U^{(j)} = U^{(i)}. \tag{25}$$

III. DEMOCRATIC HYPERSPHERICAL COORDINATES

Jacobi coordinates for a three-body system have the property that the particles are not treated equivalently. Hyperspherical coordinates which treat all particles on an equal basis were first considered by Smith[21] and Dragt.[16] For arbitrary masses we parametrize the c.m.-position vectors by[22]

$$\mathbf{r}_i = (\mu(m_j + m_k)/m_iM)^{1/2} rM(\alpha,\beta,\gamma) \times \begin{pmatrix} \cos \psi \cos(\varphi/2 + \gamma_i) \\ \sin \psi \sin(\varphi/2 + \gamma_i) \\ 0 \end{pmatrix}, \quad i = 1,2,3 \tag{26}$$

In Eq. (26) $M(\alpha,\beta,\gamma)$ is an orthogonal 3×3 matrix depending on three Euler angles α,β,γ, which transforms from the lab-fixed frame to a body-fixed frame. The hyperradius r is again given [Eq. (7)], and M is the total mass [Eq. (12)]. The intrinsic angles ψ and φ are independent of any particle labeling.

The constants γ_j in Eq. (26) depend on a decomposition of the three-body system into "one body plus a two-body subsystem" as in the case of Jacobi coordinates. If we label this one body by the index $i(i = 1,2,3)$ the set of constants γ_j $(j = 1,2,3)$ reads explicitly

$$\gamma_j^{(i)} = \beta_{ij} \sum_k \epsilon_{ijk}. \tag{27}$$

The constants β_{ij} are given by Eq. (11). Since Eq. (26) as well as all following equations containing these constants hold in any decomposition we drop the upper index for simplicity, $\gamma_j^{(i)} \equiv \gamma_j$.

The connection between the Jacobi hyperspherical intrinsic angles $\alpha^{(i)}$ and $\Theta^{(i)}$ and the democratic angles ψ and φ is found by calculating the quantities $\mathbf{r}^{(i)2} - \boldsymbol{\rho}^{(i)2},\mathbf{r}^{(i)} \cdot \boldsymbol{\rho}^{(i)}$, and $(\mathbf{r}^{(i)}\times \boldsymbol{\rho}^{(i)})_3$ in both systems. The result reads

$$\cos 2\alpha^{(i)} = \cos 2\psi \cos(\varphi + 2\gamma_i),$$
$$\sin 2\alpha^{(i)} \cos \Theta^{(i)} = -\epsilon_{ijk} \cos 2\psi \sin(\varphi + 2\gamma_i), \quad (28)$$
$$\sin 2\alpha^{(i)} \sin \Theta^{(i)} = \sin 2\psi.$$

A cyclic permutation of the three particles in the democratic frame changes only the numerical values of the constants γ_i and leaves the angles ψ and φ unchanged.

IV. SCALAR HYPERSPHERICAL HARMONICS

We consider now rotations in the six-dimensional c.m.-position space of the three-body system. We denote Cartesian vector components of this space by x_μ and canonically conjugate momentum components by p_μ,

$$[x_\mu, p_\nu] = i\delta_{\mu\nu}, \quad \mu,\nu = 1,...,6. \quad (29)$$

Infinitesimal rotations are generated by

$$\Lambda_{\mu\nu} = x_\mu p_\nu - x_\nu p_\mu. \quad (30)$$

The quadratic Casimir operator, often called the "grand angular momentum," reads

$$\Lambda^2 = \frac{1}{2} \sum_{\mu\nu} \Lambda^2_{\mu\nu}. \quad (31)$$

The eigenvalue equation

$$\Lambda^2 \mathscr{Y}_{\lambda g}(\Omega) = \lambda(\lambda + 4) \mathscr{Y}_{\lambda g}(\Omega), \quad (32)$$

with $\lambda = 0,1,2,...$ defines hyperspherical harmonics depending on a set of five angles Ω. The index g stands in general for a set of quantum numbers which labels degenerate λ states.

Scalar hyperspherical harmonics satisfy the additional requirement

$$\mathbf{L}^2 \mathscr{Y}_{\lambda g}(\Omega) = 0, \quad (33)$$

where

$$\mathbf{L} = \sum_{i=1}^{3} \mathbf{r}_i \times \mathbf{p}_i \quad (34)$$

is the total orbital angular momentum of the three bodies.

These harmonics are particularly important; a harmonic expansion of a rotational invariant potential energy surface for instance contains only these scalar harmonics. They depend only on two intrinsic angles. In terms of Jacobi hyperspherical angles these harmonics normalized to unity read[13,14,23]

$$\mathscr{Y}_{\lambda l_i}(\alpha^{(i)},\Theta^{(i)}) = N_{\lambda l_i} C^{(l_i+1)}_{\lambda/2 - l_i}(\cos 2\alpha^{(i)}) \times (\sin 2\alpha^{(i)})^{l_i} Y_{l_i 0}(\Theta^{(i)},\Phi^{(i)}), \quad (35)$$

with $l_i = 0,1,2...$ and $\lambda = 2l_i, 2l_i + 2, 2l_i + 4,...$.

The normalization constant is given by

$$N_{\lambda l_i} = \frac{1}{\pi} 2^{l_i + 1/2} l_i! \sqrt{\frac{(\lambda + 2)(\lambda/2 - l_i)!}{(\lambda/2 + l_i + 1)!}}. \quad (36)$$

The symbol $C^m_n(x)$ stands for a Gegenbauer polynomial, and $Y_{lm}(\Theta,\Phi)$ is a spherical harmonic independent of Φ for $m = 0$.

Our discussion in Sec. II has identified the angles $\alpha^{(i)},\Theta^{(i)}$, and $\Phi^{(i)}$ as parametrization of quaternions $Q^{(i)}$ with unit length. One expects therefore that scalar hyperspherical harmonics in six dimensions given by Eq. (35) are equal to R_4 harmonics. This may be seen, for instance, by comparing the explicit expressions for Λ^2 in six dimensions, disregarding derivatives with respect to external angles, with the expression for Λ^2 in four dimensions, disregarding derivatives with respect to $\Phi^{(i)}$. We circumvent here this tedious calculation and compare directly the scalar hyperspherical harmonics with the well-known R_4 harmonics.[20] The, latter read in terms of quaternionic components [see Eq. (15a)], is

$$\mathscr{Y}^{(4)}_{jl_i m_i}(Q^{(i)}) = H_{jl_i}(Q^{(i)}_0, |\mathbf{Q}^{(i)}|) \mathscr{Y}^{(3)}_{l_i m_i}(\mathbf{Q}^{(i)}), \quad (37)$$

where

$$\mathscr{Y}^{(3)}_{l_i m_i}(\mathbf{Q}^{(i)}) = |\mathbf{Q}^{(i)}|^{l_i} Y_{l_i m_i}(\hat{Q}^{(i)}) \quad (38)$$

is a harmonic function in three dimensions, $|\mathbf{Q}^{(i)}| = (\mathbf{Q}^{(i)} \cdot \mathbf{Q}^{(i)})^{1/2}$, $\hat{Q}^{(i)} = \mathbf{Q}^{(i)}/|\mathbf{Q}^{(i)}|$, and H_{jl_i} is the rotational invariant. Using Eqs. (15b), (15c), and (16) these functions read up to a constant factor

$$H_{jl_i}(\cos 2\alpha^{(i)}, \sin 2\alpha^{(i)}) \propto C^{(l_i+1)}_{2j - l_i}(\cos 2\alpha^{(i)}). \quad (39)$$

We thus find the expected result

$$\mathscr{Y}_{\lambda l_i}(\alpha^{(i)},\Theta^{(i)}) \propto \mathscr{Y}^{(4)}_{(\lambda/4) l_i 0}(Q^{(i)}). \quad (40)$$

Next we consider the relationship between R_4 harmonics and rotation matrices in three dimensions,[24,25]

$$H_{jl}(Q_0, |\mathbf{Q}|) \mathscr{Y}^{(3)}_{lm}(\mathbf{Q}) = \sum_{v,v'} (-)^{j+v} \sqrt{\frac{2j+1}{2\pi^2}} \langle jv'jv | lm \rangle D^j_{-v',v}(Q_0,\mathbf{Q}), \quad (41)$$

where $\langle \cdots | \cdots \rangle$ stands for a Clebsch–Gordan coefficient. In this relation we put $j = \lambda/4$, $l = l_i$, and $m = 0$ and express the right-hand side (rhs) in terms of the angles $\alpha^{(i)}$ and $\Theta^{(i)}$. On the left-hand side (lhs) with $v' = -v$ we express $Q^{(i)}_0$ and $\mathbf{Q}^{(i)}$ by the angles ψ and φ [see Eq. (28)]. The rotation matrix then gets the form

$$D^{(\lambda/4)}_{(v/4)(v/4)}(Q^{(i)}_0, \mathbf{Q}^{(i)}) = e^{i\epsilon_{ijk} v \gamma_i} D^{(\lambda/4)}_{(v/4)(v/4)}(-\epsilon_{ijk}\varphi, \psi\psi, -\epsilon_{ijk}\varphi), \quad (42)$$

where the last D symbol is a rotation matrix parametrized by Euler angles. This function coincides with the scalar hyperspherical harmonics parametrized by democratic angles[4,5,16–19]

$$Y_{\lambda v}(\psi,\varphi) = \sqrt{(\lambda + 2)/2\pi^3} D^{(\lambda/4)}_{(v/4)(v/4)}(-\varphi, \psi, -\varphi). \quad (43)$$

Equation (41) represents therefore a simple relation between the two sets of scalar hyperspherical harmonics given by Eqs. (35) and (43),

$$Y_{\lambda l_i}(\alpha^{(i)},\Theta^{(i)}) = i^{-l_i} \sum_v (-)^{(\lambda + v)/4} e^{i\epsilon_{ijk} v \gamma_i} \times \left\langle \frac{\lambda}{4} - \frac{v}{4} \frac{\lambda}{4} \frac{v}{4} \middle| l_i 0 \right\rangle Y_{\lambda v}(\psi, \epsilon_{ijk}\varphi). \quad (44a)$$

A relation of this structure was first communicated by Aquilanti *et al.*,[26] inspecting hydrogenic wave functions in the momentum space. Suitably standardized simultaneous eigenfunctions of the energy ($E = -1/n^2 < 0$) and of the Lenz vector $\mathbf{A}$, and those of the energy and of the orbital angular momentum $\mathbf{l}$, are connected by

$$\phi_{nlm}(\mathbf{p}) = \sum_q \left\langle \frac{n-1}{2} \frac{m+q}{2} \frac{n-1}{2} \frac{m-q}{2} \middle| lm \right\rangle \phi_{nqm}(\mathbf{p}),$$

where q is the eigenvalue of A_3. Except for a factor proportional to

$$\{1+(np)^2\}^{-2},$$

the functions $\phi_{nqm}(\mathbf{p})$ are equal to rotation matrices parametrized by Euler–Rodrigues parameters on a sphere $S_3 \subset R_4$. For the special quantum numbers $n=\lambda/2+1$, $q=\nu/2$, and $m=0$ these rotation matrices have been identified in Ref. 26 with scalar hyperspherical harmonics [Eq. (43)] whereas the R_4 harmonics $\phi_{nlm}(\mathbf{p})$ have been identified with scalar hyperspherical harmonics parametrized by Jacobi hyperspherical angles [see Eq. (35)]. The $O(4)$ symmetry of the H atom suggests then a four-dimensional rotation connecting the two particular sets of intrinsic hyperspherical angles.

Our present treatment circumvents the consideration of the H atom and shows directly that a hyperspherical frame transformation in the six-dimensional three-particle space is always accompanied with a four-dimensional rotation for the intrinsic angles. The connection between R_4 harmonics and rotation matrices used here is well known in angular momentum theory.

With the help of the orthogonality for Clebsch–Gordan coefficients relation (44a) is easily converted

$$Y_{\lambda\nu}(\psi,\epsilon_{ijk}\varphi)=(-)^{(\lambda+\nu)/4}e^{-i\epsilon_{ijk}\nu\gamma_i}\times\sum_{l_i} i^{l_i}\left\langle\frac{\lambda}{4}-\frac{\nu}{4}\,\frac{\lambda}{4}\,\frac{\nu}{4}\middle| l_i 0\right\rangle Y_{\lambda l_i}(\alpha^{(i)},\Theta^{(i)}). \tag{44b}$$

Finally, we use Eqs. (44a) and (44b) to derive a relation between two sets of hyperspherical harmonics parametrized by different Jacobi hyperspherical angles. This result reads

$$Y_{\lambda l_i}(\alpha^{(i)},\Theta^{(i)})=\sum_{l_j} T_{l_i l_j} Y_{\lambda l_j}(\alpha^{(j)},\Theta^{(j)}), \tag{45}$$

with

$$T_{l_i l_j}=i^{l_i-l_j}\sum_\nu e^{i\epsilon_{ijk}\nu(\gamma_i-\gamma_j)}\times\left\langle\frac{\lambda}{4}-\frac{\nu}{4}\,\frac{\lambda}{4}\,\frac{\nu}{4}\middle| l_i 0\right\rangle\left\langle\frac{\lambda}{4}-\frac{\nu}{4}\,\frac{\lambda}{4}\,\frac{\nu}{4}\middle| l_j 0\right\rangle. \tag{46}$$

Transformation (46) has also been derived by Smorodinskii *et al.*,[27] directly using Eq. (9).

V. HARMONIC EXPANSION OF POTENTIALS

For the purpose of illustration we apply our results to the expansion of a central field potential into a series of scalar hyperspherical harmonics. To this end we consider a potential between particles labeled by indices i and j, $V(|\mathbf{r}_i-\mathbf{r}_j|)$. In the Jacobi frame labeled by the indices $k\neq i\neq j$ only harmonics with $l_k=0$ occur in the expansion because the potential is independent of the angle $\Theta^{(k)}$. In consequence of Eq. (35) the expansion reads therefore

$$V(|\mathbf{r}_i-\mathbf{r}_j|)=\sum_{\lambda=0,2,\ldots} v_\lambda(r)C^{(1)}_{\lambda/2}(\cos 2\alpha^{(k)}). \tag{47}$$

The coefficients $v_\lambda(r)$ are obtained using the orthogonality of Gegenbauer polynomials.

The transformation of the rhs of Eq. (47) to another Jacobi frame labeled by the index i is performed with help of the addition theorem for Gegenbauer polynomials,[23] which may be cast in our situation into the form

$$C^{(1)}_{\lambda/2}(\cos 2\alpha^{(k)})=\frac{2\pi^3}{\lambda+2}\sum_{l_i=0}^{\lambda/2} Y_{\lambda l_i}(-\beta_{ki},0)Y_{\lambda l_i}(\alpha^{(i)},\Theta^{(i)}). \tag{48}$$

Here we have used the relation

$$\cos 2\alpha^{(k)}=\cos 2\beta_{ki}\cos 2\alpha^{(i)}-\sin 2\beta_{ki}\sin 2\alpha^{(i)}\cos\Theta^{(i)}$$

[see Eq. (14)]. The harmonics on the rhs of Eq. (48) are defined by Eq. (35).

Tranformation of Eq. (48) to democratic angles ψ and φ is provided by Eq. (44a). Using Eq. (28) and the orthogonality for Clebsch–Gordan coefficients we find

$$C^{(1)}_{\lambda/2}(\cos 2\psi\cos(\varphi+2\gamma_i))=\frac{2\pi^3}{\lambda+2}\sum_\nu Y_{\lambda\nu}(0,-2\gamma_i)^*\mathscr{Y}_{\lambda\nu}(\psi,\varphi), \tag{49}$$

where the harmonics on the rhs of Eq. (49) are given by Eq. (43). Expressing these harmonics explicitly in terms of reduced rotation matrices and exponentials we obtain in angular momentum notation (i.e., $j=0, 1/2, 1,\ldots$ and $m=-j, -j+1,\ldots,j$) the interesting relation

$$C^{(1)}_{2j}(\cos 2\psi\cos\Phi)=\sum_{m=-j}^{j} e^{-2im\Phi}d^{\,j}_{mm}(\cos 4\psi), \tag{50}$$

which seems to be unknown.

The development of this section leads to surprisingly simple results in the case of Coulomb potentials

$$V(|\mathbf{r}_i-\mathbf{r}_j|)=(|\mathbf{r}_i-\mathbf{r}_j|)^{-1}. \tag{51}$$

The expansion coefficients $v_\lambda(r)$ read in this case[28]

$$v_\lambda(r)=\frac{8}{\pi r}\sqrt{\frac{\mu_{ij}}{\mu}}\frac{\lambda+2}{(\lambda+1)(\lambda+3)}. \tag{52}$$

The harmonic expansion of the potential may, for example, be used for the computation of matrix elements between hyperspherical harmonics. It is evident that such matrix elements can always be expressed in terms of $O(6)$ Wigner coefficients. For scalar hyperspherical harmonics however a grand simplification occurs. Using Jacobi hyperspherical angles the integration of three R_4 harmonics leads to $O(4)$ Wigner coefficients. This technique[29] has been used recently in the treatment of the molecular ion H_2^+. Using democratic angles on the other hand, the integration of three $O(3)$-rotation matrices leads to $O(3)$ Wigner coefficients.[4,5,30]

ACKNOWLEDGMENT

I thank Dr. Peter Zoller for a critical reading of the manuscript.

[1] V. Fock, K. Nor. Vidensk. Selsk. Forh. **31**, 138 (1958).
[2] J. H. Macek, J. Phys. B **1**, 831 (1968).
[3] C. D. Lin, Phys. Rev. A **10**, 1986 (1974); **12**, 493 (1975); **14**, 30 (1977).
[4] H. Klar and M. Klar, Phys. Rev. A **17**, 1007 (1978); J. Phys. B **13**, 1057 (1980).
[5] E. Pelikan and H. Klar, Z. Phys. A **310**, 153 (1983).
[6] U. Fano, Rep. Prog. Phys. **46**, 97 (1983).

[7]A. Kupperman, J. A. Kaye, and J. P. Deyer, Chem. Phys. Lett. **74**, 257 (1980).
[8]J. Röhmelt, Chem. Phys. Lett. **74**, 263 (1980).
[9]B. R. Johnson, J. Chem. Phys. **73**, 5051 (1980); **79**, 1906 (1983); **79**, 1916 (1983).
[10]Yu. A. Simonov and A. M. Badalyan, Sov. J. Nucl. Phys. **5**, 60 (1967).
[11]J. L. Ballot and M. Fabre de la Ripelle, Ann. Phys. (NY) **127**, 62 (1980).
[12]S. P. Lee, "The Three-Nucleon Problem," thesis, London Imperial College, 1969 (unpublished).
[13]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw–Hill, New York, 1953), p. 1730 ff.
[14]D. L. Knirk, J. Chem. Phys. **60**, 66 (1974); **60**, 760 (1974); Phys. Rev. Lett. **32**, 651 (1974).
[15]Yu. F. Smirnov and K. V. Shitikova, Sov. J. Part. Nucl. **8**, 344 (1977).
[16]A. Dragt, J. Math. Phys. **6**, 533 (1965).
[17]R. C. Whitten and F. T. Smith, J. Math. Phys. **9**, 1103 (1968).
[18]W. Zickendraht, Ann. Phys. **25**, 18 (1965).
[19]H. Mayer, J. Phys. A **10**, 1562 (1975).
[20]L. C. Biedenharn and J. D. Louck, *Angular Momentum of Mathematics and Its Application* (Addison–Wesley, Reading, MA, 1981), Vol. 8.
[21]F. T. Smith, J. Math. Phys. **3**, 375 (1962).
[22]H. Klar, Lecture Notes, University of Innsbruck, 1983 (unpublished).
[23]A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw–Hill, New York, 1953), Vol. 2.
[24]See Ref. 6, p. 352.
[25]J. D. Talman, *Special Functions* (Benjamin, New York, 1968), p. 172 ff.
[26]V. Aquilanti, G. Grossi, and A. Laganà, J. Chem. Phys. **76**, 1587 (1982).
[27]Ya. A. Smorodinskii and V. D. Efros, Sov. J. Nucl. Phys. **17**, 107 (1973).
[28]A. Erdélyi, W. Mangnus, F. Oberhettinger, and F. G. Triconi, *Tables of Integral Transforms* (McGraw–Hill, New York, 1954), Vol. 2, p. 280.
[29]E. Pelikan and V. Aquilanti (private communication, 1984).
[30]V. B. Mandelzweig, Phys. Lett. A **80**, 361 (1980).

Generalization of the splitting theorem to arbitrary one- and two-particle operators

Tomislav P. Živković
Rudjer Bošković Institute, P. O. Box 1016, 41001 Zagreb, Yugoslavia

(Received 18 September 1984; accepted for publication 18 January 1985)

The splitting theorem, previously derived for symmetric operators, is generalized to arbitrary one- and two-particle operators. The notion of alternant systems, as defined in various semiempirical theories of alternant hydrocarbons, is accordingly generalized to arbitrary Hamiltonians. A simple test to decide whether or not a given Hamiltonian is alternant is obtained. A configuration interaction space X_n generated by n electrons moving over $2n$ orthonormalized orbitals is considered. Eigenstates $\Psi \in X_n$ of alternant Hamiltonians are contained in complementary spaces X_n^+ and X_n^-, subspaces of the space X_n. Each state $\Psi^{\pm} \in X_n^{\pm}$ is called alternantlike, and it has characteristic properties of eigenstates associated with neutral alternant hydrocarbons, such as uniform charge density distribution, vanishing bond orders between vertices of the same parity, etc. The complete set of linear properties associated with one- and two-particle operators and common to all alternantlike states $\Psi^{\pm} \in X_n^{\pm}$ is obtained.

I. INTRODUCTION

The purpose of this paper is to generalize the so-called splitting theorem[1,2] and its consequences to arbitrary one- and two-particle operators. This theorem applies to the configuration interaction (CI) space X_n generated by n fermions (electrons) moving over $2n$ orthonormalized spin orbitals.[1–3] The theorem was originally derived for one-particle operators[1] in connection with the molecular orbital resonance theory (MORT)[1–4] approach, and then generalized to two-particle symmetrical operators.[2] In brief, the theorem states that the space X_n can be partitioned into two complementary subspaces X_n^+ and X_n^- which are of the same dimension and such that each state $\Psi^+ \in X_n^+$ as well as each state $\Psi^- \in X_n^-$ is "alternantlike" in the sense that it has all the essential properties of π-electron eigenstates of neutral alternant hydrocarbons (AH).

In order to formulate the theorem, the so-called "reduced" operators are defined. These operators are of two kinds, "alternant" and "antialternant." Alternant reduced operators block diagonalize in the space X_n into subspaces X_n^+ and X_n^-, i.e., they have vanishing matrix elements between all the states $\Psi^+ \in X_n^+$ and $\Psi^- \in X_n^-$ contained in different subspaces. On the other hand, antialternant reduced operators have vanishing matrix elements between all the states contained in the same subspace, either in X_n^+ or in X_n^-. Each linear combination of reduced alternant operators is defined as an alternant operator, while each linear combination of reduced antialternant operators is defined as an antialternant operator. Alternant operators defined in this way have alternantlike eigenstates which in turn have all the nice properties (like uniform charge density distribution, vanishing bond orders between vertices of the same parity, etc.) traditionally associated with alternant systems. Reduced operators thus serve as elementary building blocks of alternant and antialternant operators, and in particular they lead to the constructive and efficient definition of alternant systems.

The plan of the exposition in this paper is as follows: We first derive the complete set of reduced operators. This leads to an explicit construction of arbitrary alternant and antialternant operators, and to the generalization of the splitting theorem to arbitrary operators. Using the notion of the so-called "weakly" alternant operators, we then derive the most general definition of alternant systems: Quantum systems described by weakly alternant Hamiltonians, and only such systems, are alternant. This definition appears to be quite natural, since Hermitian weakly alternant operators are shown to have a complete set of alternantlike eigenstates, and also each operator having a complete set of alternantlike eigenstates is shown to be a weakly alternant operator. Finally, all "linear" properties of alternantlike states are derived. Among these properties are uniform charge density distribution, vanishing bond orders between the vertices of the same parity, etc.

The above consequences of the splitting theorem present, among other things, a systematic and qualitatively new approach to the long-standing problem of alternant systems. It has been shown by various authors, that within a range of models π-electron eigenstates of neutral alternant hydrocarbons have uniform charge density distribution over all carbon atoms, and vanishing bond orders between atoms of the same parity.[5–7] Traditionally, these remarkable properties of alternant systems are derived from the so-called pairing theorem within various molecular orbital (MO) approaches. There have been many attempts to construct "alternant" Hamiltonians having eigenstates with such properties.[5–7] The most general *explicit* solution was given by McLachlan.[6] He has shown that eigenstates of the Parieser–Parr–Pople (PPP) Hamiltonian associated with an AH system satisfy the pairing theorem, and hence the above properties follow in the case of neutral AH systems. Koutecký has extended the notion of alternant systems to some more general symmetric Hamiltonians.[7] However, his approach is rather implicit, and it does not permit an easy construction of alternant Hamiltonians, nor does it present any simple test to decide

 0022-2488/85/071626-14$02.50

whether a given Hamiltonian is alternant or not.

In addition to many other results, the splitting theorem as formulated here gives the complete answer to the above problem of the construction and identification of alternant systems. These systems are defined in such a way that no further generalization is possible without seriously altering the properties of the corresponding eigenstates. The set of all alternantlike states is identified, as well as all one- and two-particle linear properties of these states.

In the present paper we are restricted to the one- and two-particle operators and to the CI space X_n. Hence, unless otherwise stated, all the statements such as "each operator," "an arbitrary state," "all linear properties," etc., should be interpreted as "each one- and two-particle operator" (or their linear combination), "an arbitrary state in the CI space X_n," "all one- and two-particle linear properties," etc., respectively. It is possible to extend the conclusions of this paper to more general operators and spaces.[8] However, the present approach is quite general, since each observable is represented as at most a two-particle operator, while rather complex quantum chemical systems can be described within the space X_n.

II. THE SPLITTING THEOREM

Let η_i^+ and η_i be fermion creation and annihilation operators, respectively. Define reduced operators $\hat{I}$, $\hat{R}_{ij}$, and $\hat{R}_{ij,kl}$

$$\begin{aligned}
&\hat{R}_{ij} = \hat{A}_{ij} - \delta_{ij},\\
&\hat{R}_{ij,kl} = \hat{A}_{ij,kl} \quad (i\neq j\neq k\neq l),\\
&\hat{R}_{ik,jk} = 2\hat{A}_{ik,jk} + \hat{A}_{ij} \quad (i\neq j\neq k),\\
&\hat{R}_{ij,ij} = 2\hat{A}_{ij,ij} + \hat{A}_{ii} + \hat{A}_{jj} - 1 \quad (i\neq j),
\end{aligned} \tag{1a}$$

where

$$\begin{aligned}
&\hat{A}_{ij} = \eta_i^+\eta_j + \eta_j^+\eta_i,\\
&\hat{A}_{ij,kl} = \eta_i^+\eta_j^+\eta_k\eta_l + \eta_l^+\eta_k^+\eta_j\eta_i.
\end{aligned} \tag{1b}$$

A unit operator $\hat{I}$ is a reduced operator as well.[2]

The above reduced operators, as well as operators $\hat{A}_{ij}$ and $\hat{A}_{ij,kl}$, are symmetric, Hermitian, and real.[2] In addition, operators $\hat{A}_{ij}$ and $\hat{A}_{ij,kl}$ satisfy symmetry relations

$$\hat{A}_{ij} = \hat{A}_{ji}, \quad \hat{A}_{ij,kl} = -\hat{A}_{ij,lk} = \hat{A}_{kl,ij}, \tag{2a}$$

and hence we define

$$\hat{R}_{ij} = \hat{R}_{ji}, \quad \hat{R}_{ij,kl} = -\hat{R}_{ij,lk} = \hat{R}_{kl,ij} \tag{2b}$$

in order to generalize the definition (1) of reduced operators to arbitrary indices i,j,k, and l. This facilitates mathematical manipulations involving summations over different indices. In addition, (2b) implies

$$\hat{R}_{ii,kl} = \hat{R}_{ij,kk} = 0, \tag{3}$$

i.e., only operators $\hat{R}_{ij,kl}$ with different indices i and j as well as different indices k and l need be considered.

Each symmetric operator can be represented as a linear combination of reduced operators $\hat{I}$, $\hat{R}_{ij}$, and $\hat{R}_{ij,kl}$, and this representation is unique up to the symmetry relations (2b) (i.e., provided operators $\hat{R}_{ij}$ and $\hat{R}_{ji}$ are considered to be one and the same operator, etc.).[2] For the sake of reference, and as suggested by the graphical representation of the above operators, we use the following terminology: operators $\hat{R}_{ii}$ are vertex operators, operators $\hat{R}_{ij} (i\neq j)$ are bond operators, operators $\hat{R}_{ij,ij} = -\hat{R}_{ij,ji} (i\neq j)$ are vertex–vertex operators, operators $\hat{R}_{ik,jk} = -\hat{R}_{ik,kj} (i\neq j\neq k)$ are bond–vertex operators, while operators $\hat{R}_{ij,kl} (i\neq j\neq k\neq l)$ are bond–bond operators. Analogous terminology is used to denote operators $\hat{A}_{ij}$ and $\hat{A}_{ij,kl}$. Accordingly, indices (i), (j), (k), and (l) are referred to as vertices.[3]

Assume now that there are $2n$ (an even number) creation and $2n$ annihilation operators, and partition the set $B = \{i\}$ of $2n$ indices (vertices) $i = 1,...,2n$ into subsets B^0 and B^* containing n vertices each. We refer to the set B^0 as "source" and to the set B^* as "sink."[1,2] Given the partition $B\to\{B^0,B^*\}$, the set of all symmetric reduced operators $\hat{I}$, $\hat{R}_{ij}$, and $\hat{R}_{ij,kl}$ can be partitioned into "alternant" and "antialternant" operators. A reduced operator $\hat{R}_{ij}$ is alternant if vertices (i) and (j) are of the opposite parity, and antialternant otherwise. Similarly, a reduced operator $\hat{R}_{ij,kl}$ is alternant if an even number among four vertices (i), (j), (k), and (l) is source, and antialternant otherwise.[2] A unit operator is an alternant operator as well. Further, each linear combination of reduced alternant operators is defined to be an alternant operator, while each linear combination of reduced antialternant operators is defined to be an antialternant operator.[2] An arbitrary symmetric operator can be now uniquely represented as a linear combination of an alternant and an antialternant operator.[2]

Let now $|0\rangle$ be a vacuum state

$$\eta_i|0\rangle = 0, \quad i = 1,...,2n, \tag{4}$$

and consider the n-particle space X_n spanned by all vectors $|\Delta_\nu\rangle$ of the form

$$|\Delta_\nu\rangle = \eta_{i1}^+\eta_{i2}^+\cdots\eta_{in}^+|0\rangle. \tag{5}$$

Given the partition $B\to\{B^0,B^*\}$, there is a unique splitting of the space X_n into complementary subspaces X_n^+ and X_n^- with special properties: these subspaces are of the same dimension, and in connection with alternant and antialternant operators they satisfy[2] the following theorem.

Theorem 1 (the splitting theorem): (a) Let $\hat{O}_{\rm al}$ be an alternant operator, and let $\Psi^+\in X_n^+$ and $\Psi^-\in X_n^-$. Then

$$\langle\Psi^+|\hat{O}_{\rm al}|\Psi^-\rangle = \langle\Psi^-|\hat{O}_{\rm al}|\Psi^+\rangle = 0. \tag{6a}$$

(b) Let $\hat{O}_{\rm nal}$ be an antialternant operator, and let either $\Psi_1,\Psi_2\in X_n^+$ or $\Psi_1,\Psi_2\in X_n^-$. Then

$$\langle\Psi_1|\hat{O}_{\rm nal}|\Psi_2\rangle = 0. \tag{6b}$$

Each state $\Psi^+\in X_n^+$, as well as each state $\Psi^-\in X_n^-$ has characteristic properties of π-electron eigenstates associated with neutral alternant hydrocarbons (AH), and hence we call these states "alternantlike."[1,2] Normal operators, and in particular Hermitian operators, have the complete set of eigenstates.[9] It follows that each Hermitian alternant operator has the complete set of alternantlike eigenstates.[2] Since observables are represented by Hermitian operators, the consideration of Hermitian operators alone is no restriction on the generality. These, as well as many other results follow in a straightforward manner from the splitting theorem, and they justify the above definition of alternantlike states and alternant and antialternant operators. The actual construction of spaces X_n^+ and X_n^- can be found elsewhere.[1,2]

The splitting theorem was originally derived for the one-particle operators,[1] and then generalized to one- and two-particle symmetric operators.[2] We will now generalize this theorem to arbitrary one- and two-particle operators. Since an arbitrary operator can be represented as a linear combination of a symmetric and an antisymmetric operator, the generalization consists in finding all the antisymmetric operators satisfying the splitting theorem. Consider operators $\hat{B}_{ij}$ and $\hat{B}_{ij,kl}$

$$\begin{aligned} \hat{B}_{ij} &= \sqrt{-1}(\eta_i^+\eta_j - \eta_j^+\eta_i), \\ \hat{B}_{ij,kl} &= \sqrt{-1}(\eta_i^+\eta_j^+\eta_k\eta_l - \eta_l^+\eta_k^+\eta_j\eta_i), \end{aligned} \tag{7}$$

which form a base in a space of all one- and two-particle antisymmetric operators. Operators (7) are Hermitian and antisymmetric, i.e., in the base (5) they satisfy

$$\begin{aligned} \langle\Delta_\nu|\hat{B}_{ij}|\Delta_\mu\rangle^* &= \langle\Delta_\mu|\hat{B}_{ij}|\Delta_\nu\rangle, \\ \langle\Delta_\nu|\hat{B}_{ij,kl}|\Delta_\mu\rangle^* &= \langle\Delta_\mu|\hat{B}_{ij,kl}|\Delta_\nu\rangle, \\ \langle\Delta_\nu|\hat{B}_{ij}|\Delta_\mu\rangle &= -\langle\Delta_\mu|\hat{B}_{ij}|\Delta_\nu\rangle, \\ \langle\Delta_\nu|\hat{B}_{ij,kl}|\Delta_\mu\rangle &= -\langle\Delta_\mu|\hat{B}_{ij,kl}|\Delta_\nu\rangle. \end{aligned} \tag{8a}$$

In addition, they satisfy symmetry relations

$$\hat{B}_{ij} = -\hat{B}_{ji}, \quad \hat{B}_{ij,kl} = -\hat{B}_{ij,lk} = -\hat{B}_{kl,ij} \tag{8b}$$

to be compared with symmetry relations (2a) satisfied by symmetric operators $\hat{A}_{ij}$ and $\hat{A}_{ij,kl}$. In particular, relations (8b) imply

$$\hat{B}_{ii} = 0, \quad \hat{B}_{ii,kl} = \hat{B}_{ij,kk} = \hat{B}_{ij,ij} = 0. \tag{9}$$

In conjuncture with symmetric operators (1b) antisymmetric operators (7) span the space of all one- and two-particle operators. We will now construct antisymmetric reduced operators $\hat{P}$ out of operators $\hat{B}$, similarly as symmetric reduced operators $\hat{R}$ are constructed out of operators $\hat{A}$.

Notice first that one-particle operators $\hat{B}_{ij}$ already satisfy properties (6). Namely, one can show that $\langle\Psi^+|\hat{B}_{ij}|\Psi^-\rangle = \langle\Psi^-|\hat{B}_{ij}|\Psi^+\rangle = 0$ whenever vertices (i) and (j) are of the same parity, and $\langle\Psi_1^+|\hat{B}_{ij}|\Psi_2^+\rangle = \langle\Psi_1^-|\hat{B}_{ij}|\Psi_2^-\rangle = 0$ whenever vertices (i) and (j) are of the opposite parity.[3] Hence operators $\hat{B}_{ij}$ with indices i and j of the same parity have required properties of alternant operators, while operators $\hat{B}_{ij}$ with indices i and j of the opposite parity have required properties of antialternant operators. It follows that each operator $\hat{R}_{ij}\hat{B}_{kl}$ has also definite symmetry properties with respect to complementary spaces X_n^+ and X_n^-, i.e., it satisfies either relation (6a) or relation (6b). For example, if both operators $\hat{R}_{ij}$ and $\hat{B}_{kl}$ satisfy (6a), and if $\Psi^+ \in X_n^+$, then $\langle\Psi^-|\hat{B}_{kl}|\Psi^+\rangle = 0$ for an arbitrary state $\Psi^- \in X_n^-$, and hence the state $\Psi' = \hat{B}_{kl}\Psi^+$ has no component in the space X_n^-, i.e., $\Psi' \in X_n^+$. Similarly $\hat{R}_{ij}\Psi' \in X_n^+$, and hence $\langle\Psi^-|\hat{R}_{ij}\hat{B}_{kl}|\Psi^+\rangle = 0$. In the same way one finds $\langle\Psi^+|\hat{R}_{ij}\hat{B}_{kl}|\Psi^-\rangle = 0$, i.e., $\hat{R}_{ij}\hat{B}_{kl}$ satisfies (6a). Using the anticommutation algebra of creation and annihilation operators η_i^+ and η_i one finds

$$\hat{B}_{ij,kl} = \tfrac{1}{2}[\hat{A}_{ik}\hat{B}_{lj} + \hat{B}_{ki}\hat{A}_{jl}] \quad (i\neq j\neq l\neq k). \tag{10a}$$

It can now be easily shown that operator $\hat{B}_{ij,kl}(i\neq j\neq k\neq l)$ satisfies (6a) if an odd number out of four vertices (i), (j), (k), and (l) is source, and (6b) otherwise. Since further

$$-\hat{B}_{ij}\hat{R}_{kk} = 2\hat{B}_{ik,jk} + \hat{B}_{ij}, \tag{10b}$$

operators $\hat{P}_{ik,jk} = 2\hat{B}_{ik,jk} + \hat{B}_{ij}$ have also definite symmetry properties with respect to complementary spaces X_n^+ and X_n^-, and hence antisymmetric reduced operators can be chosen to be

$$\begin{aligned} \hat{P}_{ij} &= \hat{B}_{ij} \quad (i\neq j), \\ \hat{P}_{ij,kl} &= \hat{B}_{ij,kl} \quad (i\neq j\neq k\neq l), \\ \hat{P}_{ik,jk} &= 2\hat{B}_{ik,jk} + \hat{B}_{ij} \quad (i\neq j\neq k). \end{aligned} \tag{7'}$$

In view of the symmetry relations (8b) one can define

$$\hat{P}_{ij} = -\hat{P}_{ji}, \quad \hat{P}_{ij,kl} = -\hat{P}_{ji,kl} = -\hat{P}_{kl,ij} \tag{8b'}$$

in order to generalize the definition (7′) of antisymmetric reduced operators to arbitrary indices i, j, k, and l. Operators (7′) are Hermitian and antisymmetric. In addition (8b′) implies

$$\hat{P}_{ii} = 0, \quad \hat{P}_{ii,kl} = \hat{P}_{ij,kk} = 0, \quad \hat{P}_{ij,ij} = 0, \tag{9'}$$

and hence reduced operators (9′) need not be considered.

Each reduced operator $\hat{P}_{ij}$ and $\hat{P}_{ij,kl}$ satisfies either relation (6a) or relation (6b). Reduced operators satisfying relation (6a) are defined to be alternant, while reduced operators satisfying relation (6b) are defined to be antialternant. One finds that a reduced operator $\hat{P}_{ij}$ is alternant if vertices (i) and (j) are of the same parity, and antialternant otherwise. Similarly, a reduced operator $\hat{P}_{ij,kl}$ is alternant if it contains an odd number of vertices of the same parity, and antialternant otherwise. This rule is exactly opposite to the rule for reduced symmetric operators $\hat{R}_{ij}$ and $\hat{R}_{ij,kl}$. Hence if $\hat{R}_{ij}$ is alternant (antialternant) then $\hat{P}_{ij}$ is antialternant (alternant), and vice versa. Also, if $\hat{R}_{ij,kl}$ is alternant (antialternant) then $\hat{P}_{ij,kl}$ is antialternant (alternant), and vice versa.

The set of all antisymmetric reduced operators (7′) is complete in the space of antisymmetric operators, i.e., each antisymmetric operator can be represented as their linear combination. In addition, this representation is unique up to the symmetry relations (8b′). The first part of this statement follows from the fact that operators $\hat{B}_{ij,kl}(i\neq j\neq k\neq l)$ and $\hat{B}_{ik,jk}(i\neq j\neq k)$ span the space of all two-particle antisymmetric operators, and they can be represented as linear combinations of reduced operators (7′)

$$\begin{aligned} \hat{B}_{ij,kl} &= \hat{P}_{ij,kl}, \quad \text{bond–bond}, \\ \hat{B}_{ik,jk} &= \tfrac{1}{2}(\hat{P}_{ik,jk} - \hat{P}_{ij}), \quad \text{bond–vertex}, \end{aligned} \tag{11}$$

while operators $\hat{P}_{ij} = \hat{B}_{ij}$ span the space of all one-particle antisymmetric operators. The second part follows from the fact that operators (7) are linearly independent up to the symmetry relations (8b), and hence relations (7′) imply that reduced operators $\hat{P}_{ij}$ and $\hat{P}_{ij,kl}$ are linearly independent up to the symmetry relations (8b′). It follows that reduced operators (1a) and (7′) are complete and linearly independent in the space of all operators. One can now generalize the definition of alternant and antialternant operators. By definition, each linear combination of reduced alternant operators (1a) and/or (7′) is an alternant operator, while each linear combination of reduced antialternant operators (1a) and/or (7') is an antialternant operator. Each alternant operator $\hat{O}_{\rm al}$ is hence of the form

$$\hat{O}_{\rm al} = \alpha + \sum_{i,j}^{-} \alpha_{ij}^{s} \hat{R}_{ij} + \sum_{i,j}^{+} \alpha_{ij}^{a} \hat{P}_{ij} + \sum_{i,j,k,l}^{+} \alpha_{ij,kl}^{s} \hat{R}_{ij,kl} + \sum_{i,j,k,l}^{-} \alpha_{ij,kl}^{a} \hat{P}_{ij,kl}, \tag{12a}$$

while each antialternant operator $\hat{O}_{\rm nal}$ is of the form

$$\hat{O}_{\rm nal} = \sum_{i,j}^{+} \alpha_{ij}^{s} \hat{R}_{ij} + \sum_{i,j}^{-} \alpha_{ij}^{a} \hat{P}_{ij} + \sum_{i,j,k,l}^{-} \alpha_{ij,kl}^{s} \hat{R}_{ij,kl} + \sum_{i,j,k,l}^{+} \alpha_{ij,kl}^{a} \hat{P}_{ij,kl}, \tag{12b}$$

where α, α_{ij}^{s}, α_{ij}^{a}, $\alpha_{ij,kl}^{s}$, and $\alpha_{ij,kl}^{a}$ are arbitrary coefficients, and the following convention concerning different summations is used: double summations Σ_{ij}^{+} and Σ_{ij}^{-} are performed over vertices (i) and (j) of the same and of the opposite parity, respectively. Quadrupole summation Σ_{ijkl}^{+} is performed over all sets of four vertices (i), (j), (k), and (l) such that an even number of these vertices is source, while quadrupole summation Σ_{ijkl}^{-} is performed over all sets of four vertices (i), (j), (k), and (l) such that an odd number of these vertices is source. Without loss of generality one can assume

$$\alpha_{ij}^{s} = \alpha_{ji}^{s}, \quad \alpha_{ij}^{a} = -\alpha_{ji}^{a}, \quad \alpha_{ij,kl}^{s} = -\alpha_{ji,kl}^{s} = \alpha_{kl,ij}^{s}, \quad \alpha_{ij,kl}^{a} = -\alpha_{ji,kl}^{a} = -\alpha_{kl,ij}^{a} \tag{12c}$$

in order to follow symmetry properties (2b) and (8b') of the correpsonding reduced operators. All alternant and antialternant operators (12a) and (12b) satisfy relations (6a) and (6b), respectively. This completes the extension of the splitting theorem to arbitrary one- and two-particle operators.

In the above formulation of the splitting theorem, there is still one question which remains open. We have shown that alternant and antialternant operators satisfy relations (6a) and (6b), respectively. It is, however, not yet clear whether or not alternant and antialternant operators are *all* operators satisyfing these relations. In order to answer this question, let us define "weakly" alternant and "weakly" antialternant operators. By definition, each operator satisfying (6a) is weakly alternant, while each operator satisfying (6b) is weakly antialternant. Obviously, each alternant operator is also weakly alternant, while each antialternant operator is also weakly antialternant. The above question can now be formulated in the following way: Is each weakly alternant operator necessarily an alternant operator, and similarly, is each weakly antialternant operator necessarily an antialternant operator?

Let $\hat{O}$ be a weakly alternant operator. Since the set of all reduced operators is linearly independent and complete, there is a unique decomposition

$$\hat{O} = \hat{O}_{\rm al} + \hat{O}_{\rm nal}, \tag{13}$$

where $\hat{O}_{\rm al}$ is an alternant operator, while $\hat{O}_{\rm nal}$ is an antialternant operator. By definition, operator $\hat{O}$ satisfies (6a). Further, according to the splitting theorem, $\hat{O}_{\rm al}$ also satisfies (6a), and hence $\hat{O}_{\rm nal}$ should satisfy (6a) as well. However, $\hat{O}_{\rm nal}$ is an antialternant operator satisfying (6b), and one thus finds that $\hat{O}_{\rm nal}$ should vanish over the space X_n. This does not yet imply that $\hat{O}_{\rm nal}$ is zero, i.e., that $\hat{O} = \hat{O}_{\rm al}$. Reduced operators $\hat{R}_{ij}$, $\hat{R}_{ij,kl}$, $\hat{I}$, $\hat{P}_{ij}$, and $\hat{P}_{ij,kl}$ are linearly independent provided the whole space X generated by the creation operators η_i^{+} from the vacuum state $|0\rangle$ is considered. This space contains the vacuum state, all the one-particle states, all the two-particle states, etc. The n-particle space X_n is the nontrivial subspace of the space X, and reduced operators are not linearly independent on X_n alone. For example, the number of particle operators $\hat{N} = \Sigma_i \eta_i^{+} \eta_i$ is constant over X_n, i.e.,

$$\hat{N}\Psi = n\Psi, \tag{14}$$

whenever $\Psi \in X_n$. Hence the antialternant operator $\hat{O}_{\rm nal} = \lambda(\hat{N} - n) = \frac{1}{2}\lambda\, \Sigma_i \hat{R}_{ii}$, where λ is an arbitrary constant, vanishes over X_n. This example shows that an antialternant operator vanishing over X_n is not necessarily zero. The problem to find *all* weakly alternant operators thus reduces to the problem to find *all* antialternant operators vanishing over X_n. Similarly, the problem to find *all* weakly antialternant operators reduces to the problem to find *all* alternant operators vanishing over X_n. This problem can be further simplified, since one easily finds that if an operator $\hat{O}$ vanishes over X_n, both its symmetric and its antisymmetric components should vanish over X_n. Hence it is sufficient to find all symmetric and all antisymmetric operators with the above properties. In the Appendix we prove the following theorem.

Theorem 2: (a) The necessary and sufficient condition for a symmetric alternant operator $\hat{O}_{\rm al}^{s}$ to vanish over X_n is that it is of the form

$$\hat{O}_{\rm al}^{s} = \sum_{i} C_i + \sum_{i<j} (C_i + C_j)\hat{R}_{ij,ji} + \sum_{i<j}^{+} C_{ij}^{s} \sum_{k} \hat{R}_{ik,kj}, \tag{15}$$

where C_i and C_{ij}^{s} are arbitrary coefficients.

(b) The necessary and sufficient condition for a symmetric antialternant operator $\hat{O}_{\rm nal}^{s}$ to vanish over X_n is that it is of the form

$$\hat{O}_{\rm nal}^{s} = C^{s} \sum_{i} \hat{R}_{ii} + \sum_{i<j}^{-} C_{ij}^{s} \sum_{k} \hat{R}_{ik,kj}, \tag{16}$$

where C^{s} and C_{ij}^{s} are arbitrary coefficients.[10]

(c) The necessary and sufficient condition for an antisymmetric alternant operator $\hat{O}_{\rm al}^{a}$ to vanish over X_n is that it is of the form

$$\hat{O}_{\rm al}^{a} = \sum_{i<j}^{-} C_{ij}^{a} \sum_{k} \hat{P}_{ik,kj}, \tag{17}$$

where C_{ij}^{a} are arbitrary coefficients.

(d) The necessary and sufficient condition for an antisymmetric antialternant operator $\hat{O}_{\rm nal}^{a}$ to vanish over X_n is that it is of the form

$$\hat{O}_{\rm nal}^{a} = \sum_{i<j}^{+} C_{ij}^{a} \sum_{k} \hat{P}_{ik,kj}, \tag{18}$$

where the C_{ij}^{a} are arbitrary coefficients.

From the above theorem, Corollary 1 now follows.

Corollary 1: (a) Each weakly alternant operator is of the form

$$\hat{O} = \hat{O}_{\rm al} + \hat{O}^{s}_{\rm nal} + \hat{O}^{a}_{\rm nal}, \tag{19}$$

where $\hat{O}_{\rm al}$ is an alternant operator, while $\hat{O}^{s}_{\rm nal}$ and $\hat{O}^{a}_{\rm nal}$ are antialternant operators whose general forms are given by relations (16) and (18), respectively.

(b) Each weakly antialternant operator is of the form

$$\hat{O} = \hat{O}_{\rm nal} + \hat{O}^{s}_{\rm al} + \hat{O}^{a}_{\rm al}, \tag{20}$$

where $\hat{O}_{\rm nal}$ is an antialternant operator, while $\hat{O}^{s}_{\rm al}$ and $\hat{O}^{a}_{\rm al}$ are alternant operators whose general forms are given by relations (15) and (17), respectively.

Corollary 1 answers the above question to find *all* operators satisfying (6a) and *all* operators satisfying (6b). However, what we obtained is in fact much more than is usually needed. Namely, operators $\hat{O}^{s}_{\rm nal}$ and $\hat{O}^{a}_{\rm nal}$ vanish over X_n, and hence weakly alternant operator (19) and alternant operator $\hat{O}_{\rm al}$ have the same eigenvalues and eigenvectors in X_n, i.e., as far as the space X_n is considered, these operators are identical. Hence we have the following theorem.

Theorem 3: (a) Let the operator $\hat{O}$ be weakly alternant, i.e., let it satisfy

$$\langle \Psi^+ | \hat{O} | \Psi^- \rangle = \langle \Psi^- | \hat{O} | \Psi^+ \rangle = 0, \tag{21a}$$

for each $\Psi^+ \in X_n^+$ and $\Psi^- \in X_n^-$. Then there exists an alternant operator $\hat{O}_{\rm al}$ such that

$$\hat{O}\Psi = \hat{O}_{\rm al}\Psi, \tag{21b}$$

whenever $\Psi \in X_n$.

(b) Let the operator $\hat{O}$ be weakly antialternant, i.e., let it satisfy

$$\langle \Psi_1 | \hat{O} | \Psi_2 \rangle = 0, \tag{21c}$$

if either $\Psi_1, \Psi_2 \in X_n^+$ or $\Psi_1, \Psi_2 \in X_n^-$. Then there exists an antialternant operator $\hat{O}_{\rm nal}$ such that

$$\hat{O}\Psi = \hat{O}_{\rm nal}\Psi, \tag{21d}$$

whenever $\Psi \in X_n$.

According to the above theorem, as far as the space X_n is considered, alternant operators are *all* operators satisfying (6a), while antialternant operators are *all* operators satisfying (6b).

III. PARTITION OF AN ARBITRARY OPERATOR ON ITS ALTERNANT AND ANTIALTERNANT COMPONENT

In the next section it will be shown that all alternant systems are described by alternant and weakly alternant Hamiltonians. Further, using relations (12a), (16), (18), and (19), one can explicitly construct all alternant and weakly alternant Hamiltonians. These Hamiltonians are represented as linear combinations of reduced operators. We have thus obtained an explicit and constructive answer to the problem of the formation of alternant systems. The inverse question is equally important: How can one recognize alternant and weakly alternant Hamiltonians, i.e., how can one decide whether or not a given Hamiltonian describes an alternant system? A Hamiltonian is usually not written in the form of the linear combination of reduced operators, and hence one cannot immediately answer this question. The related but not identical question is to find a decomposition of an arbitrary Hamiltonian on its alternant and antialternant components. The explicit form of this decomposition presents a rationale for the intuitive picture where an *arbitrary* system is considered to be an alternant system perturbed by an antialternant perturbation. This picture leads to an effective perturbation expansion with many interesting properties.[8] We will now answer the above questions for the general case of an arbitrary operator, which can be easily specialized to the case of Hermitian operators.

An arbitrary one- and two-particle operator can be written in the form

$$\hat{O} = \lambda + \sum_{i,j} \lambda_{ij} \eta_i^+ \eta_j + \sum_{i<j,k<l} \lambda_{ij,kl} \eta_i^+ \eta_j^+ \eta_l \eta_k, \tag{22a}$$

where λ, λ_{ij}, and $\lambda_{ij,kl}$ are arbitrary coefficients. Without loss of generality one can assume

$$\lambda_{ij,kl} = -\lambda_{ji,kl} = -\lambda_{ij,lk} \tag{22b}$$

as implied by the fermion anticommutation relations. This permits the extension of the above summations to arbitrary indices.

We will also use the notation

$$\lambda_{ij} = \lambda^{s}_{ij} + \lambda^{a}_{ij}, \quad \lambda_{ij,kl} = \lambda^{s}_{ij,kl} + \lambda^{a}_{ij,kl}, \tag{22c}$$

where superscripts (s) and (a) refer to symmetric and to antisymmetric components, respectively, of matrices $\{\lambda_{ij}\}$ and $\{\lambda_{ij,kl}\}$:

$$\begin{aligned} &\lambda^{s}_{ij} = (\lambda_{ij} + \lambda_{ji})/2, \quad \lambda^{a}_{ij} = (\lambda_{ij} - \lambda_{ji})/2, \\ &\lambda^{s}_{ij,kl} = (\lambda_{ij,kl} + \lambda_{kl,ij})/2, \\ &\lambda^{a}_{ij,kl} = (\lambda_{ij,kl} - \lambda_{kl,ij})/2. \end{aligned} \tag{22d}$$

These components satisfy

$$\begin{aligned} &\lambda^{s}_{ij} = \lambda^{s}_{ji}, \quad \lambda^{s}_{ij,kl} = \lambda^{s}_{kl,ij} = -\lambda^{s}_{ji,kl}, \\ &\lambda^{a}_{ij} = -\lambda^{a}_{ji}, \quad \lambda^{a}_{ij,kl} = -\lambda^{a}_{kl,ij} = -\lambda^{a}_{ji,kl}. \end{aligned} \tag{22e}$$

Since further

$$\begin{aligned} &\eta_i^+ \eta_j = (\hat{A}_{ij} - \sqrt{-1}\hat{B}_{ij})/2, \\ &\eta_i^+ \eta_j^+ \eta_k \eta_l = (\hat{A}_{ij,kl} - \sqrt{-1}\hat{B}_{ij,kl})/2, \end{aligned} \tag{23}$$

one obtains

$$\hat{O} = \lambda + \frac{1}{2}\left[\sum_{i,j} \lambda^{s}_{ij} \hat{A}_{ij} + \sum_{i<j,k<l} \lambda^{s}_{ij,kl} \hat{A}_{ij,lk}\right] + \sqrt{-1}\,\frac{1}{2}\left[\sum_{i,j} \lambda^{a}_{ij} \hat{B}_{ji} + \sum_{i<j,k<l} \lambda^{a}_{ij,kl} \hat{B}_{ij,kl}\right]. \tag{24}$$

Note that operators $\hat{A}$ and $\hat{B}$ are Hermitian, and hence the operator $\hat{O}$ is Hermitian whenever coefficients λ, λ^{s}_{ij}, and $\lambda^{s}_{ij,kl}$ are real, while coefficients λ^{a}_{ij} and $\lambda^{a}_{ij,kl}$ are imaginary. Using relations (1a) and (7'), one can express operator $\hat{O}$ as a linear combination of reduced operators

$$
\begin{aligned}
\hat{O} = \lambda + \frac{1}{2}\Big\{ & \sum_i \Big[\lambda^s_{ii} + \frac{1}{4}\sum_j \lambda^s_{ij,ij}\Big] \\
& + \sum_i \Big[\lambda^s_{ii} + \frac{1}{2}\sum_j \lambda^s_{ij,ij}\Big] \hat{R}_{ii} \\
& + \sum_{i\neq j} \Big[\lambda^s_{ij} + \frac{1}{2}\sum_k \lambda^s_{ik,jk}\Big] \hat{R}_{ij} \\
& + \frac{1}{2}\sum_{i<j} \lambda^s_{ij,ij}\, \hat{R}_{ij,ji} \\
& + \frac{1}{2}\sum_k \sum_{i\neq j} \lambda^s_{ik,jk}\, \hat{R}_{ik,kj} + \sum_{i<j,k<l} \lambda^s_{ij,kl}\, \hat{R}_{ij,lk} \Big\} \\
& + \sqrt{-1}\,\frac{1}{2}\Big\{ \sum_{i,j} \Big[\lambda^a_{ij} + \frac{1}{2}\sum_k \lambda^a_{ik,jk}\Big] \hat{P}_{ji} \\
& + \frac{1}{2}\sum_k \sum_{i,j} \lambda^a_{ik,jk}\, \hat{P}_{ik,jk} + \sum_{i<j,k<l} \lambda^a_{ij,kl}\, \hat{P}_{ij,kl} \Big\},
\end{aligned} \tag{25}
$$

where the quadrupole summations Σ_{ijkl} are performed over mutually disjunct vertices (i), (j), (k), and (l), i.e., only over bond–bond operators. Operator $\hat{O}$ can now be written in the form

$$\hat{O} = \hat{O}_{\rm al} + \hat{O}_{\rm nal}, \tag{26}$$

where $\hat{O}_{\rm al}$ is an alternant operator, while $\hat{O}_{\rm nal}$ is an antialternant operator

$$
\begin{aligned}
\hat{O}_{\rm al} = \lambda + \frac{1}{2}\Big\{ & \sum_i \Big[\lambda^s_{ii} + \frac{1}{4}\sum_j \lambda^s_{ij,ij}\Big] \\
& + \sum_{i,j}^{-} \Big[\lambda^s_{ij} + \frac{1}{2}\sum_k \lambda^s_{ik,jk}\Big] \hat{R}_{ij} \\
& + \frac{1}{2}\sum_{i<j} \lambda^s_{ij,ij}\, \hat{R}_{ij,ji} \\
& + \frac{1}{2}\sum_k \sum_{i\neq j}^{+} \lambda^s_{ik,jk}\, \hat{R}_{ik,kj} + \sum_{i<j,k<l}^{+} \lambda^s_{ij,kl}\, \hat{R}_{ij,kl} \Big\} \\
& + \sqrt{-1}\,\frac{1}{2}\Big\{ \sum_{i,j}^{+} \Big[\lambda^a_{ij} + \frac{1}{2}\sum_k \lambda^a_{ik,jk}\Big] \hat{P}_{ji} \\
& + \frac{1}{2}\sum_k \sum_{i,j}^{-} \lambda^a_{ik,jk}\, \hat{P}_{ik,jk} \\
& + \sum_{i<j,k<l}^{-} \lambda^a_{ij,kl}\, \hat{P}_{ij,kl} \Big\},
\end{aligned} \tag{27a}
$$

$$
\begin{aligned}
\hat{O}_{\rm nal} = \frac{1}{2}\Big\{ & \sum_i \Big[\lambda^s_{ii} + \frac{1}{2}\sum_j \lambda^s_{ij,ij}\Big] \hat{R}_{ii} \\
& + \sum_{i\neq j}^{+} \Big[\lambda^s_{ij} + \frac{1}{2}\sum_k \lambda^s_{ik,jk}\Big] \hat{R}_{ij} \\
& + \frac{1}{2}\sum_k \sum_{i,j}^{-} \lambda^s_{ik,jk}\, \hat{R}_{ik,kj} + \sum_{i<j,k<l}^{-} \lambda^s_{ij,kl}\, \hat{R}_{ij,lk} \Big\} \\
& + \sqrt{-1}\,\frac{1}{2}\Big\{ \sum_{i,j}^{-} \Big[\lambda^a_{ij} + \frac{1}{2}\sum_k \lambda^a_{ik,jk}\Big] \hat{P}_{ji} \\
& + \frac{1}{2}\sum_k \sum_{i,j}^{+} \lambda^a_{ik,jk}\, \hat{P}_{ik,jk} + \sum_{i<j,k<l}^{+} \lambda^a_{ij,kl}\, \hat{P}_{ij,kl} \Big\}.
\end{aligned} \tag{27b}
$$

In Eqs. (27) quadrupole summations Σ^{+}_{ijkl} and Σ^{-}_{ijkl} are performed over mutually distinct vertices (i), (j), (k), and (l), i.e., over bond–bond operators.

Operator $\hat{O}_{\rm al}$, is a sum of a symmetric and an antisymmetric component, and similarly operator $\hat{O}_{\rm nal}$ is a sum of a symmetric and an antisymmetric component. Symmetric components of $\hat{O}_{\rm al}$ and $\hat{O}_{\rm nal}$ involve only symmetric parts of matrices $\{\lambda_{ij}\}$ and $\{\lambda_{ij,kl}\}$, while antisymmetric components of $\hat{O}_{\rm al}$ and $\hat{O}_{\rm nal}$ involve only antisymmetric parts of matrices $\{\lambda_{ij}\}$ and $\{\lambda_{ij,kl}\}$.

Relations (27) present a simple and efficient algorithm to obtain alternant and antialternant components of an arbitrary operator (22). In addition, since reduced operators are linearly independent, these relations automatically yield necessary and sufficient conditions for an operator (22) to be (weakly) alternant or (weakly) antialternant. For example, according to Corollary 1, a necessary and sufficient condition for an operator (22) to be weakly alternant is $\hat{O}_{\rm nal} = \hat{O}^s_{\rm nal} + \hat{O}^a_{\rm nal}$, which is equivalent to the following set of conditions[10]:

(a) $\lambda^s_{ii} + \frac{1}{2}\sum_j \lambda^s_{ij,ij} = 2C^s$;

(b) $\lambda^s_{ij} + \frac{1}{2}\sum_k \lambda^s_{ik,jk} = 0$,

i and j are of the same parity, $i\neq j$;

(c) $\lambda^a_{ij} + \frac{1}{2}\sum_k \lambda^a_{ik,jk} = 0$,

i and j are of the opposite parity;

(d) $\lambda^s_{ik,jk} = 2C^s_{ij}$,

i and j are of the opposite parity, $i\neq j\neq k$; (28)

(e) $\lambda^a_{ik,jk} = 2C^a_{ji}$,

i and j are of the same parity, $i\neq j\neq k$;

(f) $\lambda^s_{ij,kl} = 0$,

odd number of source vertices, $i\neq j\neq k\neq l$;

(g) $\lambda^a_{ij,kl} = 0$,

even number of source vertices, $i\neq j\neq k\neq l$;

where C^s, $C^s_{ij} = C^s_{ji}$, and $C^a_{ij} = -C^a_{ji}$ are arbitrary coefficients. The above conditions do not contain coefficients λ^s_{ij} (i and j are of the opposite parity), λ^a_{ij} (i and j are of the same parity), $\lambda^s_{ij,kl}$ (even number of source vertices, $i\neq j\neq k\neq l$), and $\lambda^a_{ij,kl}$ (odd number of source vertices, $i\neq j\neq k\neq l$). Hence all these coefficients are completely arbitrary. In a similar way one obtains necessary and sufficient conditions for the operator (22) to be alternant. These conditions coincide with conditions (28) except for the conditions (28a), (28d), and (28e) which become

(a) $\lambda^s_{ii} + \frac{1}{2}\sum_j \lambda^s_{ij,ij} = 0$;

(d) $\lambda^s_{ik,jk} = 0$,

i and j are of the opposite parity; (28′)

(e) $\lambda^a_{ik,jk} = 0$,

i and j are of the same parity;

that is, coefficients C^s, C^s_{ij}, and C^a_{ij} in (28) are set to zero.

Analogously one obtains explicit conditions for the operator $\hat{O}$ to be weakly antialternant and to be antialternant, respectively.

IV. ALTERNANT SYSTEMS

Since alternantlike states have characteristic properties of π-electron eigenstates associated with neutral alternant hydrocarbons, it is natural to consider each Hamiltonian having the complete set of alternantlike eigenstates as describing some alternant system. We will take this intuitive picture as a definition of alternant systems: A quantum system is "alternant" if it is described by some Hamiltonian having the complete set of alternantlike eigenstates. The completeness is here defined with respect to the space X_n. In view of this definition, we would like to find all Hamiltonians describing alternant systems.

Consider the Hamiltonian operator $\hat{H}$ written in the form (22). Since H is Hermitian, λ and symmetric coefficients λ_{ij}^{s} and $\lambda_{ij,kl}^{s}$ are real, while antisymmetric coefficients λ_{ij}^{a} and $\lambda_{ij,kl}^{a}$ are imaginary. Further, each Hermitian operator has the complete set of eigenstates in the space X_n. From the splitting theorem and property (6a) of weakly alternant operators, it now follows that each Hermitian weakly alternant operator $\hat{H} = \hat{H}_{\rm al} + \hat{O}_{\rm nal}^{s} + \hat{O}_{\rm nal}^{a}$ has a complete set of alternantlike eigenstates, i.e., according to the above definition it describes some alternant system. In addition, weakly alternant operators are most general operators with such a property. Namely, if the operator $\hat{O}$ is not weakly alternant, then there are some states $\Psi^{-} \in X_n^{-}$ and $\Psi^{+} \in X_n^{+}$ such that $\langle \Psi^{-} | \hat{O} | \Psi^{+} \rangle \neq 0$. If now the operator $\hat{O}$ has at the same time the complete set of alternantlike eigenstates Ψ_i

$$\hat{O}\Psi_i = \epsilon_i \Psi_i, \quad \text{where} \quad \Psi_i \in X_n^{+} \quad \text{or} \quad \Psi_i \in X_n^{-},$$

then the state Ψ^{+} can be expanded in terms of eigenstates $\Psi_i \in X_n^{+}$, and using this expansion one obtains $\langle \Psi^{-} | \hat{O} | \Psi^{+} \rangle = 0$, contrary to the above expression. This proves Corollary 2.

Corollary 2: The necessary and sufficient condition for a Hermitian operator $\hat{H}$ to have the complete set of alternantlike eigenstates is that it is weakly alternant.

In other words, as far as Hermitian operators are considered, weakly alternant operators and *only* weakly alternant operators have the complete set of alternantlike eigenstates. Corollary 2 implies the following corollary.

Corollary 3: Each quantum system which is described by some weakly alternant Hamiltonian is an alternant system. Also, only such systems which are described by weakly alternant Hamiltonians are alternant systems.

If the Hamiltonian H is written in the form (22), then necessary and sufficient conditions for this Hamiltonian to describe an alternant system are explicitly given by relations (28). The interpretation of these conditions is straightforward: conditions (28a) imply that the effective potential energy of an electron is constant over all vertices (spin-orbitals) (i). Conditions (28b) imply that the real component of the effective resonance integral between vertices of the same parity should be zero. This is the generalization of the usual assumption that in an alternant system the resonance integral vanishes between atoms of the same parity. Conditions (28c) imply that the imaginary component of the effective resonance integral vanishes between vertices of the opposite parity. Conditions (28d) and (28e) refer to the real and imaginary components of three center integrals. Thus, according to the relations (28d), whenever vertices (i) and (j) are of the same parity and $k \neq i \neq j$, coefficients $\lambda_{ik,jk}^{s}$ (which represent real components of three center integrals) are independent on the vertex (k), i.e., $\lambda_{ik,jk}^{s} = \lambda_{ik',jk'}^{s}$. Similarly, conditions (28f) and (28g) refer to real and imaginary components, respectively, of four center integrals. If the Hamiltonian H is real, then conditions (28c), (28e), and (28g) referring to imaginary components of various integrals are automatically satisfied. Conditions (28a), (28b), (28d), and (28f) referring to symmetric coefficients λ_{ij}^{s} and $\lambda_{ij,kl}^{s}$ were derived already elsewhere.[2,10] However, conditions (28), as obtained in this paper, can be applied to an arbitrary Hamiltonian, and hence they encompass the most general definition of alternant systems. As emphasized in the previous section, these conditions do not contain coefficients λ_{ij}^{s} (i and j are of the opposite parity), λ_{ij}^{a} (i and j are of the same parity), etc. Hence all these coefficients (integrals) can be set arbitrary. This is a very substantial generalization of the notion of alternant systems. From the above discussion it is obvious that no further generalization is possible.

One further point should be emphasized. Though each weakly alternant Hamiltonian describes some alternant system, in order to describe an arbitrary system it is sufficient to consider only alternant Hamiltonians. Namely, from Theorem 3 it follows that each alternant system can be described by some alternant Hamiltonian. According to Theorem 2, even the space of all alternant Hamiltonians is too large, since operators $\hat{O}_{\rm al}^{a}$ and $\hat{O}_{\rm al}^{s}$ vanish over X_n, and hence they can be "extracted" from an arbitrary alternant Hamiltonian without altering eigenstates and eigenvalues.

V. DENSITY MATRICES OF ALTERNANT SYSTEMS

The importance of alternantlike states lies in the fact that these states have characteristic properties of neutral alternant hydrocarbons.[1,2] Hence these states can be associated with alternant systems, as we did in the preceding section. We will now discuss properties of alternantlike states in more detail. In fact, we will find *all* one- and two-particle properties which can be expressed as expectation values of some observables.

According to the splitting theorem, each alternantlike state $\Psi^{\pm} \in X_n^{\pm}$ satisfies

$$\langle \Psi^{\pm} | \hat{O}_{\rm nal} | \Psi^{\pm} \rangle = 0, \tag{29a}$$

where $\hat{O}_{\rm nal}$ is an arbitrary antialternant operator. Since relation (29a) holds for all alternantlike states, it expresses some general property of these states. This property can also be considered to be characteristic, since it is in general not shared by arbitrary states $\Psi \in X_n$ (unless, for example, $\hat{O}_{\rm nal}$ vanishes over X_n, in which case $\hat{O}_{\rm nal} = \hat{O}_{\rm nal}^{s} + \hat{O}_{\rm nal}^{a}$, etc). Each antialternant operator $\hat{O}_{\rm nal}$ is thus associated with some property common to all alternantlike states. Further, according to Theorem 2 each state $\Psi \in X_n$ satisfies

$$\langle \Psi | \hat{O}_{\rm al}^{s} + \hat{O}_{\rm al}^{a} + \hat{O}_{\rm nal}^{s} + \hat{O}_{\rm nal}^{a} | \Psi \rangle = 0. \tag{29b}$$

Operators $\hat{O}_{\rm al}^{s}$, $\hat{O}_{\rm al}^{a}$, $\hat{O}_{\rm nal}^{s}$, and $\hat{O}_{\rm nal}^{a}$ are hence associated with some properties common to all states $\Psi \in X_n$, and in particular to the alternantlike states $\Psi^{\pm} \in X_n^{\pm}$.

One would like to know how general are relations (29), i.e., can one express each property common to all alternantlike states in this form? In the case of linear properties, the answer is affirmative. These properties can be expressed in the form[11]

$$\lambda_0 + \lambda_1 \langle \Psi | \hat{O}_1 | \Psi \rangle + \lambda_2 \langle \Psi | \hat{O}_2 | \Psi \rangle + \cdots = 0, \tag{30a}$$

where $\hat{O}_1$, $\hat{O}_2$,... are linear operators. Relation (30a) can be written in the form $\langle \Psi | \hat{O} | \Psi \rangle = 0$, where $\hat{O} = \lambda_0 + \lambda_1 \hat{O}_1 + \lambda_2 \hat{O}_2 + \cdots$. In particular, if Ψ is an alternantlike state, one has

$$\langle \Psi^{\pm} | \hat{O} | \Psi^{\pm} \rangle = 0. \tag{30b}$$

Hence the problem to find *all* linear properties of alternantlike states is equivalent to the problem to find all operators $\hat{O}$ satisfying (30b) for an arbitrary alternantlike state $\Psi^{\pm}$. One can now show that if some operator $\hat{O}$ satisfies (30b) for each alternantlike state $\Psi^{\pm}$, then it satisfies

$$\langle \Psi_1^{\pm} | \hat{O} | \Psi_2^{\pm} \rangle = 0, \tag{30c}$$

for each pair $\Psi_1^{\pm}$ and $\Psi_2^{\pm}$ of alternantlike states. But this last condition is exactly the definition of weakly antialternant operators, and hence according to Corollary 1 the operator $\hat{O} = \hat{O}_{\text{nal}} + \hat{O}_{\text{al}}^{s} + \hat{O}_{\text{al}}^{a}$ is the most general operator satisfying (30b). Each operator satisfying (30b) is hence a linear combination of operators satisfying (29a) and (29b). This shows that the set of all weakly antialternant operators encompasses *all* linear properties common to alternantlike states. It is, in fact, sufficient to consider some complete set of linearly independent weakly antialternant operators, since otherwise the associated properties depend on each other. Consider first the relation (29a) referring to antialternant operators. Since each antialternant operator can be represented as a linear combination of reduced antialternant operators, condition (29a) imposed on an arbitrary antialternant operator is equivalent to the set of conditions involving all reduced antialternant operators. These last conditions can be conveniently transformed into the set of conditions involving matrix elements of one- and two-particle density matrices.

Define first matrix elements of one- and two-particle density matrices γ and Γ, respectively,[12–14]

$$\gamma_{ij}(\Psi) = \langle \Psi | \eta_i^{+} \eta_j | \Psi \rangle, \tag{31a}$$

$$\Gamma_{ij,kl}(\Psi) = \langle \Psi | \eta_i^{+} \eta_j^{+} \eta_l \eta_k | \Psi \rangle / 2, \tag{31b}$$

where the state Ψ is assumed to be normalized. We further use the notation

$$\gamma_{ij} = \gamma_{ij}^{s} + \gamma_{ij}^{a}, \quad \Gamma_{ij,kl} = \Gamma_{ij,kl}^{s} + \Gamma_{ij,kl}^{a}, \tag{32a}$$

where superscripts (s) and (a) refer to symmetric and antisymmetric components, respectively. We have

$$\begin{aligned}
\gamma_{ij}^{s} &= (\gamma_{ij} + \gamma_{ji})/2 = \langle \Psi | \hat{A}_{ij} | \Psi \rangle / 2, \\
\gamma_{ij}^{a} &= (\gamma_{ij} - \gamma_{ji})/2 = \sqrt{-1} \langle \Psi | \hat{B}_{ji} | \Psi \rangle / 2, \\
\Gamma_{ij,kl}^{s} &= (\Gamma_{ij,kl} + \Gamma_{kl,ij})/2 = \langle \Psi | \hat{A}_{ij,lk} | \Psi \rangle / 4, \\
\Gamma_{ij,kl}^{a} &= (\Gamma_{ij,kl} - \Gamma_{kl,ij})/2 = \sqrt{-1} \langle \Psi | \hat{B}_{ij,kl} | \Psi \rangle / 4.
\end{aligned} \tag{32b}$$

Since matrices γ and Γ are Hermitian,

$$\gamma_{ij}^{*} = \gamma_{ji}, \quad \Gamma_{ij,kl}^{*} = \Gamma_{kl,ij}, \tag{33}$$

symmetric and antisymmetric components coincide with real and imaginary components, respectively. From the relations (1), (7), (29a), and (31) it now follows that

$$\begin{aligned}
&\text{(a)} \quad \gamma_{ii}(\Psi^{\pm}) = \tfrac{1}{2}; \\
&\text{(b)} \quad \gamma_{ij}^{s}(\Psi^{\pm}) = 0, \\
&\qquad i \text{ and } j \text{ are of the same parity}, \quad i \neq j; \\
&\text{(c)} \quad \gamma_{ij}^{a}(\Psi^{\pm}) = 0, \\
&\qquad i \text{ and } j \text{ are of the opposite parity}; \\
&\text{(d)} \quad \Gamma_{ik,jk}^{s}(\Psi^{\pm}) = \gamma_{ij}^{s}(\Psi^{\pm})/4, \\
&\qquad i \text{ and } j \text{ are of the opposite parity}, \ i \neq j \neq k; \\
&\text{(e)} \quad \Gamma_{ik,jk}^{a}(\Psi^{\pm}) = \gamma_{ij}^{a}(\Psi^{\pm})/4, \\
&\qquad i \text{ and } j \text{ are of the same parity}, \quad i \neq j \neq k; \\
&\text{(f)} \quad \Gamma_{ij,kl}^{s}(\Psi^{\pm}) = 0, \\
&\qquad \text{odd number of vertices of the same} \\
&\qquad \text{parity}, \quad i \neq j \neq k \neq l; \\
&\text{(g)} \quad \Gamma_{ij,kl}^{a}(\Psi^{\pm}) = 0, \\
&\qquad \text{even number of vertices of the same} \\
&\qquad \text{parity}, \quad i \neq j \neq k \neq l.
\end{aligned} \tag{34}$$

Relations (34) are satisfied by all alternantlike states $\Psi^{\pm} \in X_n^{\pm}$. They express common properties of these states, and they are characteristic in the sense that they are not satisfied by all states $\Psi \in X_n$.

In quantum mechanical treatment of molecular systems one usually considers atomic orbits as "building blocks," and to each atomic orbital one associates one spin-α and one spin-β spin-orbital.[1,2] Within this model one can define spin-independent one- and two-particle density matrices ρ and P, respectively, as

$$\begin{aligned}
\rho_{ij} &= \gamma_{ij}^{\alpha} + \gamma_{ij}^{\beta}, \\
P_{ij,kl} &= \Gamma_{ij,kl}^{\alpha\alpha} + \Gamma_{ij,kl}^{\alpha\beta} + \Gamma_{ij,kl}^{\beta\alpha} + \Gamma_{ij,kl}^{\beta\beta},
\end{aligned} \tag{35}$$

where

$$\begin{aligned}
&\gamma_{ij}^{\alpha} = \langle \Psi | a_i^{+} a_j | \Psi \rangle, \quad \gamma_{ij}^{\beta} = \langle \Psi | b_i^{+} b_j | \Psi \rangle, \\
&\Gamma_{ij,kl}^{\alpha\alpha} = \langle \Psi | a_i^{+} a_j^{+} a_l a_k | \Psi \rangle / 2, \\
&\Gamma_{ij,kl}^{\alpha\beta} = \langle \Psi | a_i^{+} b_j^{+} b_l a_k | \Psi \rangle / 2, \\
&\Gamma_{ij,kl}^{\beta\alpha} = \langle \Psi | b_i^{+} a_j^{+} a_l b_k | \Psi \rangle / 2, \\
&\Gamma_{ij,kl}^{\beta\beta} = \langle \Psi | b_i^{+} b_j^{+} b_l b_k | \Psi \rangle / 2,
\end{aligned} \tag{36}$$

while a_i^{+} (a_i) and b_i^{+} (b_i) are spin-α creation (annihilation) and spin-β creation (annihilation) operators, respectively. In accord with relations (32) one can now define symmetric and antisymmetric components of density matrices ρ and P as

$$\rho_{ij} = \rho_{ij}^{s} + \rho_{ij}^{a}, \quad P_{ij,kl} = P_{ij,kl}^{s} + P_{ij,kl}^{a}, \tag{37a}$$

where

$$\begin{aligned}
&\rho_{ij}^{s} = (\rho_{ij} + \rho_{ji})/2, \quad \rho_{ij}^{a} = (\rho_{ij} - \rho_{ji})/2, \\
&P_{ij,kl}^{s} = (P_{ij,kl} + P_{kl,ij})/2, \quad P_{ij,kl}^{a} = (P_{ij,kl} - P_{kl,ij})/2.
\end{aligned} \tag{37b}$$

Since the total number of spin-orbitals is $2n$, there are n spin-α orbitals $\chi_i = w_i\alpha$ and n spin-β orbitals $\bar{\chi}_i = w_i\,\beta$, where the w_i are orthonormalized atomic orbitals, while α and β are spin-α and spin-β states, respectively.[1,2] Source and sink vertices can now be defined in such a way that if χ_i is source, then $\bar{\chi}_i$ is sink, and vice versa. In other words, the two spin-orbitals associated with the same atomic orbital w_i are of the opposite parity.[2] By convention, the parity of the atomic orbital w_i is chosen to coincide with the parity of the corresponding spin-α orbital χ_i (see Ref. 2). Within this picture relations (34) imply

$$
\begin{aligned}
&\text{(a)}\quad \rho_{ii}(\Psi^{\pm}) = 1;\\
&\text{(b)}\quad \rho^{s}_{ij}(\Psi^{\pm}) = 0,\\
&\qquad i \text{ and } j \text{ are of the same parity},\quad i \neq j;\\
&\text{(c)}\quad \rho^{a}_{ij}(\Psi^{\pm}) = 0,\\
&\qquad i \text{ and } j \text{ are of the opposite parity};\\
&\text{(d)}\quad P^{s}_{ik,jk}(\Psi^{\pm}) = \rho^{s}_{ij}(\Psi^{\pm})/2,\\
&\qquad i \text{ and } j \text{ are of the opposite parity},\quad i \neq j \neq k;\\
&\text{(e)}\quad P^{a}_{ik,jk}(\Psi^{\pm}) = \rho^{a}_{ij}(\Psi^{\pm})/2, \qquad\qquad (38)\\
&\qquad i \text{ and } j \text{ are of the same parity},\quad i \neq j \neq k;\\
&\text{(f)}\quad P^{s}_{ij,kl}(\Psi^{\pm}) = 0,\\
&\qquad \text{odd number of vertices of the same}\\
&\qquad \text{parity},\quad i \neq j \neq k \neq l;\\
&\text{(g)}\quad P^{a}_{ij,kl}(\Psi^{\pm}) = 0,\\
&\qquad \text{even number of vertices of the same}\\
&\qquad \text{parity},\quad i \neq j \neq k \neq l.
\end{aligned}
$$

In connection with relations (34) and (38) a few points should be emphasized.

(a) Indices i, j, k, and l refer to spin-orbitals in relations (34) and to atomic orbitals (i.e., a pair of spin-orbitals) in relations (38). Relations (38) are a consequence of relations (34), and they are valid within a particular though rather general model. Other similar models can be also formulated,[1] and we give here relations (38) in order to make more transparent the physical content of relations (34). In any case relations (34) are more fundamental, and they contain all the relevant information.

(b) In the treatment of neutral alternant hydrocarbons (AH's), and within the π-electron model where each carbon atom donates one π-electron to the system, the partition on sink and source vertices can be chosen to coincide with the usual partition on starred and unstarred carbon atoms. In this case relations (38a) express the uniform π-electron density distribution over all carbon atoms, while relations (38b) express the vanishing of bond orders between atomic orbitals (and hence atoms) of the same parity.[1,2] These are well-known properties of neutral AH's, and they are traditionally derived from the pairing theorem.[5–7] These properties are common to *all* alternantlike states $\Psi^{\pm} \in X_n^{\pm}$, and not just to particular eigenstates of selected alternant Hamiltonians, as usually considered by other authors.[5–7] Properties (38a) and (38b) are derived from properties (34a) and (34b), respectively, which are more fundamental. Thus (34a) implies uniform spin-α and uniform spin-β density of $\frac{1}{2}$ over all spin-orbitals, while (34b) implies vanishing bond orders between all spin-orbitals of the same parity. Note that, e.g., relation (34b) contains more information than relation (38b), since according to (34b) real components of "cross" bond orders $\rho^{\alpha\beta}_{ij} = \langle \Psi^{\pm} | a_i^{+} b_j | \Psi^{\pm} \rangle$ and $\rho^{\beta\alpha}_{ij} = \langle \Psi^{\pm} | b_i^{+} a_j | \Psi^{\pm} \rangle$ vanish between spin-orbitals $\chi_i = w_i\alpha$ and $\bar{\chi}_j = w_j\,\beta$ of the opposite parity (the corresponding atomic orbitals w_i and w_j are then of the same parity), and this information is not contained in the relation (38b). Relation (34b) is thus more fundamental.

(c) Apart from relations (38a) and (38b), all other relations in (38) express some properties which, as far as I know, were not obtained by other authors. All these properties, as well as the set of more basic properties (34), are common to *all* alternantlike states. Also, relations (34) express all the characteristic properties of alternantlike states, as far as one- and two-particle operators are considered.

(d) If the state $\Psi \in X_n$ is real, then antisymmetric components γ^{a} and Γ^{a} of density matrices vanish, and hence relations (34c), (34e), and (34g) are automatically satisfied. Remaining relations (34a), (34b), (34d), and (34f) refer to symmetric (i.e., real) components of density matrices. In the case of symmetric Hamiltonians it is sufficient to consider these relations alone,[2] unless there is some accidental degeneracy which may lead to complex eigenstates. Even in this last case, it is always possible to choose the complete set of real eigenstates, i.e., to consider only symmetric components of density matrices. However, in a more general case when arbitrary complex states are considered, antisymmetric components γ^{a} and Γ^{a} usually do not vanish, and all relations (34) have a full physical content. The same is true for the spin-independent density matrices ρ and P.

(e) If the state $\Psi \in X_n$ is a single-determinental function [as usually assumed in various self-consistent field (SCF) approaches], then a two-particle density matrix Γ factorizes[2,13,14]

$$\Gamma_{ij,kl} = \tfrac{1}{2}(\gamma_{ik}\gamma_{jl} - \gamma_{il}\gamma_{jk}), \tag{39a}$$

and hence

$$
\begin{aligned}
\Gamma^{s}_{ij,kl} &= \tfrac{1}{2}(\gamma^{s}_{ik}\gamma^{s}_{jl} + \gamma^{a}_{ik}\gamma^{a}_{jl} - \gamma^{s}_{il}\gamma^{s}_{jk} - \gamma^{a}_{il}\gamma^{a}_{jk}),\\
\Gamma^{a}_{ij,kl} &= \tfrac{1}{2}(\gamma^{a}_{ik}\gamma^{s}_{jl} + \gamma^{s}_{ik}\gamma^{a}_{jl} - \gamma^{a}_{il}\gamma^{s}_{jk} - \gamma^{s}_{il}\gamma^{a}_{jk}).
\end{aligned} \tag{39b}
$$

In this case relations (34d)–(34g) involving matrix elements of two-particle density matrices can be derived from relations (34a)–(34c) involving matrix elements of one-particle density matrix. For example, if $i \neq j \neq k \neq l$ and if, in addition, there is an odd number of source vertices, then either i and k or j and l are of the same parity, and hence (34b) implies $\gamma^{s}_{ik}\gamma^{s}_{jl} = 0$. Similarly, either vertices (i) and (l) or vertices (j) and (k) are of the same parity, and hence $\gamma^{s}_{il}\gamma^{s}_{jk} = 0$. Analogously (34c) implies $\gamma^{a}_{ik}\gamma^{a}_{jl} = \gamma^{a}_{il}\gamma^{a}_{jk} = 0$ and hence $\Gamma^{s}_{ij,kl} = 0$, in accord with the relation (34f). In a similar way all other relations involving matrix elements of a two-particle density matrix can be derived. Analogously spin-independent density matrices ρ and P can be treated.

In conclusion, properties (34) and (38) are properties of *all* alternantlike states $\Psi^{\pm} \in X_n^{\pm}$, and not only the properties of some eigenstates of alternant Hamiltonians. In some special cases these relations simplify. Thus if the state $\Psi^{\pm}$ is real, relations (34c), (34e), and (34g) are automatically satisfied since antisymmetric components γ^a and Γ^a of density matrices are zero. Similarly, if the state $\Psi^{\pm}$ is a single-determinental function, then the two-particle density matrix Γ factorizes according to (39), and relations (34d)–(34g) can be derived from relations (34a)–(34c), i.e., they are not independent. However, in a general case all relations (34) and (38) are independent. Relations (34) contain all characteristic linear properties of alternantlike states. These relations are ultimately equivalent to the condition (29a) imposed on an arbitrary antialternant operator.

Beside properties (34), which are derived from the relation (29a) and which are characteristic in the sense that they are, in general, not shared by all states $\Psi \in X_n$, there is an analogous set of properties which can be derived from the relation (29b), and which are shared by all states $\Psi \in X_n$. Theorem 2 and the relation (29b) imply well-known relations

$$\sum_i \gamma_{ii} = n, \tag{40a}$$

$$2\sum_k \Gamma^s_{ik,jk} = (n-1)\gamma^s_{ij}, \quad i \neq j, \tag{40b}$$

$$2\sum_k \Gamma^a_{ik,jk} = (n-1)\gamma^a_{ij}, \tag{40c}$$

$$2\sum_j \Gamma_{ij,ij} = (n-1)\gamma_{ii}. \tag{40d}$$

The last three relations can be contracted into

$$2\sum_k \Gamma_{ik,jk} = (n-1)\gamma_{ij}. \tag{40e}$$

Relations (40) are equivalent to the requirement (29b). The physical interpretation of these relations is straightforward. Thus relation (40a) expresses the fact that there are n particles in the system. Concerning the relation (40d), note that $2\Gamma_{ij,ij} = \langle \Psi | \eta_i^+ \eta_j^+ \eta_j \eta_i | \Psi \rangle$ is the so-called pair correlation function[12] which is the probability to find one particle at the vertex (i) simultaneously with another particle at the vertex (j). If one sums this probability over all vertices (j), and since there are $(n-1)$ remaining particles in the system, one should obtain $(n-1)$ multiplied by the probability γ_{ii} to find one particle at the vertex (i). This is exactly the content of the relations (40d). In the similar way relations (40b) and (40c) can be interpreted.

In a sense relations (40) are trivial, since they can be easily derived from the definition (31) of density matrices and from the fact that the operator $\hat{N} = \Sigma \eta_i^+ \eta_i$ satisfies (14). However, we have shown here that the set of conditions (40) is *equivalent* to the requirement that an arbitrary operator $\hat{O} = \hat{O}^s_{\rm al} + \hat{O}^s_{\rm nal} + \hat{O}^a_{\rm al} + \hat{O}^a_{\rm nal}$ vanishes over X_n. This shows that relations (40) express the complete set of linear properties common to all states $\Psi \in X_n$. In conjuncture with relations (34), relations (40) express the complete set of linear properties common to all alternantlike states $\Psi^{\pm} \in X_n^{\pm}$. However, not all conditions (40) are independent on conditions (34). One easily finds that condition (40a) follows from conditions (34a), conditions (40b) involving vertices (i) and (j) of the opposite parity follow from conditions (34d), while conditions (40c) involving vertices (i) and (j) of the same parity follow from conditions (34e). Independent states have only conditions (40b) and (40c) involving vertices (i) and (j) of the same and of the opposite parity, respectively, and conditions (40d). We have thus obtained the complete set of linear properties common to all alternantlike states.

VI. SUMMARY AND CONCLUSION

The main result of this paper is the generalization of the splitting theorem to arbitrary one- and two-particle operators. This theorem is formulated in the configuration interaction space X_n which is generated by n electrons moving over $2n$ orthonormalized spin orbitals. In brief, the theorem states that the space X_n can be partitioned into complementary subspaces X_n^+ and X_n^- such that matrix elements of "alternant" operators vanish between the states contained in different subspaces, while matrix elements of "antialternant" operators vanish between the states contained in the same subspace. As a consequence all the states $\Psi^+ \in X_n^+$, as well as all the states $\Psi^- \in X_n^-$ (jointly called "alternantlike" states), have characteristic properties of π-electron eigenstates associated with neutral alternant hydrocarbons. In addition, each Hermitian alternant operator has the complete set of alternantlike eigenstates, and this is particularly true for alternant Hamiltonians.

In order to formulate the splitting theorem, one has to define the so-called "reduced" operators. These operators are building blocks of alternant and antialternant operators, and they are given by relations (1a) and (7'). Reduced operators $\hat{I}$, $\hat{R}_{ij}$, and $\hat{R}_{ij,kl}$ [Eq. (1a)] are symmetric Hermitian, while reduced operators $\hat{P}_{ij}$ and $\hat{P}_{ij,kl}$ [Eq. (7')] are antisymmetric Hermitian. The notion of alternant and antialternant operators, as well as the splitting $X_n \rightarrow \{X_n^+, X_n^-\}$, depends on the partition $B \rightarrow \{B^0, B^*\}$ of the set B containing all the vertices (i) (or likewise all the corresponding spin orbitals $\eta_i^+|0\rangle$) into subsets B^0 and B^* called "source" and "sink," respectively. Subsets B^0 and B^* are required to contain the same number of elements. Once the partition $B \rightarrow \{B^0, B^*\}$ is fixed, each reduced operator is uniquely defined to be either alternant or antialternant: a unit operator $\hat{I}$, reduced operators $\hat{R}_{ij}$ with vertices (i) and (j) of the opposite parity, reduced operators $\hat{P}_{ij}$ with vertices (i) and (j) of the same parity, reduced operators $\hat{R}_{ij,kl}$ containing an even number of source vertices among four vertices (i), (j), (k), and (l), as well as reduced operators $\hat{P}_{ij,kl}$ containing an odd number of source vertices among four vertices (i), (j), (k), and (l), are alternant. Reduced operators $\hat{R}_{ij}$ with vertices (i) and (j) of the same parity, reduced operators $\hat{P}_{ij}$ with vertices (i) and (j) of the opposite parity, reduced operators $\hat{R}_{ij,kl}$ containing an odd number of source vertices, and reduced operators $\hat{P}_{ij,kl}$ containing an even number of source vertices are antialternant. Each linear combination of reduced alternant operators is now an alternant operator, while each linear combination of reduced antialternant operators is an antialternant operator. Since the set of all reduced operators (1a) and (7') is linearly independent and complete, an arbitrary operator can be uniquely written as a sum of an alternant and an antialternant operator. This decomposition is explicitly given by relations (27). From these relations one easily obtains necessary and sufficient conditions for an arbitrary one- and two-particle operator to be either alternant or antialternant.

According to the splitting theorem, matrix elements of alternant operators vanish between the states contained in different subspaces X_n^+ and X_n^-. However, alternant operators are not all operators with such a property. One hence defines "weakly alternant" and likewise "weakly antialternant" operators. Weakly alternant is each operator with vanishing matrix elements between the states contained in different subspaces, while weakly antialternant is each operator with vanishing matrix elements between the states contained in the same subspace. One finds that each weakly alternant operator is a linear combination of an alternant operator and an antialternant operator vanishing over X_n. Similarly, each weakly antialternant operator is found to be a linear combination of an antialternant operator and an alternant operator vanishing over X_n (Corollary 1). An explicit expression of all operators vanishing over X_n is obtained (Theorem 2). Hence one can easily formulate necessary and sufficient conditions for an arbitrary operator to be either weakly alternant [Eq. (28)], or to be weakly antialternant. In addition, in the case of the Hermitian operators, weakly alternant and only weakly alternant operators have the complete set of alternantlike eigenstates. This property of weakly alternant operators suggests a natural definition of alternant systems: each quantum system which is described by some weakly alternant Hamiltonian is an alternant system. Also, only such systems which are described in this way are alternant. This definition is general in that it can be applied to each Hamiltonian. Using relations (28) one can easily verify whether a given Hamiltonian describes some alternant system or not. In addition, on the space X_n each weakly alternant operator coincides with some alternant operator (Theorem 3). Hence all operators describing alternant systems can be constructed as linear combinations of reduced alternant operators.

In conclusion to the above summary, the splitting theorem leads to a simple and constructive definition of alternant systems. The role of reduced alternant operators is here crucial since they serve as building blocks of arbitrary alternant operators. However, the role of reduced antialternant operators is equally important. Since matrix elements of these operators vanish between all the alternantlike states contained in the same subspace, these operators express some general properties of alternantlike states. It is convenient to express these properties in terms of conditions imposed on matrix elements of one- and two-particle density matrices. One thus obtains the set of all one- and two-particle linear properties characteristic of alternantlike states [Eqs. (34)]. Among other things, Eqs. (34) contain the generalization of the well-known properties of π-electron eigenstates associated with neutral alternant hydrocarbons, like uniform charge density distribution and vanishing bond orders between vertices of the same parity. In conjuncture with properties (40) which are common to all states $\Psi \in X_n$, Eqs. (34) complete the set of one- and two-particle properties common to all alternantlike states. It should be noted that there are some linear properties associated with three-particle, four-particle, etc., operators, and they are expressible in terms of matrix elements of higher-order density matrices.[8] However, relations (34) and (40) are complete as far as at most two-particle operators, or equivalently at most matrix elements of a two-particle density matrix, are considered.

The results presented above are some far-reaching generalizations of the characteristic properties of neutral alternant hydrocarbons, and they are not restricted to π-electron systems alone. The definition of reduced operators is completely general, and any one- and two-particle operator can be split into its alternant and antialternant component. Concerning the space of states, we are restricted here to the CI space X_n. This is not the most general CI space. However, beside π-electron systems, it is flexible enough to describe rather complicated σ-electron systems. In addition, using the notion of dummy vertices, all the results obtained in this paper can be generalized to arbitrary finite-dimensional CI spaces, including some infinite-dimensional spaces.[8] Note that the notion of alternant and antialternant operators, as well as the splitting $X_n \to \{X_n^+, X_n^-\}$ depends on the partition $B \to \{B^0, B^*\}$. This flexibility in the notion of alternant and antialternant operators, and alternantlike states is useful, since the partition $B \to \{B^0, B^*\}$ can be chosen in such a way as to fulfill a particular purpose. For example, one can wish to minimize the antialternant component of the Hamiltonian operator. In this case the alternant component of the Hamiltonian can be treated as an unperturbed Hamiltonian and the antialternant component as a perturbation. This leads to a rather efficient perturbation expansion with many interesting properties.[8] In particular, in the case of π-electron systems of alternant hydrocarbons the partition $B \to \{B^0, B^*\}$ always can be chosen to coincide with the usual partition on starred and unstarred atoms. In this case the antialternant component of the Hamiltonian operator vanishes, etc.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation, Grant No. F6F006F, and the Research Council for Scientific Work of Croatia (SIZ-II).

APPENDIX: PROOF OF THEOREM 2

The proof of this theorem consists of two parts. First we show that conditions (15)–(18) are sufficient for the operators $\hat{O}^s_{\rm al} - \hat{O}^a_{\rm nal}$ to vanish over the space X_n. Next we show that these conditions are also necessary.

(a) We show that the conditions are sufficient. According to the relation (15) the operator $\hat{O}^s_{\rm al}$ is as required a symmetric alternant operator, since it is a linear combination of reduced symmetric alternant operators. Similarly operators $\hat{O}^s_{\rm nal}$, $\hat{O}^a_{\rm al}$, and $\hat{O}^a_{\rm nal}$, as given by relations (16), (17), and (18), are symmetric antialternant, antisymmetric alternant, and antisymmetric antialternant, respectively.

Using the anticommutation algebra of fermion operators η_i^+ and η_i, one finds that relations (15)–(18) are equivalent to

$$\hat{O}^s_{\rm al} = 2\left[\sum_i C_i \hat{R}_{ii} + \sum_{i<j}^{+} C^s_{ij} \hat{A}_{ij}\right](\hat{N} - n), \tag{A1}$$

$$\hat{O}^{s}_{\rm nal} = 2\left[C^{s} + \sum_{i<j}^{-} C^{s}_{ij}\hat{A}_{ij}\right](\hat{N}-n), \tag{A2}$$

$$\hat{O}^{a}_{\rm al} = 2\left[\sum_{i<j}^{-} C^{a}_{ij}\hat{B}_{ij}\right](\hat{N}-n), \tag{A3}$$

$$\hat{O}^{a}_{\rm nal} = 2\left[\sum_{i<j}^{+} C^{a}_{ij}\hat{B}_{ij}\right](\hat{N}-n). \tag{A4}$$

Since $(\hat{N}-n)\Psi = 0$ for each state $\Psi \in X_n$, all the above operators vanish over X_n. This proves that conditions (15)–(18) are sufficient.

(b) In order to prove that conditions (15)–(18) are necessary we proceed in a few steps. An arbitrary operator $\hat{O}$ can be written in the form

$$\hat{O} = \lambda + \sum_{ij}\lambda_{ij}\,\eta_i^{+}\eta_j + \sum_{i<j,k<l}\lambda_{ij,kl}\eta_i^{+}\eta_j^{+}\,\eta_l\eta_k. \tag{A5}$$

In the first step we show that the requirement that this operator vanishes over X_n implies that it contains no bond–bond operator, i.e., that $\lambda_{ij,kl} = 0$ whenever $i \neq j \neq k \neq l$. Hence each operator $\hat{O}$ vanishing over X_n can be written as such a linear combination of reduced operators which does not contain bond–bond operators $\hat{R}_{ij,kl}$ and $\hat{P}_{ij,kl}$.

In the next step we consider operators $\hat{O}^{s}_{\rm al}$, $\hat{O}^{s}_{\rm nal}$, $\hat{O}^{a}_{\rm al}$, and $\hat{O}^{a}_{\rm nal}$ represented as linear combinations of appropriate reduced operators. For example, operator $\hat{O}^{s}_{\rm al}$ is a linear combination of reduced symmetric alternant operators excluding bond–bond operators, etc. The requirements $\hat{O}^{s}_{\rm al}|\Delta_\nu\rangle = 0$, $\hat{O}^{s}_{\rm nal}|\Delta_\nu\rangle = 0$, $\hat{O}^{a}_{\rm al}|\Delta_\nu\rangle = 0$, and $\hat{O}^{a}_{\rm nal}|\Delta_\nu\rangle = 0$, where $|\Delta_\nu\rangle \in X_n$ are some selected vectors in the space X_n, lead to the set of conditions involving coefficients in the corresponding expansions of operators $\hat{O}^{s}_{\rm al}$, $\hat{O}^{s}_{\rm nal}$, $\hat{O}^{a}_{\rm al}$, and $\hat{O}^{a}_{\rm nal}$ in terms of reduced operators. These conditions are sufficient to fix operators $\hat{O}^{s}_{\rm nal}$, $\hat{O}^{a}_{\rm al}$, and $\hat{O}^{a}_{\rm nal}$ in the forms (16), (17), and (18), respectively. Concerning the operator $\hat{O}^{s}_{\rm al}$, one further step is needed since the conditions obtained are rather implicit. This step is accomplished by mathematical induction reasoning from the general n-particle to the general $(n+1)$-particle case.

(b1) Step one: The space X_n is spanned by vectors

$$|\Delta_\nu\rangle = |\nu_1,\nu_2,\ldots,\nu_{2n}\rangle, \tag{A6}$$

where

$$\sum_i \nu_i = n \tag{A7}$$

and $\nu_i = 1$ if the orbital χ_i is occupied, while $\nu_i = 0$ otherwise.[2,15] Vectors (A6) with the condition (A7) are the same as vectors (5) written in a slightly modified notation. One can show that[15]

$$\eta_k^{+}|\ldots,\nu_k,\ldots\rangle = (-1)^{\Sigma_k}(1-\nu_k)|\ldots,\nu_k+1,\ldots\rangle,$$
$$\eta_k|\ldots,\nu_k,\ldots\rangle = (-1)^{\Sigma_k}\nu_k|\ldots,\nu_k-1,\ldots\rangle, \tag{A8}$$

where

$$\Sigma_k = \nu_1+\nu_2+\cdots+\nu_{k-1}. \tag{A9}$$

The vanishing of the operator $\hat{O}$ over the space X_n implies

$$\hat{O}|\Delta_\nu\rangle = 0, \tag{A10}$$

for each vector (A6) satisfying (A7). In particular $\hat{O}|\Delta_0\rangle = 0$, where

$$|\Delta_0\rangle = |1,1,\ldots,1_n,0,0,\ldots\rangle. \tag{A11}$$

Consider the action of the operator $\eta_i^{+}\eta_j^{+}\eta_l\eta_k$ $(i\neq j\neq k\neq l)$ on the vector $|\Delta_0\rangle$. Relations (A8) imply

$$\eta_i^{+}\eta_j^{+}\,\eta_l\eta_k|\Delta_\nu\rangle = \pm|1,\ldots,1,0_l,1,\ldots,1,0_k,1,\ldots,1_n,0,\ldots,1_i,0,\ldots,0,1_j,0,\ldots\rangle = \pm|\Delta_{l,k\to i,j}\rangle \tag{A11$'$}$$

whenever $l,k \leqslant n$ and $i,j > n$. None of the other operators contained in the expression (A5) creates the state $|\Delta_{l,k\to i,j}\rangle$ from the state $|\Delta_0\rangle$. Hence $\hat{O}|\Delta_0\rangle = 0$ implies $\lambda_{ij,kl} = 0$ whenever $i\neq j\neq k\neq l$, $l,k\leqslant n$ and $i,j>n$. The last two conditions are irrelevant, since the state Δ_0 can be replaced by some other state Δ_ν. This leads to the condition $\lambda_{ij,kl} = 0$ whenever $i\neq j\neq k\neq l$.

(b2) Step two: Consider as an example the antisymmetric alternant operator $\hat{O}^{a}_{\rm al}$. This operator is a linear combination of reduced alternant operators $\hat{P}_{ij}$ and $\hat{P}_{ij,kl}$, and since it is required to vanish over X_n, the first step implies that it is of the form

$$\hat{O}^{a}_{\rm al} = \sum_{ij}^{+}\lambda^{a}_{ij}\hat{P}_{ij} + \sum_k\sum_{i<j}^{-}\lambda^{a}_{ik,jk}\hat{P}_{ik,kj}. \tag{A12}$$

From relations (A8) one derives

$$\hat{P}_{ij}|\Delta_0\rangle = \sqrt{-1}\begin{cases}(-1)^{n+i+1}|\Delta_{i\to j}\rangle, & i\leqslant n,\ \ j>n,\\ (-1)^{n+j}|\Delta_{j\to i}\rangle, & j\leqslant n,\ \ i>n,\\ 0, & \text{otherwise,}\end{cases} \tag{A13}$$

$$\hat{P}_{ik,kj}|\Delta_0\rangle = \begin{cases}\hat{P}_{ij}|\Delta_0\rangle, & k\leqslant n,\\ -\hat{P}_{ij}|\Delta_0\rangle, & k>n,\end{cases} \tag{A14}$$

where

$$|\Delta_{i\to j}\rangle = |1,\ldots,1,0_i,1,\ldots,1_n,0,\ldots,0,1_j,0,\ldots\rangle. \tag{A15}$$

In particular, $\hat{P}_{ij}|\Delta_0\rangle = \pm\sqrt{-1}|\Delta_{i\to j}\rangle$ whenever $i\leqslant n$ and $j>n$. Since vertices (i) and (j) are required to be of the same parity, none of the remaining reduced operators contained in (A12) creates the state $|\Delta_{i\to j}\rangle$ from the state $|\Delta_0\rangle$, and hence the condition

$$\hat{O}^{a}_{\rm al}|\Delta_0\rangle = 0 \tag{A16}$$

implies $\lambda^{a}_{ij} = 0$ whenever $i\leqslant n$ and $j>n$. With an appropriate replacement of the vector $|\Delta_0\rangle$ with some other vector $|\Delta_\nu\rangle \in X_n$ one finds $\lambda^{a}_{ij} = 0$. The first term in (A12) is hence zero. Relations (A13), (A14), and (A16) further imply

$$\sum_{k=1}^{n}\lambda^{a}_{ik,jk} - \sum_{k=n+1}^{2n}\lambda^{a}_{ik,jk} = 0. \tag{A17}$$

If the vector $|\Delta_0\rangle$ is replaced with the vector $|\Delta_{1\to n+1}\rangle$, the analogous relation reads

$$\sum_{k=2}^{n+1}\lambda^{a}_{ik,jk} - \sum_{k=n+2}^{2n}\lambda^{a}_{ik,jk} - \lambda^{a}_{i1,j1} = 0. \tag{A17$'$}$$

Relations (A17) imply $\lambda^{a}_{i1,j1} = \lambda^{a}_{i\,n+1,j\,n+1}$ which is easily generalized to $\lambda^{a}_{ik,jk} = C^{a}_{ij}$. Operator $\hat{O}^{a}_{\rm al}$ is hence necessarily of the form (17). In a similar way relations (16) and (18) can be derived. Concerning the symmetric alternant operator $\hat{O}^s_{\rm al}$ the same reasoning leads to the conclusion that $O^s_{\rm al}$ is necessarily of the form

$$\hat{O}^s_{\rm al} = \lambda + \sum_{i<j} \lambda^s_{ij,ij} \hat{R}_{ij,ji} + \sum_{i<j}^{+} C^a_{ij} \sum_k \hat{R}_{ik,kj}, \tag{A18}$$

where the coefficients C^a_{ij} are arbitrary, while coefficients λ and $\lambda^s_{ij,ij} = \alpha_{ij}$ are required to satisfy the condition that the operator $\hat{O}'$,

$$\hat{O}' = \lambda + \sum_{i<j} \alpha_{ij} \hat{R}_{ij,ji}, \tag{A19}$$

vanishes over X_n.

(b3) Step three: It remains to show that the vanishing of the operator (A19) over X_n implies

$$\lambda = \sum_{i=1}^{2n} C_i, \quad \alpha_{ij} = C_i + C_j, \quad i,j \leqslant 2n, \tag{A20}$$

where the C_i are arbitrary. The proof goes by induction. From relation (A8) it follows that

$$\hat{R}_{ij,ji} |\Delta_\nu\rangle = (2\nu_i - 1)(2\nu_j - 1)|\Delta_\nu\rangle, \quad i \neq j. \tag{A21}$$

Hence $\hat{O}'|\Delta_\nu\rangle = 0$ implies

$$\lambda + \sum_{i<j}^{2n} \alpha_{ij} \mu_i \mu_j = 0, \tag{A22}$$

where

$$\mu_i = 2\nu_i - 1 = \pm 1, \quad \sum_{i=1}^{2n} \mu_i = 0. \tag{A22$'$}$$

Relations (A22) should hold for all possible choices of $2n$ integers μ_i satisfying (A22′). If $n = 1$ relations (A22) imply $\lambda - \alpha_{12} = 0$, and hence the general solution can be written in the form (A21), i.e., $\alpha_{12} = C_1 + C_2$ and $\lambda = C_1 + C_2$. The case $n = 2$ leads to the set of conditions

$$\begin{aligned} &\lambda + \alpha_{12} - \alpha_{13} - \alpha_{14} - \alpha_{23} - \alpha_{24} + \alpha_{34} = 0, \\ &\lambda - \alpha_{12} + \alpha_{13} - \alpha_{14} - \alpha_{23} + \alpha_{24} - \alpha_{34} = 0, \\ &\lambda - \alpha_{12} - \alpha_{13} + \alpha_{14} + \alpha_{23} - \alpha_{24} - \alpha_{34} = 0, \end{aligned} \tag{A23}$$

which is equivalent to

$$\lambda = \alpha_{14} + \alpha_{23} = \alpha_{13} + \alpha_{24} = \alpha_{12} + \alpha_{34}. \tag{A23$'$}$$

In (A23′) coefficients α_{14}, α_{23}, α_{13}, and α_{12} can be chosen arbitrary, and relations (A23′) then determine coefficients λ, α_{24}, and α_{34}. However, one easily finds that relations

$$\begin{aligned} &C_1 + C_2 = \alpha_{12}, \quad C_1 + C_3 = \alpha_{13}, \\ &C_1 + C_4 = \alpha_{14}, \quad C_2 + C_3 = \alpha_{23} \end{aligned} \tag{A23$''$}$$

have the solution in unknowns C_1, C_2, C_3, and C_4 for an arbitrary choice of coefficients α_{12}, α_{13}, α_{14}, and α_{23}. Hence one finds that $\lambda = \Sigma C_i$ and $\alpha_{ij} = C_i + C_j$, where the C_i are arbitrary, contains all solutions to (A23). Similarly the case $n = 3$ can be treated to show again that (A20) is the most general solution. However, with the increase of n the algebra becomes quite involved, and in order to show that each solution to (A22) can be written in the form (A20), we proceed by induction.

Assume that for some $n \geqslant 3$ each solution to (A22) can be written in the form (A20), and consider the case $n' = n + 1$. According to (A22) coefficients λ and α_{ij} satisfy

$$\lambda + \sum_{i<j}^{2n+2} \alpha_{ij} \mu_i \mu_j = 0, \tag{A22$''$}$$

for all μ_i ($i = 1,...,2n+2$) such that $\mu_i = \pm 1$ and $\Sigma_i^{2n+2} \mu_i = 0$. In particular, (A22″) should hold for all sets of $(2n + 2)$ integers $\mu_i = \pm 1$ satisfying the following conditions:

$$\mu_{2n+1} = 1, \quad \mu_{2n+2} = -1, \quad \sum_{i=1}^{2n} \mu_i = 0, \tag{A24a}$$

$$\mu_{2n+1} = -1, \quad \mu_{2n+2} = 1, \quad \sum_{i=1}^{2n} \mu_i = 0, \tag{A24b}$$

$$\mu_1 = 1, \quad \mu_2 = -1, \quad \sum_{i=3}^{2n+2} \mu_i = 0, \tag{A24c}$$

$$\mu_1 = -1, \quad \mu_2 = 1, \quad \sum_{i=3}^{2n+2} \mu_i = 0. \tag{A24d}$$

Inserting integers μ_i satisfying conditions (A24a) and (A24b) into (A22″) one obtains

$$\begin{aligned} &\lambda + \sum_{i<j}^{2n} \alpha_{ij} \mu_i \mu_j + \sum_{i=1}^{2n} (\alpha_{i,2n+1} - \alpha_{i,2n+2})\mu_i - \alpha_{2n+1,2n+2} = 0, \\ &\lambda + \sum_{i<j}^{2n} \alpha_{ij} \mu_i \mu_j - \sum_{i=1}^{2n} (\alpha_{i,2n+1} - \alpha_{i,2n+2})\mu_i - \alpha_{2n+1,2n+2} = 0, \end{aligned} \tag{A25}$$

and hence

$$\lambda - \alpha_{2n+1,2n+2} + \sum_{i<j}^{2n} \alpha_{ij} \mu_i \mu_j = 0, \tag{A26a}$$

$$\sum_{i=1}^{2n} (\alpha_{i,2n+1} - \alpha_{i,2n+2})\mu_i = 0. \tag{A26b}$$

According to (A24) integers μ_i in (A26) satisfy the conditions (A22′), and hence (by assumption) the general solution to (A26a) can be written in the form

$$\lambda - \alpha_{2n+1,2n+2} = \sum_{i=1}^{2n} C_i, \quad \alpha_{ij} = C_i + C_j, \quad i \leqslant 2n, \tag{A27}$$

where the C_i are arbitrary. In a similar way relations (A23c) and (A24d) lead to the relations

$$\lambda - \alpha_{12} + \sum_{\substack{3 \\ i<j}}^{2n+2} \alpha_{ij} \mu_i \mu_j = 0, \tag{A28a}$$

$$\sum_{i=3}^{2n+2} (\alpha_{i,1} - \alpha_{i,2})\mu_i = 0. \tag{A28b}$$

Relation (A28a) is again of the type (A22), and hence each solution to this relation can be written in the form

$$\lambda - \alpha_{12} = \sum_{i=3}^{2n+2} C'_i, \quad \alpha_{ij} = C'_i + C'_j, \quad 3 \leqslant i < 2n+2, \tag{A29}$$

where the C'_i are arbitrary. Combining (A27) and (A29) one obtains

$$C_i + C_j = C'_i + C'_j, \quad 3 \leqslant i,j \leqslant 2n, \tag{A30}$$

which implies $C_i = C'_i (3 \leqslant i \leqslant 2n)$. Since C_{2n+1} and C_{2n+2} are not fixed by the relation (A27), we can as well write $C_{2n+2} = C'_{2n+2}$ and $C_{2n+1} = C'_{2n+1}$. Relations (A27) and (A29) now imply that the coefficients α_{ij} are necessarily of the form $\alpha_{ij} = C_i + C_j$ whenever either $1 \leqslant i,j \leqslant 2n$ or $3 \leqslant i,j \leqslant 2n+2$. This includes all coefficients α_{ij} except for coefficients $\alpha_{1,2n+1}$, $\alpha_{1,2n+2}$, $\alpha_{2,2n+1}$, and $\alpha_{2,2n+2}$. After some algebra one finds that these coefficients are also given in the form $C_i + C_j$ and finally one derives $\lambda = \Sigma C_i$, which proves that for each n, a general solution to (A22) is of the form (A20) with C_i arbitrary.

[1]T. P. Živković, Croat. Chem. Acta **57**, 367 (1984).

[2]T. P. Živković, J. Math. Phys. **25**, 2749 (1984).

[3]T. P. Živković, Croat. Chem. Acta **56**, 525 (1983); **56**, 29 (1983).

[4]T. P. Živković, Theor. Chim. Acta **61**, 363 (1982); **62**, 335 (1983); **65**, 151 (1984); Int. J. Quant. Chem. **23**, 679 (1983); Chem. Phys. Lett. **107**, 272 (1984).

[5]C. A. Coulson and G. S. Rushbrooke, Proc. Cambridge Philos. Soc. **36**, 193 (1940); C. A. Coulson and H. C. Longuet-Higgins, Proc. R. Soc. London Ser. A **191**, 39 (1941); J. A. Pople, Trans. Faraday Soc. **49**, 1375 (1953); A. Brickstock and J. A. Pople, Trans. Faraday Soc. **50**, 901 (1954); A. T. Amos, Mol. Phys. **5**, 91 (1962); W. Moffitt, J. Chem. Phys. **26**, 424 (1957); R. B. Potts, J. Chem. Phys. **21**, 758 (1953); H. V. McIntosh, J. Mol. Spectrosc. **8**, 169 (1962); G. G. Hall and A. T. Amos, in *Physical Chemistry Vol. III, Electronic Structure of Atoms and Molecules,* edited by H. Eyring, P. Henderson, and W. Jost (Academic, New York, 1969).

[6]A. D. McLachlan, Mol. Phys. **4**, 49 (1961); **2**, 271 (1959).

[7]J. Koutecký, J. Chem. Phys. **44**, 3702 (1966).

[8]T. P. Živković (to be published).

[9]See, for example, J. Mathews and R. L. Walker, *Mathematical Methods of Physics* (Benjamin, Menlo Park, CA, 1970), Sec. 6.

[10]In Ref. 2 where only symmetric operators were considered, the term $\Sigma_{i<j} C^s_{ij} \Sigma_k \hat{R}_{ik,kj}$ in Eq. (16) was erroneously omitted. Hence the condition (28d) was erroneously derived to read $\lambda^s_{ik,jk} = 0$ instead of $\lambda^s_{ik,jk} = 2C^s_{ij}$ with $C^s_{ij} = C^s_{ji}$ arbitrary.

[11]Relation (30a) is in fact the definition of linear properties, i.e., we define each property expressible in this form to be linear.

[12]In Ref. 2 only real components of matrices γ and Γ were considered. This is sufficient if only symmetric Hamiltonians are treated, since these Hamiltonians have the complete set of real eigenstates, and hence the corresponding density matrices are real. Note that a two-particle density matrix Γ as defined in Ref. 13 differs by a factor of two from a two-particle density matrix as defined in Ref. 14. We normalize here a two-particle density matrix according to Ref. 14.

[13]R. McWeeny and B. T. Sutcliffe, *Methods of Molecular Quantum Mechanics* (Academic, London, 1969), Chap. 4.

[14]R. Zahradník and R. Polák, *Elements of Quantum Chemistry* (Plenum, New York, 1980).

[15]Richard D. Mattuck, *A Guide to Feynman diagrams in the Many-Body Problem* (McGraw–Hill, Maidenhead, England, 1967).

Finite-dimensional representations of the Lie superalgebra sl(1,3) in a Gel'fand–Zetlin basis. I. Typical representations

Tchavdar D. Palev[a)]
Institut für Theoretische Physik der Technischen Universität Clausthal, Clausthal-Zellerfeld, West Germany

(Received 28 November 1984; accepted for publication 11 January 1985)

In a series of two papers all finite-dimensional irreducible representations of the special linear Lie superalgebra sl(1,3) are constructed. Explicit formulas for the generators in an orthonormal Gel'fand–Zetlin basis of the even subalgebra gl(3) are given. This paper develops a background for constructing the representations. Expressions for the transformation properties of the basis under the action of the generators are written down within all typical sl(1,3) modules.

I. INTRODUCTION

In this paper and the one that follows[1] we study all finite-dimensional irreducible representations (IR's) of the Lie superalgebra sl(1,3). For any such representation we write down explicit formulas for the transformations of the corresponding sl(1,3) module (= representation space) V under the action of the algebra. To this end we consider V as a representation space of the even subalgebra $gl(3)\subset sl(1,3)$ and represent it as a direct sum of its irreducible gl(3) submodules V_l,

$$V=\sum_l \oplus V_l. \tag{1.1}$$

As a basis Γ_l within every V_l we choose the Gel'fand–Zetlin basis and define an orthonormal basis

$$\Gamma=\cup_l \Gamma_l=\{f_i|i=1,2,\dots,n\} \tag{1.2}$$

in V, which will be also called a Gel'fand–Zetlin basis (GZ basis). Within every irreducible sl(1,3) module V we compute the matrix elements

$$(f_i,E_k f_j)=(\alpha_k)^*_{ij} \tag{1.3}$$

of the sl(1,3) generators E_k. More precisely, we write down the transformation properties of the GZ basis under the action of the generators,

$$E_k f_i=\sum_{j=1}^n (\alpha_k)_{ji} f_j. \tag{1.4}$$

The algebra sl(1,3) is an example, one of the simplest examples, of a basic Lie superalgebra (LS), i.e., (1) it is simple, (2) its even subalgebra is reductive, and (3) its Killing form is nondegenerate. All basic Lie superalgebras (LS's) are by now classified.[2] First of all, every simple Lie algebra (LA) is a basic LS, since in this case (2) and (3) are consequences of (1). The basic LS's that are not Lie algebras (LA's) resolve into four countable classes [$A(m,n)$, $B(m,n)$, $C(n)$, and $D(m,n)$], two exceptional LS's [$F(4)$ and $G(3)$], and a one-parameter family of exceptional LS's [$D(2,1;\alpha)$]. In these notations, $sl(1,3)=A(0,2)$.

The structure of the basic LS's resembles in many respects the simple LA's. Every such algebra A can be represented as a direct-space sum

$$A=N^-\oplus H\oplus N^+ \tag{1.5}$$

of its Cartan subalgebra H, which is the Cartan subalgebra of the even part, and the subalgebras N^- and N^+ spanned on the negative and the positive root vectors, respectively. The root vectors e_α are in one-to-one correspondence with their roots α, which are elements from the dual space of $H:\alpha\in H^*$. The correspondence $e_\alpha\leftrightarrow\alpha$ is determined from the relation

$$[\![h,e_\alpha]\!]=\alpha(h)e_\alpha,\quad \forall h\in H, \tag{1.6}$$

where $[\![\ ,\ ]\!]$ denotes the product in A. One defines in a natural way such concepts as simple roots, canonical basis, Cartan matrix, etc.[2]

In the last years much attention has been paid to the representation theory of the basic LS's.[3–7] Several important (mainly classification) results have been proved. Nevertheless, the results in this direction are far from being complete. An essential new feature of the representation theory is that the finite-dimensional modules of a basic LS A are not necessarily direct sums of irreducible (i.e., simple) A modules. Apart from the algebra $B(0,n)$, every basic LS has indecomposable (i.e., not fully reducible) representations. Examples of such representations are available [see, for instance, Ref. 7; in Ref. 1 we also consider some indecomposable representations of sl(1,3)]. However, at present there exists no theory to tell us how to construct all indecomposable representations. Even the power of these representations is not known (i.e., how many of them do exist). In contrast to this, the structure of the irreducible representations is much better understood. The biggest success in this direction is the full classification of all finite-dimensional IR's for any basic LS.[3] All such representations are with highest weight. The known-for-the-LA's method of induced representations is generalized in a natural way for LS's and gives all finite-dimensional irreducible representations. In particular, let $\bar V(\Lambda)$ be an A module with a highest weight Λ, induced from the irreducible module $V_0(\Lambda)$ of the even subalgebra $A_0\subset A$ with the additional requirement that $V_0(\Lambda)$ is annihilated by all positive root vectors of A (see Sec. IV). Then, every irreducible finite-dimensional A module $V(\Lambda)$ with highest weight Λ either (1) coincides with $\bar V(\Lambda)$ or (2) is a factor module of it. In case (1), i.e., if $\bar V(\Lambda)=V(\Lambda)$, the representation is called typical. In case (2), $\bar V(\Lambda)$ is indecomposable. The factor module $V(\Lambda)=\bar V(\Lambda)/\bar I(\Lambda)$ with respect to the maximal (nontrivial) invariant subspace $\bar I(\Lambda)$ is simple. The corresponding irreducible representation is said to be nontypical. One can give also an independent [of the particular space $\bar V(\Lambda)$] defin-

[a)] Present address: Institute of Nuclear Research and Nuclear Energy, boul. Lenin 72, 1184 Sofia, Bulgaria.

 0022-2488/85/071640-21$02.50

ition of the typical representations: The irreducible finite-dimensional A module $V(\Lambda)$ is typical if

$$(\Lambda+\rho,\alpha)\neq 0, \quad \text{for all odd roots obeying } (\alpha,\alpha)=0. \tag{1.7}$$

Here $(\ ,\ )$ is the Killing form on A and ρ is the half sum of all even positive roots minus the half sum of all odd positive roots. An equivalent statement is that $V(\Lambda)$ is typical if it splits within any finite-dimensional irreducible representation [$V(\Lambda)$ is always a direct summand]. Because of this property, the structure of all typical representations turns out to be much simpler. In Ref. 3, Kac, who introduced the above classification, has established several other properties of the typical representations. Nevertheless, the representation theory of the IR's and even the theory of the much simpler typical representations is far from being complete. In particular, the important (from a physical point of view) problem to compute the matrix elements (1.3) of the generators has been solved only for certain representations or for some lower-rank LS's.[5,6] A systematic study of all typical and nontypical IR's (and of some other indecomposable representations) is available only for the LS sl(1,2) (see Ref. 7).

In the present paper and in Ref. 1, we make a further small step toward the solution of the general representation problem. We study all finite-dimensional IR's of the LS sl(1,3), decompose them with respect to the even subalgebra gl(3), and write down the representations of the generators (1.4) in the GZ basis (1.2).

In Sec. II we recall the definition of the LS sl(1,3) and list some of its properties, which can be found in Refs. 2 and 3 or are a consequence of the results contained therein (see also Ref. 6). The general representation theory of the even subalgebra gl(3) is well known. Here we shall need, however, some particular topics of it. They are collected in Sec. III, where we also introduce the notation. Using the technique of the induced representations, in Sec. IV we introduce the main representation space $\overline{V}([m]_3)$, which carries all finite-dimensional IR's of sl(1,3). In a basis that arises in a natural way, the induced basis, we write down the representations of the generators (1.4) and select those of them that are simple. This solves the problem to construct all typical representations within some basis. The induced basis is, however, not convenient for computing matrix elements (1.3), since on the level of Sec. IV, the scalar product between the vectors of the induced basis is not known, the metric is given in terms of the GZ basis. Moreover, as it will become clear, the vectors from the induced basis have, in general, nonzero projections in more than one gl(3)-irreducible submodule of $\overline{V}([m]_3)$. Therefore, this basis is inconvenient for specifying the gl(3) submodules. These submodules, on the other hand, will play an important role in the determination of the nontypical IR's.[1] Therefore, in Sec. V we pass to the orthonormed GZ basis, which is by construction gl(3)-reduced. In Sec. V A we decompose the induced sl(1,3) modules into irreducible gl(3) submodules and derive relations between the induced and the GZ basis. In Sec. V B representations of the generators are given in the GZ basis.

II. THE LIE SUPERALGEBRA sl(1,3)

We consider sl(1,3) as a subalgebra of the general linear LS l(1,3). The latter can be defined as the set of all 4×4 matrices

$$\underbrace{\begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03}\\ a_{10} & a_{11} & a_{12} & a_{13}\\ a_{20} & a_{21} & a_{22} & a_{23}\\ a_{30} & a_{31} & a_{32} & a_{33}\end{pmatrix}}_{l(1,3)} = \underbrace{\begin{pmatrix} a_{00} & 0 & 0 & 0\\ 0 & a_{11} & a_{12} & a_{13}\\ 0 & a_{21} & a_{22} & a_{23}\\ 0 & a_{31} & a_{32} & a_{33}\end{pmatrix}}_{l_0(1,3)} \oplus \underbrace{\begin{pmatrix} 0 & a_{01} & a_{02} & a_{03}\\ a_{10} & 0 & 0 & 0\\ a_{20} & 0 & 0 & 0\\ a_{30} & 0 & 0 & 0\end{pmatrix}}_{l_1(1,3)}. \tag{2.1}$$

With respect to the usual operations between matrices, l(1,3) is a linear space, which is a sum of its linearly independent subspaces $l_0(1,3)$ and $l_1(1,3)$. The mapping $l(1,3)\times l(1,3)\to l(1,3)$, which is a linear extension of the relation

$$[a,b]=ab-(-1)^{\alpha\beta}ba,$$
$$\forall\, a\in l_\alpha(1,3),\quad b\in l_\beta(1,3),\quad \alpha,\beta=0,1, \tag{2.2}$$

turns l(1,3) into a Lie superalgebra with an even subalgebra $l_0(1,3)$ and an odd part $l_1(1,3)$.

Let e_{AB}, $A,B=0,1,2,3$ be a matrix with 1 on the A th row and the B th column and zero elsewhere. All these matrices constitute a basis in l(1,3). The Killing form $(\ ,\ )'$ of l(1,3), restricted to its Cartan subalgebra

$$H'=\text{lin.env.}\{e_{AA}\,|A=0,1,2,3\} \tag{2.3}$$

is degenerate and reads

$$(e_{AA},e_{BB})'=4g_{AB}-2(-1)^{a_A+a_B}, \tag{2.4}$$

where

$$a_0=1,\quad a_1=a_2=a_3=0, \tag{2.5}$$

and

$$g_{AB}=(-1)^{a_A}\delta_{AB}. \tag{2.6}$$

The special linear LS sl(1,3) is a subalgebra of l(1,3), consisting of all matrices $(a_{AB})\in l(1,3)$, such that $a_{00}-a_{11}-a_{22}-a_{33}=0$, i.e.,

$$sl(1,3)=\{a\,|a\in l(1,3),\ \operatorname{tr} ga=0\}. \tag{2.7}$$

The grading on sl(1,3) is the one induced from l(1,3). As a Cartan subalgebra H of sl(1,3), we choose

$$H=\text{lin.env.}\{g_{AA}e_{AA}-g_{BB}e_{BB}\,|A,B=0,1,2,3\}. \tag{2.8}$$

Then sl(1,3) can be represented as a direct sum of its subspaces

$$sl(1,3)=N^-\oplus H\oplus N^+, \tag{2.9}$$

where

$$N^+ = \text{lin.env.}\{e_{AB}|A<B=0,1,2,3\}, \tag{2.10}$$

$$N^- = \text{lin.env.}\{e_{AB}|A>B=0,1,2,3\}, \tag{2.11}$$

are subalgebras spanned on all positive root vectors e_{AB}, $A<B=0,1,2,3$ and on all negative root vectors e_{AB}, $A>B=0,1,2,3$, respectively. On H, the form $(\ ,\ )'$ coincides with the Killing form $(\ ,\ )$ of sl(1,3). Since the second term in (2.4) vanishes on H, we neglect it and write

$$(e_{AA},e_{BB}) = 4g_{AB}. \tag{2.12}$$

This equality defines a nondegenerate bilinear form on H', which restricted to H coincides with the Killing form of sl(1,3).

Choose as a basis in H' the matrices

$$\epsilon_A = e_{AA}, \quad A=0,1,2,3. \tag{2.13}$$

As an ordered basis in the conjugate space $\overset{*}{H}{}'$ we take the dual to (2.13) basis of linear functionals

$$\epsilon^0,\ \epsilon^1,\ \epsilon^2,\ \epsilon^3, \tag{2.14}$$

i.e., by definition $\epsilon^A(\epsilon_B)=\delta_{AB}, A,B=0,1,2,3$. Then the relation

$$\overset{*}{f}(h) = (f,h), \quad \forall h\in H', \tag{2.15}$$

defines one-to-one correspondence $f\leftrightarrow\overset{*}{f}$ between H' and $\overset{*}{H}{}'$.

If $f_i\leftrightarrow\overset{*}{f}_i$, $i=1,2$, then

$$(\overset{*}{f}_1,\overset{*}{f}_2) = (f_1,f_2) \tag{2.16}$$

determines a nondegenerate bilinear form of $\overset{*}{H}{}'$:

$$(\epsilon^A,\epsilon^B) = g_{AB}/4. \tag{2.17}$$

The correspondence [following from (1.6)] between the root vectors and their roots is

$$e_{AB}\leftrightarrow\epsilon^A-\epsilon^B, \quad A\neq B=0,1,2,3. \tag{2.18}$$

For the linear functional ρ, defined in (1.7), one obtains

$$\rho = -\tfrac{3}{2}\epsilon^0+\tfrac{3}{2}\epsilon^1+\tfrac{1}{2}\epsilon^2-\tfrac{1}{2}\epsilon^3. \tag{2.19}$$

In particular

$$(\rho,\epsilon^0-\epsilon^k) = (k-1)/4, \quad k=1,2,3. \tag{2.20}$$

The matrices

$$\begin{aligned} &e_0=e_{01}, && e_1=e_{12}, && e_2=e_{23},\\ &f_0=e_{10}, && f_1=e_{21}, && f_2=e_{32},\\ &h_0=e_{00}+e_{11}, && h_1=e_{11}-e_{22}, && h_2=e_{22}-e_{33} \end{aligned} \tag{2.21}$$

generate sl(1,3) and satisfy the relations ($[x,y]=xy-yx$; $\{x,y\}=xy+yx$)

$$\begin{aligned} &\{e_i,f_j\}=\delta_{ij}h_i, \quad [h_i,h_j]=0,\\ &[h_i,e_j]=\alpha_{ij}e_j, \quad [h_i,f_j]=-\alpha_{ij}f_j, \quad i,j=0,1,2. \end{aligned} \tag{2.22}$$

The Cartan matrix $A=(\alpha_{ij})$, carrying all information about the structure of the LS, reads

$$A=\begin{pmatrix} 0 & 1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 2 \end{pmatrix}. \tag{2.23}$$

III. SOME PROPERTIES OF gl(3) AND NOTATION

The significance of the LA gl(3) for the representation theory of sl(1,3) stems from the observation that it is isomorphic to the even part of sl(1,3). The matrices

$$E_{ij} = e_{ij}+\delta_{ij}e_{00}\in \text{sl}(1,3), \quad i,j=1,2,3, \tag{3.1}$$

constitute a basis in the even subalgebra

$$\text{sl}_0(1,3)=\left\{\begin{pmatrix} a_{00} & 0 & 0 & 0\\ 0 & a_{11} & a_{12} & a_{13}\\ 0 & a_{21} & a_{22} & a_{23}\\ 0 & a_{31} & a_{32} & a_{33}\end{pmatrix}\Bigg| \sum_{A=0}^{3} g_{AA}a_{AA}=0\right\} \tag{3.2}$$

of sl(1,3) and satisfy the Weyl commutation relations of gl(3):

$$[E_{ij},E_{kl}] = \delta_{jk}E_{il}-\delta_{li}E_{kj}. \tag{3.3}$$

Therefore, the mapping

$$E_{ij}=e_{ij}+\delta_{ij}e_{00}\leftrightarrow e_{ij}, \quad i,j=1,2,3 \tag{3.4}$$

defines an isomorphism of the even subalgebra $\text{sl}_0(1,3)$ on gl(3).

Turn now to the representations of gl(3). Throughout the paper we use the Gel'fand and Zetlin notation for the finite-dimensional IR's of gl(n) (see Ref. 8), accepting also the abbreviations of Ref. 9. Every such representation is labeled by a lexical n-tuple

$$[m]_n \equiv [m_{1n},m_{2n},\dots,m_{nn}], \tag{3.5}$$

i.e., by n in general complex numbers $m_{1n},m_{2n},\dots,m_{nn}$, which have the same imaginary part, and $m_{in}-m_{i+1,n}$ are non-negative integers:

$$\begin{aligned} &\text{Im}\, m_{1n}=\text{Im}\, m_{2n}=\cdots=\text{Im}\, m_{nn},\\ &m_{in}-m_{i+1,n}\in\mathbb{Z}_+, \quad i=1,2,\dots,n-1. \end{aligned} \tag{3.6}$$

The representations, corresponding to different n-tuples $[m]_n\neq[m']_n$ are, in general, inequivalent. Every irreducible gl(n) module

$$V([m]_n)\equiv V([m_{1n},m_{2n},\dots,m_{nn}]), \tag{3.7}$$

corresponding to a representation $[m]_n$, is a direct sum of nonequivalent irreducible gl($n-1$) submodules,

$$V([m]_n)=\sum \oplus V([m]_{n-1}). \tag{3.8}$$

The sum in (3.8) is over all $[m]_{n-1}$ such that

$$\begin{aligned} &\text{Im}\, m_{i,n-1}=\text{Im}\, m_{1n}, \quad i=1,2,\dots,n-1,\\ &\text{Re}(m_{in}-m_{i,n-1})\in\mathbb{Z}_+, \quad \text{Re}(m_{i,n-1}-m_{i+1,n})\in\mathbb{Z}_+. \end{aligned} \tag{3.9}$$

On the ground of this observation and taking into account that every irreducible gl(1) module $V(m_{11})$ is one dimensional, Gel'fand and Zetlin have introduced an orthonormal basis in $V([m]_n)$, known now as the Gel'fand–Zetlin basis.[8] In the case of gl(3), every basis vector, also called a Gel'fand–Zetlin pattern (GZ pattern)

$$\left|\begin{matrix} m_{13} & m_{23} & m_{33}\\ & m_{12} \quad m_{22} & \\ & m_{11} & \end{matrix}\right\rangle \equiv \left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle \equiv |(m)_3\rangle\,, \tag{3.10}$$

is labeled by six numbers $[m_{13},m_{23},m_{33}]\equiv[m]_3$, $[m_{12},m_{22}]\equiv[m]_2$, and m_{11}, indicating to which gl(3) module $V([m]_3)$, gl(2) submodule $V([m]_2)$, and gl(1) submodule $V(m_{11})$ the vector (3.10) belongs, respectively, i.e., by construction

$$\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle \in V(m_{11}) \subset V([m]_2) \subset V([m]_3). \tag{3.11}$$

The numbers m_{13}, m_{23}, m_{33} are fixed and label $V([m]_3)$. The basis vectors (3.10) within the same $V([m]_3)$ are distinguished by m_{12}, m_{22}, m_{11}, which assume any values consistent with the "betweenness" condition[9]

$$\begin{aligned}&\operatorname{Im} m_{ij} = \text{const},\\ &\operatorname{Re}(m_{i,j+1} - m_{ij}) \in \mathbb{Z}_+, \quad \operatorname{Re}(m_{ij} - m_{i+1,j+1}) \in \mathbb{Z}_+ .\end{aligned} \tag{3.12}$$

By $\Gamma([m]_3)$ we sometimes denote the set of all GZ-basis vectors in $V([m]_3)$, i.e., all GZ patterns (3.10).

The dimension of $V([m]_3)$ is given with the relation[10]

$$\begin{aligned}&\operatorname{Dim} V([m]_3)\\ &\quad = \tfrac{1}{2}(m_{13} - m_{33} + 2)(m_{23} - m_{33} + 1)(m_{13} - m_{23} + 1).\end{aligned} \tag{3.13}$$

For instance, $V([k, k, k])$ is a one-dimensional space with a basis vector

$$\left|\begin{matrix}k & & k & & k\\ & k & & k & \\ & & k & & \end{matrix}\right\rangle. \tag{3.14}$$

The gl(3) module $V([0, -1, -1])$ is three-dimensional and has a basis

$$\left|\begin{matrix}0, -1, -1\\ -\delta_{3k}, -1\\ \delta_{1k} - 1\end{matrix}\right\rangle, \quad k = 1,2,3. \tag{3.15}$$

The space $V([-1, -1, -2])$ is also of dimension 3 with a basis

$$\left|\begin{matrix}-1, -1, -2\\ -1, \delta_{3k}, -2\\ -1 - \delta_{1k}\end{matrix}\right\rangle, \quad k = 1,2,3. \tag{3.16}$$

Throughout the paper we shall use also the following abbreviations:

$$[m + c]_n \equiv [m]_n + c = [m_{1n} + c, m_{2n} + c, \ldots, m_{nn} + c], \tag{3.17}$$

$$[m]_n^{\pm i} = [m_{1n} \pm \delta_{1i}, m_{2n} \pm \delta_{2i}, \ldots, m_{nn} \pm \delta_{ni}], \tag{3.18}$$

$$[m]_n + c^j = [m_{1n} + \delta_{1j}c, m_{2n} + \delta_{2j}c, \ldots, m_{nn} + \delta_{nj}c]. \tag{3.19}$$

For instance,

$$[-1, -1, -1]^1 = [-1, -1, -1] + 1^1 = [0, -1, -1],$$

$$[m - 1]_2^2 = [m_{12} - 1, m_{22}].$$

Let

$$\theta(x) = \begin{cases}1, & \text{for } x \geqslant 0,\\ 0, & \text{for } x < 0.\end{cases} \tag{3.20}$$

Then

$$[m]_n \pm \theta(x)^j = [m_{1n} \pm \theta(x)\delta_{1j}, \ldots, m_{nn} \pm \theta(x)\delta_{nj}]. \tag{3.21}$$

By the construction of the GZ basis in any sl(1,3) module V, it is clear that the representation of the even generators $E_{ij}, i,j = 1,2,3$ within any gl(3) submodule $V([m]_3)$ of V will be given with the known expressions for E_{ij}, considered as gl(3) generators in $V([m]_3)$ (see Ref. 8):

$$E_{kk}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = (m_{1k} + m_{2k} + \cdots + m_{kk} - m_{1,k-1} - m_{2,k-1} - m_{k-1,k-1})\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle,$$

$$E_{12}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = |(l_{12} - l_{11})(l_{22} - l_{11})|^{1/2}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11} + 1\end{matrix}\right\rangle,$$

$$E_{21}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = |(l_{12} - l_{11} + 1)(l_{22} - l_{11} + 1)|^{1/2}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11} - 1\end{matrix}\right\rangle,$$

$$E_{q3}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \sum_{j=1}^{2} S(j,q)\left|\frac{(\delta_{1q}l_{3-j,2} + \delta_{2q}l_{j2} - l_{11} + q - 1)\Pi_{k=1}^3(l_{k3} - l_{j2})}{(l_{12} - l_{22} - j + 1)(l_{12} - l_{22} - j + 2)}\right|^{1/2}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11} + \delta_{1q}\end{matrix}\right\rangle, \quad q = 1,2, \tag{3.22}$$

$$E_{3q}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \sum_{j=1}^{2} S(j,q)\left|\frac{(\delta_{1q}l_{3-j,2} + \delta_{2q}l_{j2} - l_{11} + 2 - q)\Pi_{k=1}^3(l_{k3} - l_{j2} + 1)}{(l_{12} - l_{22} + j - 1)(l_{12} - l_{22} + j - 2)}\right|^{1/2}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11} - \delta_{1q}\end{matrix}\right\rangle,$$

where

$$S(i,j) = \begin{cases}1, & \text{for } i \leqslant j,\\ -1, & \text{for } i > j,\end{cases} \tag{3.23}$$

and

$$l_{ij} = m_{ij} - i. \tag{3.24}$$

The Cartan subalgebra (2.8) is also a Cartan subalgebra of gl(3). Choose as a basis in H the generators

$$E_{ii} = e_{ii} + e_{00}, \quad i = 1,2,3, \tag{3.25}$$

and let E^1, E^2, E^3 be the dual to (3.25) basis in H. From (3.22) one derives, for any $h = \Sigma_{i=1}^3 \xi^i E_{ii}$,

$$h \left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle = \lambda(h) \left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle, \tag{3.26}$$

with

$$\lambda = m_{11}E^1 + (m_{12} + m_{22} - m_{11})E^2 + (m_{13} + m_{23} + m_{33} - m_{12} - m_{22})E^3. \tag{3.27}$$

Hence, every GZ pattern is a weight vector and the correspondence weight vector→weight is

$$\left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle \to m_{11}E^1 + (m_{12} + m_{22} - m_{11})E^2 + (m_{13} + m_{23} + m_{33} - m_{12} - m_{22})E^3. \tag{3.28}$$

The vector

$$x_\Lambda = \left| \begin{matrix} m_{13} & m_{23} & m_{33} \\ & m_{13} \quad m_{23} & \\ & m_{13} & \end{matrix} \right\rangle \tag{3.29}$$

is annihilated by all positive root vectors E_{ij}, $i<j = 1,2,3$,

$$E_{ij}x_\Lambda = 0, \quad i<j = 1,2,3. \tag{3.30}$$

Therefore, x_Λ is the highest weight vector of $V([m]_3)$ and

$$x_\Lambda \to \Lambda = \sum_{i=1}^{3} m_{i3}E^i \tag{3.31}$$

gives the corresponding highest weight.

For later use the representations of gl(3) in the tensor product spaces

$$V([c,c,c]) \otimes V([m]_3), \tag{3.32}$$

$$V([0,-1,-1]) \otimes V([m]_3), \tag{3.33}$$

$$V([-1,-1,-2]) \otimes V([m]_3), \tag{3.34}$$

will be of particular importance. In general, any gl(3) module $V([m']_3) \otimes V([m'']_3)$ is reducible[11]:

$$V([m']_3) \otimes V([m'']_3) = \sum \oplus V([m]_3). \tag{3.35}$$

This is not always the case. The gl(3) module (3.32) is irreducible,

$$V([c,c,c]) \otimes V([m]_3) = V([m+c]_3), \tag{3.36}$$

and all vectors

$$\left| \begin{matrix} c,c,c \\ c,c \\ c \end{matrix} \right\rangle \otimes \left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle = \left| \begin{matrix} [m+c]_3 \\ [m+c]_2 \\ m_{11}+c \end{matrix} \right\rangle \tag{3.37}$$

constitute a GZ basis in this space. The spaces (3.33) and (3.34) are reducible:

$$V([0,-1,-1]) \otimes V([m]_3) = \sum_{i=1}^{3} \oplus V([m-1]_3^i), \tag{3.38}$$

$$V([-1,-1,-2]) \otimes V([m]_3) = \sum_{i=1}^{3} \oplus V([m-1]_3^{-i}). \tag{3.39}$$

It is understood that the terms $V([m-1]_3^{\pm i})$ in (3.38) and (3.39), corresponding to nonlexical 3-tuples $[m-1]_3^{\pm i}$, have to be deleted from the sum. As an orthonormed basis in (3.38), one may choose either the tensor basis

$$\left| \begin{matrix} 0,-1,-1 \\ -\delta_{3k},-1 \\ \delta_{1k}-1 \end{matrix} \right\rangle \otimes \left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle, \quad k = 1,2,3, \tag{3.40}$$

or the union (1.2) of the GZ bases

$$\left| \begin{matrix} [m-1]_3^i \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle \quad \text{in } V([m-1]_3^i), \quad i = 1,2,3. \tag{3.41}$$

The coefficients relating the tensor basis (3.40) with the GZ basis (3.41) are by definition Clebsch–Gordan coefficients (CGC's) of gl(3):

$$\left| \begin{matrix} 0,-1,-1 \\ -\delta_{3k},-1 \\ \delta_{1k}-1 \end{matrix} \right\rangle \otimes \left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle = \sum_{i=1}^{3} \sum_{m'_{12},m'_{22}m'_{11}} \left\langle \begin{matrix} 0,-1,-1 & [m]_3 \\ -\delta_{3k},-1; & [m]_2 \\ \delta_{1k}-1 & m_{11} \end{matrix} \middle| \begin{matrix} [m-1]_3^i \\ [m']_2 \\ m'_{11} \end{matrix} \right\rangle \left| \begin{matrix} [m-1]_3^i \\ [m']_2 \\ m'_{11} \end{matrix} \right\rangle, \quad k = 1,2,3. \tag{3.42}$$

Proposition 1: The CGC's in (3.42) are equal to zero if $m'_{11} \neq m_{11}, m_{11}-1$; $m'_{12} \neq m_{12}, m_{12}-1$; $m'_{22} \neq m_{22}, m_{22}-1$. More precisely,

$$\left| \begin{matrix} 0,-1,-1 \\ -\delta_{3k},-1 \\ \delta_{1k}-1 \end{matrix} \right\rangle \otimes \left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle = \sum_{i=1}^{3} \sum_{j=1}^{\theta(2-k)+1} \left\langle \begin{matrix} 0,-1,-1 & [m]_3 \\ -\delta_{3k},-1; & [m]_2 \\ \delta_{1k}-1 & m_{11} \end{matrix} \middle| \begin{matrix} [m-1]_3^i \\ [m-1]_2+\theta(2-k)^j \\ m_{11}+\delta_{1k}-1 \end{matrix} \right\rangle \left| \begin{matrix} [m-1]_3^i \\ [m-1]_2+\theta(2-k)^j \\ m_{11}+\delta_{1k}-1 \end{matrix} \right\rangle. \tag{3.43}$$

Proof: Consider first a vector

$$\left| \begin{matrix} 0,-1,-1 \\ -1,-1 \\ -1 \end{matrix} \right\rangle \otimes \left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle. \tag{3.44}$$

According to (3.11) and (3.37)

$$\left| \begin{matrix} 0,-1,-1 \\ -1,-1 \\ -1 \end{matrix} \right\rangle \otimes \left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle = V(-1) \otimes V(m_{11}) = V(m_{11}-1) \subset V([-1,-1]) \otimes V([m]_2)$$
$$= V([m-1]_2) \subset V([0,-1,-1]) \otimes V([m]_3). \tag{3.45}$$

Therefore, the CGC's in (3.42) are zeros if $m'_{11} \neq m_{11}-1, m'_{12} \neq m_{12}-1, m'_{22} = m_{22}-1$. Hence, for $k=3$, Eq. (3.42) reduces to (3.43):

$$\left|\begin{matrix}0,-1,-1\\ -1,-1\\ -1\end{matrix}\right\rangle \otimes \left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \sum_{i=1}^{3} \left\langle\begin{matrix}0,-1,-1\\ -1,-1\\ -1\end{matrix};\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\middle|\begin{matrix}[m-1]_3^i\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle \left|\begin{matrix}[m-1]_3^i\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle. \tag{3.46}$$

Consider the case $k=1,2$ in (3.42). Then

$$\left|\begin{matrix}0,-1,-1\\ 0,-1\\ \delta_{1k}-1\end{matrix}\right\rangle \otimes \left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle \in V(\delta_{1k}-1)\otimes V(m_{11}) \subset V([0,-1])\otimes V([m]_2) \subset V([0,-1,-1])\otimes V([m]_3). \tag{3.47}$$

Taking into account the decompositions

$$\begin{aligned} &V(\delta_{1k}-1)\otimes V(m_{11}) = V(m_{11}+\delta_{1k}-1),\\ &V([0,-1])\otimes V([m]_2) = V([m_{12}-1,m_{22}])\oplus V([m_{12},m_{22}-1]), \end{aligned} \tag{3.48}$$

one concludes that for $k=1,2$, (3.40) is a linear combination of vectors

$$\left|\begin{matrix}[m-1]_3^i\\ [m-1]_2^j\\ m_{11}+\delta_{1k}-1\end{matrix}\right\rangle, \quad i=1,2,3, \quad j=1,2, \quad k=1,2, \tag{3.49}$$

i.e.,

$$\left|\begin{matrix}0,-1,-1\\ 0,-1\\ \delta_{1k}-1\end{matrix}\right\rangle \otimes \left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{2} \left\langle\begin{matrix}0,-1,-1\\ 0,-1\\ \delta_{1k}-1\end{matrix};\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\middle|\begin{matrix}[m-1]_3^i\\ [m-1]_2^j\\ m_{11}+\delta_{1k}-1\end{matrix}\right\rangle \left|\begin{matrix}[m-1]_3^i\\ [m-1]_2^j\\ m_{11}+\delta_{1k}-1\end{matrix}\right\rangle. \tag{3.50}$$

The above equality coincides with (3.43) in the case $k=1,2$, and this completes the proof.

Multiplying on the left both sides of (3.43) with

$$\left|\begin{matrix}1,1,1\\ 1,1\\ 1\end{matrix}\right\rangle \in V([1,1,1]) \tag{3.51}$$

and, taking into account (3.37), one derives the following property of the CGC's:

$$\left\langle\begin{matrix}0,-1,-1\\ -\delta_{3k},-1\\ \delta_{1k}-1\end{matrix};\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\middle|\begin{matrix}[m-1]_3^i\\ [m-1]_2+\theta(2-k)^j\\ m_{11}+\delta_{1k}-1\end{matrix}\right\rangle = \left\langle\begin{matrix}1,0,0\\ 1-\delta_{3k},0\\ \delta_{1k}\end{matrix};\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\middle|\begin{matrix}[m]_3^i\\ [m]_2+\theta(2-k)^j\\ m_{11}+\delta_{1k}\end{matrix}\right\rangle. \tag{3.52}$$

Therefore, (3.43) can be written also as

$$\left|\begin{matrix}0,-1,-1\\ -\delta_{3k},-1\\ \delta_{1k}-1\end{matrix}\right\rangle \otimes \left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{\theta(2-k)+1} \left\langle\begin{matrix}1,0,0\\ 1-\delta_{3k},0\\ \delta_{1k}\end{matrix};\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\middle|\begin{matrix}[m]_3^i\\ [m]_2+\theta(2-k)^j\\ m_{11}+\delta_{1k}\end{matrix}\right\rangle \left|\begin{matrix}[m-1]_3^i\\ [m-1]_2+\theta(2-k)^j\\ m_{11}+\delta_{1k}-1\end{matrix}\right\rangle. \tag{3.53}$$

In the particular representations we consider, the CGC's are products of gl(3)- and gl(2)-scalar factors[12]

$$\begin{aligned} &\left\langle\begin{matrix}1,0,0\\ 1-\delta_{3k},0\\ \delta_{1k}\end{matrix};\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\middle|\begin{matrix}[m]_3^i\\ [m]_2+\theta(2-k)^j\\ m_{11}+\delta_{1k}\end{matrix}\right\rangle \\ &\quad = \left(\begin{matrix}1,0,0\\ 1-\delta_{3k},0\end{matrix};\begin{matrix}[m]_3\\ [m]_2\end{matrix}\middle|\begin{matrix}[m]_3^i\\ [m]_2+\theta(2-k)^j\end{matrix}\right)\left(\begin{matrix}1-\delta_{3k},0\\ \delta_{1k}\end{matrix};\begin{matrix}[m]_2\\ m_{11}\end{matrix}\middle|\begin{matrix}[m]_2+\theta(2-k)^j\\ m_{11}+\delta_{1k}\end{matrix}\right). \end{aligned} \tag{3.54}$$

The gl(3)-scalar factors (i.e., isofactors) read[13]

$$\left(\begin{matrix}1,0,0\\ 0,0\end{matrix};\begin{matrix}[m]_3\\ [m]_2\end{matrix}\middle|\begin{matrix}[m]_3^i\\ [m]_2\end{matrix}\right) = \left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{i3}-1)}{\Pi_{k\neq i=1}^{3}(l_{k3}-l_{i3})}\right|^{1/2}, \tag{3.55}$$

$$\left(\begin{matrix}1,0,0\\ 1,0\end{matrix};\begin{matrix}[m]_3\\ [m]_2\end{matrix}\middle|\begin{matrix}[m]_3^i\\ [m]_2^j\end{matrix}\right) = S(i,j)\left|\frac{\Pi_{k\neq j=1}^{2}(l_{k2}-l_{i3}-1)\Pi_{k\neq i=1}^{3}(l_{k3}-l_{j2})}{\Pi_{k\neq i=1}^{3}(l_{k3}-l_{i3})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{j2}-1)}\right|^{1/2}. \tag{3.56}$$

The gl(2)-scalar factors [i.e., the gl(2) CGC's] are

$$\begin{pmatrix}1,0 & [m]_2 \\ 0 & ;m_{11}\end{pmatrix}\left|\begin{matrix}[m]_2^j \\ m_{11}\end{matrix}\right) = \left|\frac{l_{j2}-l_{11}+1}{l_{12}-l_{22}}\right|^{1/2}, \tag{3.57}$$

$$\begin{pmatrix}1,0 & [m]_2 \\ 1 & ;m_{11}\end{pmatrix}\left|\begin{matrix}[m]_2^j \\ m_{11}+1\end{matrix}\right) = S(j,1)\left|\frac{\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11})}{l_{12}-l_{22}}\right|^{1/2}. \tag{3.58}$$

All other scalar factors vanish.

The relations (3.54)–(3.58) show that the CGC's are real numbers. Since, moreover, the tensor basis (3.40) and the GZ basis (3.41) are orthonormal, one can immediately invert the relation (3.43)

$$\left|\begin{matrix}[m-1]_3^i \\ [m-1]_2 \\ m_{11}-1\end{matrix}\right\rangle = \sum_{k=1}^{3}\sum_{j=1}^{\theta(2-k)+1}\left\langle\begin{matrix}1,0,0 & [m]_3 \\ 1-\delta_{3k},0;[m]_2-\theta(2-k)^j \\ \delta_{1k} & m_{11}-\delta_{1k}\end{matrix}\right|\left.\begin{matrix}[m]_3^i \\ [m]_2 \\ m_{11}\end{matrix}\right\rangle \times\left|\begin{matrix}0,-1,-1 \\ -\delta_{3k},-1 \\ \delta_{1k}-1\end{matrix}\right\rangle\otimes\left|\begin{matrix}[m]_3 \\ [m]_2-\theta(2-k)^j \\ m_{11}-\delta_{1k}\end{matrix}\right\rangle. \tag{3.59}$$

In the gl(3) module (3.39), we also introduce an orthonormal tensor basis (see 3.16)

$$\left|\begin{matrix}-1,-1,-2 \\ -1,\delta_{3k}-2 \\ -1-\delta_{1k}\end{matrix}\right\rangle\otimes\left|\begin{matrix}[m]_3 \\ [m]_2 \\ m_{11}\end{matrix}\right\rangle, k=1,2,3, \tag{3.60}$$

and a GZ basis

$$\left|\begin{matrix}[m-1]_3^{-i} \\ [m]_2 \\ m_{11}\end{matrix}\right\rangle,\quad i=1,2,3. \tag{3.61}$$

Then in a similar way as in the Proposition 1, one proves the following.

Proposition 2: The tensor basis in the gl(3) module (3.39), expressed in terms of the GZ basis, reads

$$\left|\begin{matrix}-1,-1,-2 \\ -1,\delta_{3k}-2 \\ -1-\delta_{1k}\end{matrix}\right\rangle\otimes\left|\begin{matrix}[m]_3 \\ [m]_2 \\ m_{11}\end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{\theta(2-k)+1}\left\langle\begin{matrix}0,0,-1 & [m]_3 \\ 0,\delta_{3k}-1;[m]_2 \\ -\delta_{1k} & m_{11}\end{matrix}\right|\left.\begin{matrix}[m]_3^{-i} \\ [m]_2-\theta(2-k)^j \\ m_{11}-\delta_{1k}\end{matrix}\right\rangle\left|\begin{matrix}[m-1]_3^{-i} \\ [m-1]_2-\theta(2-k)^j \\ m_{11}-\delta_{1k}-1\end{matrix}\right\rangle. \tag{3.62}$$

The inverse relation

$$\left|\begin{matrix}[m-1]_3^{-i} \\ [m-1]_2 \\ m_{11}-1\end{matrix}\right\rangle = \sum_{k=1}^{3}\sum_{j=1}^{\theta(2-k)+1}\left\langle\begin{matrix}0,0,-1 & [m]_3 \\ 0,\delta_{3k}-1;[m]_2+\theta(2-k)^j \\ -\delta_{1k} & m_{11}+\delta_{1k}\end{matrix}\right|\left.\begin{matrix}[m]_3^{-i} \\ [m]_2 \\ m_{11}\end{matrix}\right\rangle\left|\begin{matrix}-1,-1,-2 \\ -1,\delta_{3k}-2 \\ -1-\delta_{1k}\end{matrix}\right\rangle\otimes\left|\begin{matrix}[m]_3 \\ [m]_2+\theta(2-k)^j \\ m_{11}+\delta_{1k}\end{matrix}\right\rangle. \tag{3.63}$$

In this case the CGC's are also products of gl(3)- and gl(2)-scalar factors

$$\left\langle\begin{matrix}0,0,-1 & [m]_3 \\ 0,\delta_{3k}-1;[m]_2 \\ -\delta_{1k} & m_{11}\end{matrix}\right|\left.\begin{matrix}[m]_3^{-i} \\ [m]_2-\theta(2-k)^j \\ m_{11}-\delta_{1k}\end{matrix}\right\rangle = \begin{pmatrix}0,0,-1 & [m]_3 \\ 0,\delta_{3k}-1 & ;[m]_2\end{pmatrix}\left|\begin{matrix}[m]_3^{-i} \\ [m]_2-\theta(2-k)^j\end{matrix}\right)\begin{pmatrix}0,\delta_{3k}-1 & [m]_2 \\ -\delta_{1k} & ;[m]_{11}\end{pmatrix}\left|\begin{matrix}[m]_2-\theta(2-k)^j \\ m_{11}-\delta_{1k}\end{matrix}\right), \tag{3.64}$$

where (Ref. 12, p. 153)

$$\begin{pmatrix}0,0,-1 & [m]_3 \\ 0,0 & ;[m]_2\end{pmatrix}\left|\begin{matrix}[m]_3^{-i} \\ [m]_2\end{matrix}\right) = \left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{i3})}{\Pi_{k\neq i=1}^{3}(l_{k3}-l_{i3})}\right|^{1/2}, \tag{3.65}$$

$$\begin{pmatrix}0,0,-1 & [m]_3 \\ 0,-1 & ;[m]_2\end{pmatrix}\left|\begin{matrix}[m]_3^{-i} \\ [m]_2^{-j}\end{matrix}\right) = -S(i,j)\left|\frac{\Pi_{k\neq j=1}^{2}(l_{k2}-l_{i3})\Pi_{k\neq i=1}^{3}(l_{k3}-l_{j2}+1)}{\Pi_{k\neq j=1}^{2}(l_{k2}-l_{j2}+1)\Pi_{k\neq i=1}^{3}(l_{k3}-l_{i3})}\right|^{1/2}, \tag{3.66}$$

$$\begin{pmatrix}0,-1 & [m]_2 \\ 0 & ;m_{11}\end{pmatrix}\left|\begin{matrix}[m]_2^{-j} \\ m_{11}\end{matrix}\right) = \left|\frac{l_{j2}-l_{11}}{l_{12}-l_{22}}\right|^{1/2}, \tag{3.67}$$

$$\begin{pmatrix}0,-1 & [m]_2 \\ -1 & ;m_{11}\end{pmatrix}\left|\begin{matrix}[m]_2^{-j} \\ m_{11}-1\end{matrix}\right) = -S(j,1)\left|\frac{\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11}+1)}{l_{12}-l_{22}}\right|^{1/2}. \tag{3.68}$$

All other gl(3)- and gl(2)-scalar factors vanish.

We already have a partial answer to the representation problem. Relations (3.22) give representations of the even generators in the GZ basis of any sl(1,3) module V. It remains to determine V to be a sl(1,3)-irreducible module, to decompose it into irreducible gl(3) submodules $V([m]_3)$, $V([m']_3)$,...,i.e.,

$$V = V([m]_3) \oplus V([m']_3) \oplus \cdots, \tag{3.69}$$

and to determine expressions for the odd generators in the GZ basis in V. In the next section we perform the first step in the above program: we introduce the sl(1,3) modules which carry all irreducible representations. Moreover, we find expressions for all generators, however, in another basis—the induced one.

IV. INDUCED REPRESENTATIONS

Let

$$P_+ = \text{lin.env.}\{e_{0k} \mid k = 1,2,3\} \subset \text{sl}(1,3), \tag{4.1}$$

and let P be the subalgebra, which is a direct sum of the subspaces P_+ and $\text{sl}_0(1,3) = \text{gl}(3)$,

$$P = \text{gl}(3) \oplus P_+. \tag{4.2}$$

Following the method developed in Ref. 3, we construct here the representations of sl(1,3) induced from representations of its subalgebra P. To this end consider a finite-dimensional irreducible gl(3) module $V_0([m]_3)$ and extend it to a P module, assuming

$$P_+ V_0([m]_3) = 0. \tag{4.3}$$

Denote by U and $U_P \subset U$ the universal enveloping algebras of sl(1,3) and P, correspondingly. Let

$$\overline{V}([m]_3) = \text{Ind}_P^{\text{sl}(1,3)} V_0([m]_3) \tag{4.4}$$

be the factor space of $U \otimes V_0([m]_3)$ with respect to the subspace

$$I([m]_3) = \text{lin.env.}\{gp \otimes v - g \otimes pv \mid g \in U,\ p \in U_P\}, \tag{4.5}$$

i.e.,

$$\overline{V}([m]_3) = U \otimes V_0([m]_3)/I([m]_3). \tag{4.6}$$

The space $\overline{V}([m]_3)$ is equipped with a structure of an sl(1,3) module in a natural way:

$$g(u \otimes v) = gu \otimes v, \quad g \in \text{sl}(1,3), \quad u \in U, \quad v \in V_0([m]_3). \tag{4.7}$$

From the Poincaré–Birkhoff–Witt theorem,[3] it follows that U is a linear span of all elements of the form

$$g = (e_{10})^{\theta_1}(e_{20})^{\theta_2}(e_{30})^{\theta_3} p, \quad p \in U_P, \quad \theta_1, \theta_2, \theta_3 = 0,1. \tag{4.8}$$

The restriction $\theta = 0,1$ comes from the observation that $(e_{i0})^2 = 0$ in U. Since, for any g defined in (4.8) and $v \in V_0([m]_3)$,

$$g \otimes v = (e_{10})^{\theta_1}(e_{20})^{\theta_2}(e_{30})^{\theta_3} p \otimes v = (e_{10})^{\theta_1}(e_{20})^{\theta_2}(e_{30})^{\theta_3} \otimes pv,$$

one concludes that

$$\overline{V}([m]_3) = \text{lin.env.}\{(e_{10})^{\theta_1}(e_{20})^{\theta_2}(e_{30})^{\theta_3} \otimes v \mid v \in V_0([m]_3),\ \theta_1, \theta_2, \theta_3 = 0,1\}. \tag{4.9}$$

Let $T \subset U$ be the subalgebra spanned on all polynomials of the generators e_{10}, e_{20}, e_{30},

$$T = \text{lin.env.}\{(e_{10})^{\theta_1}(e_{20})^{\theta_2}(e_{30})^{\theta_3} \mid \theta_1, \theta_2, \theta_3 = 0,1\}. \tag{4.10}$$

The relation (4.9) shows that as a linear space the sl(1,3) module $\overline{V}([m]_3)$ is isomorphic to the tensor product of T and $V_0(m_3)$, i.e.,

$$\overline{V}([m]_3) = T \otimes V_0([m]_3). \tag{4.11}$$

As one possible basis in $\overline{V}([m]_3)$, one may choose the vectors

$$|\theta_1, \theta_2, \theta_3;(m)_3\rangle = (e_{10})^{\theta_1}(e_{20})^{\theta_2}(e_{30})^{\theta_3} \otimes |(m)_3\rangle_0, \tag{4.12}$$

with $\theta_1, \theta_2, \theta_3 = 0$ or 1 and

$$|(m)_3\rangle_0 = \left| \begin{matrix} [m]_3 \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle_0 \in \Gamma([m]_3)$$

being an arbitrary GZ pattern from the gl(3) module $V_0([m]_3)$.

The space $\overline{V}([m]_3)$ is always considered as a Z_2-graded space

$$\overline{V}([m]_3) = \overline{V}_0([m]_3) \oplus \overline{V}_1([m]_3) \tag{4.13}$$

with an even subspace $\overline{V}_0([m]_3)$ [resp. an odd subspace $\overline{V}_1([m]_3)$], spanned on all vectors (4.12) for which $\theta_1 + \theta_2 + \theta_3$ is an even (resp. an odd) number.

We call the basis (4.12) an induced basis. The representations of the sl(1,3) generators in this basis read[14] ($i,j,k = 1,2,3$)

$$\begin{aligned}
e_{i0}|\theta_1, \theta_2, \theta_3;(m)_3\rangle &= (-1)^{\theta_1 + \cdots + \theta_i}(1-\theta_i) \\
&\quad \times |\theta_1,\ldots,\theta_i + 1,\ldots,\theta_3;(m)_3\rangle, \\
e_{0i}|\theta_1, \theta_2, \theta_3;(m)_3\rangle & \\
= \sum_{k=1}^{3} \theta_k(-1)^{\theta_1 + \cdots + \theta_{k-1}} & \\
\times |\theta_1,\ldots,\theta_{k-1},0,\theta_{k+1},\ldots,\theta_3;E_{ki}(m)_3\rangle & \\
+ \theta_i(-1)^{\theta_1 + \cdots + \theta_i}\Big(\sum_{k \neq i = 1}^{3} \theta_k\Big) & \\
\times |\theta_1,\ldots,\theta_{i-1},0,\theta_{i+1},\ldots,\theta_3;(m)_3\rangle, &
\end{aligned} \tag{4.14}$$

$$\begin{aligned}
E_{ij}|\theta_1,\theta_2,\theta_3;(m)_3\rangle & \\
\equiv (e_{ij} + \delta_{ij}e_{00})|\theta_1,\theta_2,\theta_3;(m)_3\rangle & \\
= |\theta_1,\theta_2,\theta_3;E_{ij}(m)_3\rangle & \\
- \delta_{ij}(\theta_1 + \theta_2 + \theta_3 - \theta_i)|\theta_1,\theta_2,\theta_3;(m)_3\rangle & \\
- \theta_j(1-\theta_i)(-1)^{\theta_i + \cdots + \theta_j} & \\
\times |\theta_1,\ldots,\theta_i + 1,\ldots,\theta_j - 1,\ldots,\theta_3;(m)_3\rangle. &
\end{aligned}$$

A representation of sl(1,3), defined by Eqs. (4.14), is said to be induced from the representation $[m]_3$ of gl(3). This representation is either irreducible, and in this case it is typical representation, or it is indecomposable. In the latter case, $\overline{V}([m]_3)$ contains an unique maximal invariant subspace $0 \neq \overline{I}([m]_3) \neq \overline{V}([m]_3)$. The representation of sl(1,3) in the factor space $\overline{V}([m]_3)/\overline{I}([m]_3)$ is irreducible. All such representations are called nontypical.[3] In this way one may construct all irreducible representations. First, however, one has to determine all $\bar{I}([m]_3)$, which turns out to be a rather nontrivial task. The difficulty comes from the circumstance that the vectors (4.12) from the induced basis have, in general, nonzero projections both on $\bar{I}([m]_3)$ and on its orthogonal complement $\bar{V}([m]_3)\ominus\bar{I}([m]_3)$. In order to specify $\bar{I}([m]_3)$, we shall choose another basis with the property that every basis vector lies either in $\bar{I}([m]_3)$ or is orthogonal to it. To this end, observe that as a gl(3) module, $\bar{V}([m]_3)$ is completely reducible. Therefore, both $\bar{I}([m]_3)$ and $\bar{V}([m]_3)\ominus\bar{I}([m]_3)$ are gl(3) invariant. Decompose them into irreducible gl(3) submodules $V_l, l=1,...,n$:

$$\bar{I}([m]_3)=\sum_{l=1}^{k}\oplus V_l,\quad \bar{V}([m]_3)\ominus\bar{I}([m]_3)=\sum_{l=k+1}^{n}\oplus V_l. \tag{4.15}$$

Therefore,

$$\bar{V}([m]_3)=\sum_{l=1}^{n}\oplus V_l \tag{4.16}$$

and if Γ_l is a GZ basis in V_l, then the basis $\Gamma=\cup_l\Gamma_l$ in $\bar{V}([m]_3)$ will have the required properties. In the next section we perform the decomposition (4.16), choose the basis Γ, and write down the representation (4.14) in this basis. In the case of nontypical modules $\bar{V}([m]_3)$, this will simplify considerably the determination of the maximal invariant submodules $\bar{I}([m]_3)$. Moreover, the vectors $\Gamma=\cup_{l=k+1}^{n}\Gamma_l$ will give a basis in the irreducible sl(1,3) module $\bar{V}([m]_3)/\bar{I}([m]_3)$. Here we first prove a criterion for the irreducibility of the induced representations.

Proposition 3: The sl(1,3) module $\bar{V}([m]_3)$ is typical iff $m_{13}\neq 0$, $m_{23}\neq 1$, $m_{33}\neq 2$.

Proof: One can show in a straightforward way that all modules, corresponding to $m_{13}\neq 0$, $m_{23}\neq 1$, $m_{33}\neq 2$, are irreducible.[14] It is not simple, however, to prove the inverse. Therefore, we shall use a general criterion for the irreducibility (Proposition 2.9 in Ref. 3) stating (in our case) that the induced sl(1,3) module $\bar{V}([m]_3)$ is irreducible iff

$$(\Lambda+\rho,\epsilon^0-\epsilon^k)\neq 0,\quad \text{for all } k=1,2,3. \tag{4.17}$$

Here $\epsilon^0-\epsilon^i$ are the odd positive roots (2.18), ρ is given with the Eq. (2.19), and Λ is the highest weight of the gl(3) module $V_0([m]_3)$. From the Gel'fand–Zetlin formulas (3.22), one derives that any Cartan element $h=\Sigma_{i=1}^3\xi^i E_{ii}$ acts on the highest weight vector ($\mathbb{1}$ is the unity in U)

$$\bar{x}_\Lambda=|0,0,0;x_\Lambda\rangle=\mathbb{1}\otimes x_\Lambda,\quad x_\Lambda\in V_0([m]_3), \tag{4.18}$$

as

$$h\bar{x}_\Lambda=\sum_{i=1}^{3}\xi^i m_{i3}\bar{x}_\Lambda=\Big(\sum_{i=1}^{3}m_{i3}\epsilon^i\Big)(h)\bar{x}_\Lambda. \tag{4.19}$$

Therefore, as an element from $\overset{*}{H}{}'$ (with a basis 2.14) Λ reads

$$\Lambda=m_{13}\epsilon^1+m_{23}\epsilon^2+m_{33}\epsilon^3. \tag{4.20}$$

Inserting (4.20) in (4.17) and, taking into account (2.17), (2.19), and (2.20), one obtains

$$(\Lambda+\rho,\epsilon^0-\epsilon^k)=(k-m_{k3}-1)/4,\quad k=1,2,3. \tag{4.21}$$

The right-hand side differs from zero if and only if $m_{k3}\neq k-1$, $k=1,2,3$, which completes the proof.

In order to relate these results with the Kac classification, we recall (Ref. 3, Proposition 2.3) that the finite-dimensional IR's of sl(1,3) are labeled with three numbers $(\alpha_0,\alpha_1,\alpha_2)$, where α_0 is an aribtrary complex number and α_1,α_2 are arbitrary non-negative integers. These numbers are the eigenvalues of the Cartan elements h_0, h_1, h_2 [see (2.21)] on the highest weight vector (4.18). Since

$$h_0=E_{11},\quad h_1=E_{11}-E_{22},\quad h_2=E_{22}-E_{33}, \tag{4.22}$$

from (3.22) one can deduce

$$\alpha_0=m_{13},\quad \alpha_1=m_{13}-m_{23},\quad \alpha_2=m_{23}-m_{33}. \tag{4.23}$$

We conclude that the typical representations of sl(1,3) have in the Kac notation the signature

$$(\alpha_0,\alpha_1,\alpha_2)=(m_{13},m_{13}-m_{23},m_{23}-m_{33}),$$
$$m_{13}\neq 0,\quad m_{23}\neq 1,\quad m_{33}\neq 2. \tag{4.24}$$

V. REPRESENTATIONS OF sl(1,3) IN A GZ BASIS

A. Structure of $\bar{V}([m]_3)$ with respect to gl(3)

Consider $\bar{V}([m]_3)=T\otimes V_0([m]_3)$ as a gl(3) module. Since

$$\begin{aligned}[E_{ij},(e_{10})^{\theta_1}(e_{20})^{\theta_2}(e_{30})^{\theta_3}] &=\delta_{j1}\theta_1 e_{i0}(e_{20})^{\theta_2}(e_{30})^{\theta_3}+\delta_{j2}\theta_2(e_{10})^{\theta_1}e_{i0}(e_{30})^{\theta_3}\\ &\quad+\delta_{j3}\theta_3(e_{10})^{\theta_1}(e_{20})^{\theta_2}e_{i0},\end{aligned} \tag{5.1}$$

the subspace T of U is invariant with respect to the adjoint repsentation of U, restricted to gl(3),

$$[\mathrm{gl}(3),T]\subset T. \tag{5.2}$$

Therefore, T can be considered as a gl(3) module. Since, moreover, for every $g\in$gl(3) and $t\otimes v\in T\otimes V_0([m]_3)$

$$g(t\otimes v)=(\mathrm{ad}\,g)t\otimes v+t\otimes gv, \tag{5.3}$$

the representation $D_{T\otimes V_0}$ of gl(3) in $T\otimes V_0([m]_3)$ is a tensor product of the adjoint representation D_T of gl(3) in T and its representation $[m]_3$ in $V_0([m]_3)$,

$$D_{T\otimes V_0}=D_T\otimes[m]_3. \tag{5.4}$$

The gl(3) module T is reducible. Each one from its subspaces

$$\begin{aligned}T_1&=\text{lin.env.}\{(e_{10})^0(e_{20})^0(e_{30})^0\equiv\mathbb{1}\},\\ T_2&=\text{lin.env.}\{e_{10},e_{20},e_{30}\},\\ T_3&=\text{lin.env.}\{e_{10}e_{20},\ e_{10}e_{30},\ e_{20}e_{30}\},\\ T_4&=\text{lin.env.}\{e_{10}e_{20}e_{30}\}\end{aligned} \tag{5.5}$$

is gl(3) invariant and

$$T=T_1\oplus T_2\oplus T_3\oplus T_4. \tag{5.6}$$

It is a straightforward computation to show that every subspace T_i is irreducible. The Gel'fand–Zetlin labels for each space and the corresponding basis reads

$$T_1=V([0,0,0]):\ \left|\begin{matrix}0,0,0\\0,0\\0\end{matrix}\right\rangle=(e_{10})^0(e_{20})^0(e_{30})^0=\mathbb{1};$$

$$T_2=V([0,-1,-1]):\ \left|\begin{matrix}0,-1,-1\\0,-1\\0\end{matrix}\right\rangle=e_{10},$$

$$\left|\begin{matrix}0,-1,-1\\0,-1\\-1\end{matrix}\right\rangle=e_{20},$$

$$\left|\begin{matrix}0,-1,-1\\ -1,-1\\ -1\end{matrix}\right\rangle = e_{30};$$

$$\tag{5.7}$$

$$T_3 = V([-1,-1,-2]):\quad \left|\begin{matrix}-1,-1,-2\\ -1,-1\\ -1\end{matrix}\right\rangle = e_{10}e_{20},$$

$$\left|\begin{matrix}-1,-1,-2\\ -1,-2\\ -1\end{matrix}\right\rangle = e_{10}e_{30},$$

$$\left|\begin{matrix}-1,-1,-2\\ -1,-2\\ -2\end{matrix}\right\rangle = e_{20}e_{30};$$

$$T_4 = V([-2,-2,-2]):\quad \left|\begin{matrix}-2,-2,-2\\ -2,-2\\ -2\end{matrix}\right\rangle = e_{10}e_{20}e_{30}.$$

Thus, D_T is a direct sum of four gl(3) irreducible representations,

$$D_T = [0,0,0] \oplus [0,-1,-1] \oplus [-1,-1,-2] \oplus [-2,-2,-2]. \tag{5.8}$$

The corresponding gl(3) module T is a direct sum of its irreducible submodules,

$$T = V([0,0,0]) \oplus V([0,-1,-1]) \oplus V([-1,-1,-2]) \oplus V([-2,-2,-2]). \tag{5.9}$$

Inserting (5.9) in (4.11), we have

$$\overline{V}([m]_3) = D_1 \oplus D_2 \oplus D_3 \oplus D_4, \tag{5.10}$$

where

$$\begin{aligned} D_1 &= V([0,0,0]) \otimes V_0([m]_3),\\ D_2 &= V([0,-1,-1]) \otimes V_0([m]_3),\\ D_3 &= V([-1,-1,-2]) \otimes V_0([m]_3),\\ D_4 &= V([-2,-2,-2]) \otimes V_0([m]_3). \end{aligned} \tag{5.11}$$

The subspaces D_1 and D_4 are gl(3) irreducible and, according to (3.36),

$$\begin{aligned} &V([0,0,0]) \otimes V_0([m]_3) = V([m]_3),\\ &V([-2,-2,-2]) \otimes V_0([m]_3) = V([m-2]_3). \end{aligned} \tag{5.12}$$

The decomposition of D_2 and D_3 was already given [see (3.38) and (3.39)]:

$$V([0,-1,-1]) \otimes V_0([m]_3) = \sum_{i=1}^{3} \oplus V([m-1]_3^i), \tag{5.13}$$

$$V([-1,-1,-2]) \otimes V_0([m]_3) = \sum_{i=1}^{3} \oplus V([m-1]_3^{-i}). \tag{5.14}$$

Inserting (5.12)–(5.14) in (5.10), we finally conclude that the induced sl(1,3) module $\overline{V}([m]_3)$ is a direct sum of (no more than) eight gl(3)-irreducible submodules,

$$\begin{aligned} \overline{V}([m]_3) = V([m]_3) &\oplus \sum_{i=1}^{3} \oplus V([m-1]_3^i)\\ &\oplus \sum_{i=1}^{3} \oplus V([m-1]_3^{-i}) \oplus V([m-2]_3). \end{aligned} \tag{5.15}$$

In order to turn $\overline{V}([m]_3)$ into a Hilbert space, we introduce a natural metric in it: within every irreducible gl(3) submodule $V([m']_3)$, entering in the sum (5.15), we choose as a basis an orthonormal Gel'fand–Zetlin basis $\Gamma([m']_3)$ and extend the metric on $\overline{V}([m]_3)$ assuming that the sum (5.15) is orthogonal. [This is in fact a unique possibility if one requires that the Hermitian operators on each $V([m']_3)$ remain Hermitian on $\overline{V}([m]_3)$. For instance, the vectors from different irreducible submodules in (5.15) correspond to different eigenvalues of the second-order Casimir operator of gl(3) and, therefore, have to be orthogonal.] Then the union $\overline{\Gamma}([m]_3)$ of the GZ bases of all $V([m']_3)$ constitute an orthonormal basis in $\overline{V}([m]_3)$, which also will be called a GZ basis.

Since $\overline{V}([m]_3)$ is a direct sum of inequivalent gl(3) submodules, the decomposition (5.15) is unique. Therefore, if $\overline{V}([m]_3)$ is nontypical, each direct summand in (5.15) is either a subspace of $\overline{I}([m]_3)$ or is orthogonal to it. The maximal invariant subspace $\overline{I}([m]_3)$ is an orthogonal sum of gl(3)-irreducible submodules $V([m']_3)$ from the decomposition (5.15).

In view of (3.37) and (5.7), one can write down immediately the relation between the GZ basis and the induced basis in $V([m]_3)$ and $V([m-2]_3)$: in $V([m]_3)$,

$$\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \mathbf{1} \otimes \left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle_0; \tag{5.16}$$

in $V([m-2]_3)$,

$$\left|\begin{matrix}[m-2]_3\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle = e_{10}e_{20}e_{30} \otimes \left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle_0. \tag{5.17}$$

To write down the relations in $V([m-1]_3^i)$, note that according to (5.7)

$$e_{k0} = \left|\begin{matrix}0,-1,-1\\ -\delta_{3k},-1\\ \delta_{1k}-1\end{matrix}\right\rangle, \quad k = 1,2,3. \tag{5.18}$$

Inserting (5.18) in (3.59) and (3.53), we have

$$\left|\begin{matrix}[m-1]_3^i\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle = \sum_{k=1}^{3} e_{k0} \sum_{j=1}^{\theta(2-k)+1} \left\langle\begin{matrix}1,0,0 & [m]_3\\ 1-\delta_{3k},0;[m]_2-\theta(2-k)^j\\ \delta_{1k} \quad m_{11}-\delta_{1k}\end{matrix}\right|\left.\begin{matrix}[m]_3^i\\ [m]_2\\ m_{11}\end{matrix}\right\rangle \otimes \left|\begin{matrix}[m]_3\\ [m]_2-\theta(2-k)^j\\ m_{11}-\delta_{1k}\end{matrix}\right\rangle_0. \tag{5.19}$$

The inverse relation reads

$$e_{k0}\otimes\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle_0=\sum_{i=1}^{3}\sum_{j=1}^{\theta(2-k)+1}\left\langle\begin{matrix}1,0,0 & [m]_3\\ 1-\delta_{3k},0;[m]_2\\ \delta_{1k} & m_{11}\end{matrix}\middle|\begin{matrix}[m]_3^i\\ [m]_2+\theta(2-k)^j\\ m_{11}+\delta_{1k}\end{matrix}\right\rangle\left|\begin{matrix}[m-1]_3^i\\ [m-1]_2+\theta(2-k)^j\\ m_{11}+\delta_{1k}-1\end{matrix}\right\rangle. \tag{5.20}$$

In a similar way, taking into account that

$$e_{p0}e_{q0}=\left|\begin{matrix}-1,-1,-2\\ -1,-1-\theta(p+q-4)\\ -1-\theta(p+q-5)\end{matrix}\right\rangle,\quad p<q=1,2,3. \tag{5.21}$$

one obtains from (3.62) and (3.63)

$$\left|\begin{matrix}[m-1]_3^{-i}\\ [m-2]_2\\ m_{11}-1\end{matrix}\right\rangle=\sum_{p<q=1}^{3}e_{p0}e_{q0}\sum_{j=1}^{1+\theta(p+q-4)}\left\langle\begin{matrix}0,0,-1 & [m]_3\\ 0,-\theta(p+q-4);[m]_2+\theta(p+q-4)^j\\ -\theta(p+q-5) & m_{11}+\theta(p+q-5)\end{matrix}\middle|\begin{matrix}[m]_3^{-i}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle\otimes\left|\begin{matrix}[m]_3\\ [m]_2+\theta(p+q-4)^j\\ m_{11}+\theta(p+q-5)\end{matrix}\right\rangle_0, \tag{5.22}$$

$$e_{p0}e_{q0}\otimes\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle_0=\sum_{i=1}^{3}\sum_{j=1}^{1+\theta(p+q-4)}\left\langle\begin{matrix}0,0,-1 & [m]_3\\ 0,-\theta(p+q-4);[m]_2\\ -\theta(p+q-5) & m_{11}\end{matrix}\middle|\begin{matrix}[m]_3^{-i}\\ [m]_2-\theta(p+q-4)^j\\ m_{11}-\theta(p+q-5)\end{matrix}\right\rangle\left|\begin{matrix}[m-1]_3^{-i}\\ [m-1]_2-\theta(p+q-4)^j\\ m_{11}-\theta(p+q-5)-1\end{matrix}\right\rangle, \tag{5.23}$$

where $p<q=1,2,3$.

B. Transformation properties of the gl(3)-irreducible submodules under the odd generators

The representations of the even generators of sl(1,3) in the GZ basis are given with the relations (3.22). We now proceed to derive expressions for the odd generators. This task can be solved in different ways. One may use the circumstance that $\mathbf{e}^1=(e_{10},e_{20},e_{30})$ is a gl(3)-tensor operator [see T_2 in (5.7)]. So is the triple $\mathbf{e}^2=(e_{01},e_{02},e_{03})$ and it transforms as a [1,1,0] vector. Therefore, (Wigner–Eckart theorem) the matrix elements of the odd generators are products of the corresponding CGC's and the reduced matrix elements

$$\langle\Lambda_1|\mathbf{e}^k|\Lambda_2\rangle,\ \Lambda_1,\Lambda_2=[m]_3,[m-1]_3^{\pm i},\ [m-2]_3,\quad i=1,2,3,\quad k=1,2. \tag{5.24}$$

One may try to compute these 128 coefficients. Although this number can be reduced from symmetry considerations, it is still too big and in any case in the calculations one has to go back to the induced basis, i.e., to use the relations (5.19) and (5.20) and (5.22) and (5.23). Here we proceed in another way, which will lead directly to expressions for the odd generators in the GZ basis.

1. The subspace $V([m]_3)$

Consider the action of the positive root vectors e_{0k} on an arbitrary basis vector (5.16) from $V([m]_3)$. Since, by definition [see (4.3)],

$$e_{0k}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle_0=0,\quad k=1,2,3, \tag{5.25}$$

and, according to (4.5)

$$e_{0k}\left(1\otimes\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle_0\right)=1\otimes e_{0k}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle_0=0, \tag{5.26}$$

we conclude that the generators e_{0k} annihilate $V([m]_3)$:

$$e_{0k}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle=0,\quad k=1,2,3. \tag{5.27}$$

For the negative root vectors e_{k0} we must have from (5.16)

$$e_{k0}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = e_{k0}\left(\mathbf{1}\otimes\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle_0\right) = e_{k0}\otimes\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle_0 . \tag{5.28}$$

Therefore, the result is given with the right-hand side of the relation (5.20). Representing the CGC's according to (3.54) and inserting the explicit expressions (3.55)–(3.58) for the scalar factors, one obtains

$$e_{10}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{2} S(i,j)S(j,1)\left|\frac{\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11})(l_{k2}-l_{i3}-1)\Pi_{k\neq i=1}^{3}(l_{k3}-l_{j2})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)\Pi_{k\neq i=1}^{3}(l_{k3}-l_{i3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^i\\ [m-1]_2^j\\ m_{11}\end{matrix}\right\rangle, \tag{5.29}$$

$$e_{20}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \sum_{i=1}^{3}\sum_{j=1}^{2} S(i,j)\left|\frac{(l_{j2}-l_{11}+1)\Pi_{k\neq j=1}^{2}(l_{k2}-l_{i3}-1)\Pi_{k\neq i=1}^{3}(l_{k3}-l_{j2})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)\Pi_{k\neq i=1}^{3}(l_{k3}-l_{i3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^i\\ [m-1]_2^j\\ m_{11}-1\end{matrix}\right\rangle, \tag{5.30}$$

$$e_{30}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle = \sum_{i=1}^{3}\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{i3}-1)}{\Pi_{k\neq i=1}^{3}(l_{k3}-l_{i3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^i\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle. \tag{5.31}$$

The above relations show that $V([m]_3)$ is not an invariant subspace and in fact

$$e_{k0}V([m]_3)\subset\sum_{i=1}^{3}\oplus V([m-1]_3^i). \tag{5.32}$$

2. The subspaces $V([m-1]_3^i)$

Consider first the generator e_{30}. From (5.19) we have

$$e_{30}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle = -\sum_{k=1}^{2}\sum_{i=1}^{2}\left\langle\begin{matrix}1,0,0 & [m]_3\\ 1,0\ ; & [m]_2^{-i}\\ \delta_{1k} & m_{11}-\delta_{1k}\end{matrix}\right|\left.\begin{matrix}[m]_3^s\\ [m]_2\\ m_{11}\end{matrix}\right\rangle e_{k0}e_{30}\otimes\left|\begin{matrix}[m]_3\\ [m]_2^{-i}\\ m_{11}-\delta_{1k}\end{matrix}\right\rangle_0 . \tag{5.33}$$

Using (5.23) one derives that for $k=1,2$

$$e_{k0}e_{30}\otimes\left|\begin{matrix}[m]_3\\ [m]_2^{-i}\\ m_{11}-\delta_{1k}\end{matrix}\right\rangle_0 = \sum_{l=1}^{3}\sum_{j=1}^{2}\left\langle\begin{matrix}0,0,-1 & [m]_3\\ 0,-1\ ; & [m]_2^{-i}\\ -\delta_{2k} & m_{11}-\delta_{1k}\end{matrix}\right|\left.\begin{matrix}[m]_3^{-l}\\ [m]_2-1^i-1^j\\ m_{11}-1\end{matrix}\right\rangle\left|\begin{matrix}[m]_3^{-l}\\ [m-1]_2-1^i-1^j\\ m_{11}-2\end{matrix}\right\rangle. \tag{5.34}$$

Insert the last equation in (5.33):

$$e_{30}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle = -\sum_{k=1}^{2}\sum_{i=1}^{2}\sum_{l=1}^{3}\sum_{j=1}^{2}\left\langle\begin{matrix}1,0,0 & [m]_3\\ 1,0\ ; & [m]_2^{-i}\\ \delta_{1k} & m_{11}-\delta_{1k}\end{matrix}\right|\left.\begin{matrix}[m]_3^s\\ [m]_2\\ m_{11}\end{matrix}\right\rangle$$

$$\times\left\langle\begin{matrix}0,0,-1 & [m]_3\\ 0,-1\ ; & [m]_2^{-i}\\ -\delta_{2k} & m_{11}-\delta_{1k}\end{matrix}\right|\left.\begin{matrix}[m]_3^{-l}\\ [m]_2-1^i-1^j\\ m_{11}-1\end{matrix}\right\rangle\left|\begin{matrix}[m-1]_3^{-l}\\ [m-1]_2-1^i-1^j\\ m_{11}-2\end{matrix}\right\rangle. \tag{5.35}$$

The summation over i, j, k can be carried out (Appendix A) and gives

$$e_{30}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle = \sum_{l=1}^{3}\sum_{i=1}^{3}\epsilon_{sli}\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{i3}-1)}{\Pi_{k\neq i=1}^{3}(l_{k3}-l_{i3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{-l}\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle. \tag{5.36}$$

We recall that $l_{ij}=m_{ij}-i;\epsilon_{sli}$ is an antisymmetric tensor with $\epsilon_{123}=1$.

Using the commutation relation $e_{k0}=[e_{k3},e_{30}]$, one derives from (3.22) and (5.36)

$$e_{10}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle = \sum_{l=1}^{3}\sum_{i=1}^{3}\sum_{j=1}^{2}\epsilon_{sli}S(i,j)S(j,1)$$

$$\times\left|\frac{(l_{s3}-l_{j2}+1)(l_{l3}-l_{j2})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11})(l_{k2}-l_{i3}-1)}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)\Pi_{k\neq i=1}^{3}(l_{k3}-l_{i3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{-l}\\ [m-2]_2^j\\ m_{11}-1\end{matrix}\right\rangle. \tag{5.37}$$

$$e_{20}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle=\sum_{l=1}^{3}\sum_{i=1}^{3}\sum_{j=1}^{2}\epsilon_{sli}S(i,j)$$

$$\times\left|\frac{(l_{11}-l_{j2}-1)(l_{s3}-l_{j2}+1)(l_{l3}-l_{j2})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{i3}-1)}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)\Pi_{k\neq i=1}^{3}(l_{k3}-l_{i3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{-l}\\ [m-2]_2^j\\ m_{11}-2\end{matrix}\right\rangle. \quad (5.38)$$

We turn now to the operator e_{03}. Since $e_{03}\in P_+$, it annihilates $V_0([m]_3)$. Therefore, for any $|(m)_3\rangle\in V_0([m]_3)$,

$$e_{03}e_{k0}\otimes|(m)_3\rangle_0=\{e_{03},e_{k0}\}\otimes|(m)_3\rangle_0=\mathbb{1}\otimes E_{k3}|(m)_3\rangle_0. \quad (5.39)$$

From here and (5.19)

$$e_{03}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle=\sum_{k=1}^{3}\sum_{i=1}^{1+\theta(2-k)}\left\langle\begin{matrix}1,0,0 & [m]_3\\ 1-\delta_{3k},0;[m]_2-\theta(2-k)^i\\ \delta_{1k} & m_{11}-\delta_{1k}\end{matrix}\right|\left.\begin{matrix}[m]_3^s\\ [m]_2\\ m_{11}\end{matrix}\right\rangle\otimes E_{k3}\left|\begin{matrix}[m]_3\\ [m]_2-\theta(2-k)^i\\ m_{11}-\delta_{1k}\end{matrix}\right\rangle_0. \quad (5.40)$$

Write down $E_{k3}|(m)_3\rangle_0$, $k=1,2,3$ in the following compact form [see (3.22)]:

$$E_{k3}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle_0=\sum_{j=1}^{1+\theta(2-k)}\alpha_{k3}^j\begin{pmatrix}[m]_3\\ [m]_2\\ m_{11}\end{pmatrix}\left|\begin{matrix}[m]_3\\ [m]_2+\theta(2-k)^j\\ m_{11}+\delta_{1k}\end{matrix}\right\rangle_0, \quad (5.41)$$

where

$$\alpha_{13}^j\begin{pmatrix}[m]_3\\ [m]_2\\ m_{11}\end{pmatrix}=S(j,1)\left|\frac{\Pi_{k=1}^{3}(l_{k3}-l_{j2})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)}\right|^{1/2},$$

$$\alpha_{23}^j\begin{pmatrix}[m]_3\\ [m]_2\\ m_{11}\end{pmatrix}=\left|\frac{(l_{j2}-l_{11}+1)\Pi_{k=1}^{3}(l_{k3}-l_{j2})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)}\right|^{1/2}, \quad (5.42)$$

$$\alpha_{33}^1\begin{pmatrix}[m]_3\\ [m]_2\\ m_{11}\end{pmatrix}\equiv\alpha_{33}\begin{pmatrix}[m]_3\\ [m]_2\\ m_{11}\end{pmatrix}=l_{13}+l_{23}+l_{33}-l_{12}-l_{22}+3.$$

Then

$$e_{03}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle=\sum_{k=1}^{3}\sum_{i,j=1}^{1+\theta(2-k)}\left\langle\begin{matrix}1,0,0 & [m]_3\\ 1-\delta_{3k},0;[m]_2-\theta(2-k)^i\\ \delta_{1k} & m_{11}-\delta_{1k}\end{matrix}\right|\left.\begin{matrix}[m]_3^s\\ [m]_2\\ m_{11}\end{matrix}\right\rangle$$

$$\times\alpha_{k3}^j\begin{pmatrix}[m]_3\\ [m]_2-\theta(2-k)^i\\ m_{11}-\delta_{1k}\end{pmatrix}\left|\begin{matrix}[m]_3\\ [m]_2-\theta(2-k)^i+\theta(2-k)^j\\ m_{11}\end{matrix}\right\rangle. \quad (5.43)$$

The summation over i, j, k is carried out in Appendix B and gives

$$e_{03}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle=(l_{s3}+1)\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{s3}-1)}{\Pi_{k\neq s=1}^{3}(l_{k3}-l_{s3})}\right|^{1/2}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle. \quad (5.44)$$

The commutation relations

$$e_{0k}=[e_{03},e_{3k}],\quad k=1,2,3, \quad (5.45)$$

together with (3.22) and (5.44) yield

$$e_{01}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle=(l_{s3}+1)\sum_{j=1}^{2}S(s,j)S(j,1)\left|\frac{\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11}+1)(l_{k2}-l_{s3}-1)\Pi_{k\neq s=1}^{3}(l_{k3}-l_{j2}+1)}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq s=1}^{3}(l_{k3}-l_{s3})}\right|^{1/2}\left|\begin{matrix}[m]_3\\ [m]_2^{-j}\\ m_{11}-1\end{matrix}\right\rangle \quad (5.46)$$

$$e_{02}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle=(l_{s3}+1)\sum_{j=1}^{2}S(s,j)\left|\frac{(l_{j2}-l_{11})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{s3}-1)\Pi_{k\neq s=1}^{3}(l_{k3}-l_{j2}+1)}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq s=1}^{3}(l_{k3}-l_{s3})}\right|^{1/2}\left|\begin{matrix}[m]_3\\ [m]_2^{-j}\\ m_{11}\end{matrix}\right\rangle. \quad (5.47)$$

3. The subspaces $V([m-1]_3^{-i})$

Let $A_{pq} \in V_0([m]_3)$ and

$$|x\rangle = \sum_{p<q=1}^{3} e_{p0} e_{q0} \otimes A_{pq} \in \sum_{i=1}^{3} \oplus V([m-1]_3^{-i}). \tag{5.48}$$

Then

$$e_{k0} |x\rangle = \sum_{p<q=1}^{3} \epsilon_{kpq} e_{10} e_{20} e_{30} \otimes A_{pq} \in V([m-2]_3). \tag{5.49}$$

Applying the generator e_{k0} to the left- and the right-hand sides of Eq. (5.22) and using (5.49), one has

$$e_{k0} \left| \begin{matrix} [m-1]_3^{-i} \\ [m-1]_2 \\ m_{11}-1 \end{matrix} \right\rangle = \sum_{p<q=1}^{3} \epsilon_{kpq} e_{10} e_{20} e_{30}$$

$$\otimes \sum_{j=1}^{1+\theta(p+q-4)} \left\langle \begin{matrix} 0,0,-1 & [m]_3 \\ 0,-\theta(p+q-4); & [m]_2+\theta(p+q-4)^j \\ -\theta(p+q-5) & m_{11}+\theta(p+q-5) \end{matrix} \right| \left. \begin{matrix} [m]_3^{-i} \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle \left| \begin{matrix} [m]_3 \\ [m]_2+\theta(p+q-4)^j \\ m_{11}+\theta(p+q-5) \end{matrix} \right\rangle_0 . \tag{5.50}$$

Taking into account the defining relations (5.17) for the GZ basis in $V([m-2]_3)$ and inserting in (5.50) the explicit expressions for the CGC's, which follow from (3.64)–(3.68), we obtain

$$e_{10} \left| \begin{matrix} [m-1]_3^{-i} \\ [m-1]_2 \\ m_{11}-1 \end{matrix} \right\rangle = \sum_{j=1}^{2} S(i,j) S(j,1) \left| \frac{\Pi_{k\neq j=1}^2 (l_{k2}-l_{11})(l_{k2}-l_{i3}) \Pi_{k\neq i=1}^3 (l_{k3}-l_{j3})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2) \Pi_{k\neq i=1}^3 (l_{k3}-l_{i3})} \right|^{1/2} \left| \begin{matrix} [m-2]_3 \\ [m-2]_2^j \\ m_{11}-1 \end{matrix} \right\rangle, \tag{5.51}$$

$$e_{20} \left| \begin{matrix} [m-1]_3^{-i} \\ [m-1]_2 \\ m_{11}-1 \end{matrix} \right\rangle = \sum_{j=1}^{2} S(i,j) \left| \frac{(l_{j2}-l_{11}+1) \Pi_{k\neq j=1}^2 (l_{k2}-l_{i3}) \Pi_{k\neq i=1}^3 (l_{k3}-l_{j2})}{(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2) \Pi_{k\neq i=1}^3 (l_{k3}-l_{i3})} \right|^{1/2} \left| \begin{matrix} [m-2]_3 \\ [m-2]_2^j \\ m_{11}-2 \end{matrix} \right\rangle, \tag{5.52}$$

$$e_{30} \left| \begin{matrix} [m-1]_3^{-i} \\ [m-1]_2 \\ m_{11}-1 \end{matrix} \right\rangle = \left| \frac{\Pi_{k=1}^2 (l_{k2}-l_{i3})}{\Pi_{k\neq i=1}^3 (l_{k3}-l_{i3})} \right|^{1/2} \left| \begin{matrix} [m-2]_3 \\ [m-2]_2 \\ m_{11}-2 \end{matrix} \right\rangle. \tag{5.53}$$

In the derivation we have used the identity

$$\left| (l_{12}-l_{22}+\delta_{1j}-\delta_{2j}) \prod_{k\neq j=1}^{2} (l_{k2}-l_{j2}) \right|^{1/2} = |(l_{12}-l_{22}-j+1)(l_{12}-l_{22}-j+2)|^{1/2}. \tag{5.54}$$

Turn now to the generators e_{01}, e_{02}, e_{03}. The transformation properties of the vector (5.48) under e_{03} can be easily derived from the corresponding induced representation (4.14):

$$e_{03}|x> = \sum_{p\neq q=1}^{3} S(p,q) e_{p0} \otimes (\delta_{q3}-E_{q3}) A_{pq} \in \sum_{i=1}^{3} \oplus V([m-1]^i). \tag{5.55}$$

Considering as $|x>$ in the above relation the GZ-basis vector (5.22), we have

$$e_{03} \left| \begin{matrix} [m-1]_2^{-i} \\ [m-1]_2 \\ m_{11}-1 \end{matrix} \right\rangle = \sum_{p\neq q=1}^{3} S(p,q) e_{p0} \otimes \sum_{j=1}^{1+\theta(p+q-4)}$$

$$\times \left\langle \begin{matrix} 0,0,-1 & [m]_3 \\ 0,-\theta(p+q-4); & [m]_2+\theta(p+q-4)^j \\ -\theta(p+q-5) & m_{11}+\theta(p+q-5) \end{matrix} \right| \left. \begin{matrix} [m]_3^{-i} \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle (\delta_{q3}-E_{q3}) \left| \begin{matrix} [m]_3 \\ [m]_2+\theta(p+q-4)^j \\ m_{11}+\theta(p+q-5) \end{matrix} \right\rangle_0 , \tag{5.56}$$

which, in view of Eq. (5.41), gives

$$e_{03} \left| \begin{matrix} [m-1]_3^{-i} \\ [m-1]_2 \\ m_{11}-1 \end{matrix} \right\rangle = \sum_{p\neq q=1}^{3} \sum_{j=1}^{1+\theta(p+q-4)} \sum_{r=1}^{1+\theta(2-q)} S(p,q) \left[\delta_{q3} - \alpha_{q3}^r \left(\begin{matrix} [m]_3 \\ [m]_2+\theta(p+q-4)^j \\ m_{11}+\theta(p+q-5) \end{matrix} \right) \right]$$

$$\times \left\langle \begin{matrix} 0,0,-1 & [m]_3 \\ 0,-\theta(p+q-4); & [m]_2+\theta(p+q-4)^j \\ -\theta(p+q-5) & m_{11}+\theta(p+q-5) \end{matrix} \right| \left. \begin{matrix} [m]_3^{-i} \\ [m]_2 \\ m_{11} \end{matrix} \right\rangle e_{p0} \otimes \left| \begin{matrix} [m]_3 \\ [m]_2+\theta(p+q-4)^j+\theta(2-q)^r \\ m_{11}+\theta(p+q-5)+\delta_{1q} \end{matrix} \right\rangle_0 . \tag{5.57}$$

This relation gives the result in terms of the induced basis. In order to write it in the GZ basis, we use the inverse relation (5.20), which in the particular form we need it, reads

$$
e_{p0}\otimes\left|\begin{matrix}[m]_3\\ [m]_2+\theta(p+q-4)^j+\theta(2-q)^r\\ m_{11}+\theta(p+q-5)+\delta_{1q}\end{matrix}\right\rangle_0
=\sum_{s=1}^{3}\sum_{f=1}^{\theta(2-p)+1}\left\langle\begin{matrix}1,0,0 & [m]_3\\ 1-\delta_{3p},0; & [m]_2+\theta(p+q-4)^j+\theta(2-q)^r\\ \delta_{1p} & m_{11}+\theta(p+q-5)+\delta_{1q}\end{matrix}\right|\left.\begin{matrix}[m]_3^s\\ [m]_2+\theta(p+q-4)^j+\theta(2-q)^r+\theta(2-p)^f\\ m_{11}+\theta(p+q-5)+\delta_{1q}+\delta_{1p}\end{matrix}\right\rangle
$$

$$
\times\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2+\theta(p+q-4)^j+\theta(2-q)^r+\theta(2-p)^f\\ m_{11}+\theta(p+q-5)+\delta_{1q}+\delta_{1p}-1\end{matrix}\right\rangle. \tag{5.58}
$$

Inserting (5.58) in (5.57), we finally obtain

$$
e_{03}\left|\begin{matrix}[m-1]_3^{-i}\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle=\sum_{p\neq q=1}^{3}\sum_{s=1}^{3}\sum_{j=1}^{1+\theta(p+q-4)}\sum_{r=1}^{1+\theta(2-q)}\sum_{f=1}^{1+\theta(2-p)}S(p,q)\left[\delta_{q3}-\alpha_{q3}^r\begin{pmatrix}[m]_3\\ [m]_2+\theta(p+q-4)^j\\ m_{11}+\theta(p+q-5)\end{pmatrix}\right]
$$

$$
\times\left\langle\begin{matrix}0,0,-1 & [m]_3\\ 0,-\theta(p+q-4); & [m]_2+\theta(p+q-4)^j\\ -\theta(p+q-5) & m_{11}+\theta(p+q-5)\end{matrix}\right|\left.\begin{matrix}[m]_3^{-i}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle
$$

$$
\times\left\langle\begin{matrix}1,0,0 & [m]_3\\ 1-\delta_{3p},0; & [m]_2+\theta(p+q-4)^j+\theta(2-q)^r\\ \delta_{1p} & m_{11}+\theta(p+q-5)+\delta_{1q}\end{matrix}\right|\left.\begin{matrix}[m]_3^s\\ [m]_2+\theta(p+q-4)^j+\theta(2-q)^r+\theta(2-p)^f\\ m_{11}+\theta(p+q-5)+\delta_{1q}+\delta_{1p}\end{matrix}\right\rangle
$$

$$
\times\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2+\theta(p+q-4)^j+\theta(2-q)^r+\theta(2-p)^f\\ m_{11}+\theta(p+q-5)+\delta_{1q}+\delta_{1p}-1\end{matrix}\right\rangle. \tag{5.59}
$$

The summation in (5.59) gives (Appendix C)

$$
e_{03}\left|\begin{matrix}[m-1]_3^{-i}\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle=-\sum_{s=1}^{3}\sum_{j=1}^{3}(l_{j3}+1)\epsilon_{isj}\left|\frac{\Pi_{k=1}^3(l_{k2}-l_{j3})}{\Pi_{k\neq j=1}^3(l_{k3}-l_{j3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^s\\ [m]_2\\ m_{11}\end{matrix}\right\rangle. \tag{5.60}
$$

From this result and the commutation relations (5.45) one derives the representations of e_{01} and e_{02} in the GZ basis,

$$
e_{01}\left|\begin{matrix}[m-1]_3^{-i}\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle=-\sum_{s=1}^{3}\sum_{l=1}^{3}\sum_{j=1}^{2}\epsilon_{isl}(l_{l3}+1)S(l,j)S(j,1)
$$

$$
\times\left|\frac{(l_{s3}-l_{j2}+1)(l_{l3}-l_{j2})\Pi_{k\neq j=1}^2(l_{k2}-l_{11}+1)(l_{k2}-l_{l3})}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq l=1}^3(l_{k3}-l_{l3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^s\\ [m]_2^{-j}\\ m_{11}-1\end{matrix}\right\rangle, \tag{5.61}
$$

$$
e_{02}\left|\begin{matrix}[m-1]_3^{-i}\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle=-\sum_{s=1}^{3}\sum_{l=1}^{3}\sum_{j=1}^{2}\epsilon_{isl}(l_{l3}+1)S(l,j)
$$

$$
\times\left|\frac{(l_{s3}-l_{j2}+1)(l_{l3}-l_{j2})(l_{j2}-l_{11})\Pi_{k\neq j=1}^2(l_{k2}-l_{l3})}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq l=1}^3(l_{k3}-l_{l3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^s\\ [m]_2^{-j}\\ m_{11}\end{matrix}\right\rangle. \tag{5.62}
$$

4. The subspace $V([m-2]_3)$

From (5.17) one immediately concludes that $V([m-2]_3)$ is annihilated by the negative root vectors e_{k0}, $k=1,2,3$,

$$
e_{k0}\left|\begin{matrix}[m-2]_3\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle=0,\quad k=1,2,3. \tag{5.63}
$$

Consider e_{03}. The relations (5.17) and (4.14) yield

$$e_{03}\left|\begin{matrix}[m-2]_3\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle = \sum_{p<q=1}^{3}\sum_{r=1}^{3}\epsilon_{pqr}e_{p0}e_{q0}\otimes(E_{r3}-2\delta_{r3})\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle_0 . \tag{5.64}$$

Acting on $V_0([m]_3)$ with E_{r3} according to (5.41) and expressing afterwards the induced basis in terms of the GZ basis (5.23), we have

$$\begin{aligned} e_{30}\left|\begin{matrix}[m-2]_3\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle &= \sum_{l=1}^{3}\sum_{r=1}^{3}\sum_{p<q=1}^{3}\sum_{j=1}^{1+\theta(2-r)}\sum_{i=1}^{1+\theta(2-r)}\epsilon_{pqr}\left[\alpha_{r3}^{j}\begin{pmatrix}[m]_3\\ [m]_2\\ m_{11}\end{pmatrix}-2\delta_{r3}\right]\\ &\times\left\langle\begin{matrix}0,0,-1 & [m]_3\\ 0,-\theta(2-r); & [m]_2+\theta(2-r)^j\\ -\delta_{1r} & m_{11}+\delta_{1r}\end{matrix}\right|\left.\begin{matrix}[m]_3^{-l}\\ [m]_2+\theta(2-r)^j-\theta(2-r)^i\\ m_{11}\end{matrix}\right\rangle\\ &\times\left|\begin{matrix}[m-1]_3^{-l}\\ [m-1]_2+\theta(2-r)^j-\theta(2-r)^i\\ m_{11}-1\end{matrix}\right\rangle . \end{aligned} \tag{5.65}$$

The summation over p, q, r, j, i yields (Appendix D)

$$e_{03}\left|\begin{matrix}[m-2]_3\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle = \sum_{l=1}^{3}(l_{l3}+1)\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{l3})}{\Pi_{k\neq l=1}^{3}(l_{k3}-l_{l3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{-l}\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle . \tag{5.66}$$

This result together with the commutation relation $e_{k0}=[e_{k3},e_{30}]$ gives e_{01} and e_{02}:

$$\begin{aligned} e_{01}\left|\begin{matrix}[m-2]_3\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle &= \sum_{l=1}^{3}\sum_{j=1}^{2}(l_{l3}+1)S(l,j)S(j,1)\\ &\times\left|\frac{\Pi_{k\neq j=1}^{2}(l_{k2}-l_{11}+1)(l_{k2}-l_{l3})\Pi_{k\neq l=1}^{3}(l_{k3}-l_{j2}+1)}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq l=1}^{3}(l_{k3}-l_{l3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{-l}\\ [m-1]_2^{-j}\\ m_{11}-2\end{matrix}\right\rangle , \end{aligned} \tag{5.67}$$

$$e_{02}\left|\begin{matrix}[m-2]_3\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle = \sum_{l=1}^{3}\sum_{j=1}^{2}(l_{l3}+1)S(l,j)\left|\frac{(l_{j2}-l_{11})\Pi_{k\neq j=1}^{2}(l_{k2}-l_{l3})\Pi_{k\neq l=1}^{3}(l_{k3}-l_{j2}+1)}{(l_{12}-l_{22}+j-1)(l_{12}-l_{22}+j-2)\Pi_{k\neq l=1}^{3}(l_{k3}-l_{l3})}\right|^{1/2}\left|\begin{matrix}[m-1]_3^{-l}\\ [m-1]_2^{-j}\\ m_{11}-1\end{matrix}\right\rangle . \tag{5.68}$$

For any admissible triple $[m_{13}, m_{23}, m_{33}]$ the formulas (5.27), (5.29)–(5.31), (5.36)–(5.38), (5.44), (5.46), (5.47), (5.51)–(5.53), (5.60)–(5.62), and (5.66)–(5.68), derived in this section, together with the expressions (3.22) for the even generators define a representation of the Lie superalgebra sl(1,3) in an orthonormal Gel'fand–Zetlin basis. This representation is irreducible and hence typical if and only if (Proposition 3)

$$m_{13}\neq 0,\quad m_{23}\neq 1,\quad m_{33}\neq 2. \tag{5.69}$$

In this case m_{13}, m_{23}, m_{33} are the coordinates of the highest weight in the basis E^1, E^2, E^3, dual to (3.25).

If one of the conditions (5.69) is not fulfilled, the representation is indecomposable. The sl(1,3) module $\overline{V}([m]_3)$ contains a maximal invariant subspace $\overline{I}([m]_3)$, such that its orthogonal complement is not an invariant subspace. In this case the factor module $\overline{V}([m]_3)/\overline{I}([m]_3)$ carries an irreducible nontypical representation of sl(1,3). The representations of the generators for the nontypical case in a GZ basis will be given in Ref. 1.

ACKNOWLEDGMENTS

The author is thankful to Professor H. D. Doebner for the kind hospitality in the Institute for Theoretical Physics in Clausthal. The valuable discussions, permanent encouragement, and understanding in a sad (for me) period made it possible to complete the present investigation. I am thankful also to Professor S. J. Ališauskas for providing me with expressions free of mistakes for the Clebsch–Gordan coefficients.

APPENDIX A: DERIVATION OF EQ. (5.36)

In order to perform the summation in (5.35), we first observe that the sum of all terms on the right-hand side, corresponding to $i=j=1,2$, vanishes. This is a consequence of the Wigner–Eckart theorem and the decomposition (3.50), since (e_{10}, e_{20}, e_{30}) is a $(0,-1,-1)$-tensor operator under gl(3) [see T_2 in (5.7)]. One comes to the same conclusion considering the second-order Casimir operator C_2 of gl(2)$\subset$sl(1,3):

$$C_2=\tfrac{1}{2}[E_{12}E_{21}+E_{21}E_{12}+\tfrac{1}{2}(E_{11}-E_{22})^2]. \tag{A1}$$

This operator commutes with e_{30} and e_{03},

$$[C_2,e_{30}]=[C_2,e_{03}]=0. \tag{A2}$$

From (3.22) one derives

$$C_2e_{30}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle=\frac{1}{4}(m_{12}-m_{22})(m_{12}-m_{22}+2)\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle, \tag{A3}$$

$$C_2\left|\begin{matrix}[m-1]_3^{-l}\\ [m-1]_2-1^i-1^j\\ m_{11}-2\end{matrix}\right\rangle=\frac{1}{4}(m_{12}-m_{22}+\delta_{2i}+\delta_{2j}-\delta_{1i}-\delta_{1j})$$
$$\times(m_{12}-m_{22}+\delta_{2i}+\delta_{2j}-\delta_{1i}-\delta_{1j}+2)\left|\begin{matrix}[m-1]_3^{-l}\\ [m-1]_2-1^i-1^j\\ m_{11}-2\end{matrix}\right\rangle. \tag{A4}$$

Since both sides of (5.35) have to correspond to one and the same eigenvalue of C_2, one concludes that

$$\sum_{k=1}^{2}\left\langle\begin{matrix}1,0,0\ [m]_3\\ 1,0\ ;[m]_2^{-i}\\ \delta_{1k}\ \ m_{11}-\delta_{1k}\end{matrix}\middle|\begin{matrix}[m]_3^s\\ [m]_2\\ m_{11}\end{matrix}\right\rangle\left\langle\begin{matrix}0,0,-1\ [m]_3\\ 0,-1\ ;[m]_2^{-i}\\ -\delta_{2k}\ \ m_{11}-\delta_{1k}\end{matrix}\middle|\begin{matrix}[m]_3^{-l}\\ [m]_2-2^i\\ m_{11}-1\end{matrix}\right\rangle=0,\quad i=1,2. \tag{A5}$$

Therefore, (5.35) reduces to

$$e_{30}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle=-\sum_{l=1}^{3}\sum_{i=1}^{2}\sum_{k=1}^{2}\left\langle\begin{matrix}1,0,0\ [m]_3\\ 1,0\ ;[m]_2^{-i}\\ \delta_{1k}\ \ m_{11}-\delta_{1k}\end{matrix}\middle|\begin{matrix}[m]_3^s\\ [m]_2\\ m_{11}\end{matrix}\right\rangle\left\langle\begin{matrix}0,0,-1\ [m]_3\\ 0,-1\ ;[m]_2^{-i}\\ -\delta_{2k}\ \ m_{11}-\delta_{1k}\end{matrix}\middle|\begin{matrix}[m]_3^{-l}\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle\left|\begin{matrix}[m-1]_3^{-l}\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle. \tag{A6}$$

Replacing in (A6) the CGC's according to (3.54) and (3.64) as products of gl(3)- and gl(2)-scalar factors and taking into account the identity

$$\sum_{k=1}^{2}\begin{pmatrix}1,0\ [m]_2^{-i}\\ \delta_{1k};m_{11}-\delta_{1k}\end{pmatrix}\!\!\left|\begin{matrix}[m]_2\\ m_{11}\end{matrix}\right)\begin{pmatrix}0,-1\ [m]_2^{-i}\\ -\delta_{2k};m_{11}-\delta_{1k}\end{pmatrix}\!\!\left|\begin{matrix}[m-1]_2\\ m_{11}-1\end{matrix}\right)=(-1)^{i-1}, \tag{A7}$$

which follows from (3.57), (3.58), (3.67), and (3.68), we obtain

$$e_{30}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle=\sum_{l=1}^{3}\sum_{i=1}^{2}(-1)^i\left(\begin{matrix}1,0,0\ [m]_3\\ 1,0\ ;[m]_2^{-i}\end{matrix}\middle|\begin{matrix}[m]_3^s\\ [m]_2\end{matrix}\right)\left(\begin{matrix}0,0,-1\ [m]_3\\ 0,-1\ ;[m]_2^{-i}\end{matrix}\middle|\begin{matrix}[m]_3^{-l}\\ [m-1]_2\end{matrix}\right)\left|\begin{matrix}[m-1]_3^{-l}\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle. \tag{A8}$$

From (3.56) and (3.66) one derives that

$$\left(\begin{matrix}1,0,0\ [m]_3\\ 1,0\ ;[m]_3^{-i}\end{matrix}\middle|\begin{matrix}[m]_3^s\\ [m]_2\end{matrix}\right)=S(s,i)\left|\frac{\Pi_{k\neq i=1}^{2}(m_{k2}-m_{s3}-k+s-1)\Pi_{k\neq s=1}^{3}(m_{k3}-m_{i2}-k+i+1)}{\Pi_{k\neq i=1}^{2}(m_{k2}-m_{i2}-k+i)\Pi_{k\neq s=1}^{3}(m_{k3}-m_{s3}-k+s)}\right|^{1/2}, \tag{A9}$$

$$\left(\begin{matrix}0,0,-1\ [m]_3\\ 0,-1\ ;[m]_2^{-i}\end{matrix}\middle|\begin{matrix}[m]_3^{-l}\\ [m-1]_2\end{matrix}\right)=-S(l,3-i)\left|\frac{(m_{i2}-m_{l3}-i+l-1)\Pi_{k\neq l=1}^{3}(m_{k3}-\delta_{1i}m_{22}-\delta_{2i}m_{12}-k-i+4)}{(m_{12}-m_{22}+1)\Pi_{k\neq l=1}^{3}(m_{k3}-m_{l3}-k+l)}\right|^{1/2}. \tag{A10}$$

Inserting (A9) and (A10) in (A8) and writing down the sum over $i=1,2$, explicitly after some transformations we obtain

$$e_{30}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle=\sum_{l\neq s=1}^{3}\left|\frac{(m_{r3}-m_{12}-r+2)(m_{r3}-m_{22}-r+3)}{(m_{r3}-m_{s3}-r+s)(m_{r3}-m_{l3}-r+l)}\right|^{1/2}$$
$$\times\left\{\frac{S(s,1)S(l,2)|m_{s3}-m_{22}-s+3||m_{l3}-m_{12}-l+2|}{|m_{12}-m_{22}+1||m_{l3}-m_{s3}-l+s|}\right.$$
$$\left.-\frac{S(s,2)S(l,1)|m_{s3}-m_{12}-s+2||m_{l3}-m_{22}-l+3|}{|m_{12}-m_{22}+1||m_{l3}-m_{s3}-l+s|}\right\}\left|\begin{matrix}[m-1]_3^l\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle, \tag{A11}$$

where $r\neq s$ and $r\neq l$. It is a straightforward computation to show that the expression in the curly brackets is equal to ϵ_{slr}. Thus,

$$e_{30}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle = \sum_{l=1}^{3}\epsilon_{slr}\left|\frac{(m_{r3}-m_{12}-r+2)(m_{r3}-m_{22}-r+3)}{m_{r3}-m_{s3}-r+s)(m_{r3}-m_{l3}-r+l)}\right|^{1/2}\left|\begin{matrix}[m-1]_3^l\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle$$

$$= \sum_{l=1}^{3}\epsilon_{slr}\left|\frac{\Pi_{k=1}^{2}(m_{k2}-m_{r3}-k+r-1)}{\Pi_{k\neq r=1}^{3}(m_{k3}-m_{r3}-k+r)}\right|^{1/2}\left|\begin{matrix}[m-1]_3^l\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle. \tag{A12}$$

Writing (A12) in terms of $l_{ij}=m_{ij}-i$, one immediately obtains (5.36).

APPENDIX B: DERIVATION OF EQ. (5.44)

Consider the sum (5.43). The vectors

$$e_{03}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle,\quad \left|\begin{matrix}[m]_3\\ m_{12}-1,m_{22}+1\\ m_{11}\end{matrix}\right\rangle,\quad \left|\begin{matrix}[m]_3\\ m_{12}+1,m_{22}-1\\ m_{11}\end{matrix}\right\rangle \tag{B1}$$

are eigenvectors of C_2 (A1), corresponding to different eigenvalues. Therefore, the second and the third vectors in (B1) may enter the sum (5.43) only with zero coefficients. Hence, this sum reduces to

$$e_{30}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle = \left\{\sum_{k=1}^{2}\sum_{j=1}^{2}\left\langle\begin{matrix}1,0,0 & [m]_3\\ 1,0 & ;[m]_2^{-j}\\ \delta_{1k} & m_{11}-\delta_{1k}\end{matrix}\right|\left.\begin{matrix}[m]_3^s\\ [m]_2\\ m_{11}\end{matrix}\right\rangle\alpha_{k3}^{j}\begin{pmatrix}[m]_3\\ [m]_2^{-j}\\ m_{11}-\delta_{1k}\end{pmatrix}\right.$$

$$\left.+\left\langle\begin{matrix}1,0,0 & [m]_3\\ 0,0 & ;[m]_2\\ 0 & m_{11}\end{matrix}\right|\left.\begin{matrix}[m]_3^s\\ [m]_2\\ m_{11}\end{matrix}\right\rangle\alpha_{33}\begin{pmatrix}[m]_3\\ [m]_2\\ m_{11}\end{pmatrix}\right\}\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle. \tag{B2}$$

Insert in (B2) the expressions for the CGC's in terms of (3.54) and (3.55)–(3.58). After some transformations one obtains

$$e_{03}\left|\begin{matrix}[m-1]_3^s\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle = \left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{s3}-1)}{\Pi_{k\neq s=1}^{3}(l_{k3}-l_{s3})}\right|^{1/2}(l_{13}+l_{23}+l_{33}-l_{12}-l_{22}+3+\Delta_s)\left|\begin{matrix}[m]_3\\ [m]_2\\ m_{11}\end{matrix}\right\rangle, \tag{B3}$$

where

$$\Delta_s=\frac{S(s,1)|\Pi_{k\neq s=1}^{3}(l_{k3}-l_{12}+1)|+S(s,2)|\Pi_{k\neq s=1}^{3}(l_{k3}-l_{22}+1)|}{l_{12}-l_{22}}. \tag{B4}$$

According to (3.9), all $l_{ij}=m_{ij}-i$ have one and the same imaginary part. Hence, without loss of generality, we may assume that the l_{ij} in (B4) are real numbers. Consider Δ_1, i.e., the case with $s=1$. From (3.9) we know that the m_{ij} in the basis vector

$$\left|\begin{matrix}[m-1]_3^1\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle = \left|\begin{matrix}m_{13},m_{23}-1,m_{33}-1\\ m_{12}-1,m_{22}-1\\ m_{11}-1\end{matrix}\right\rangle \tag{B5}$$

obey the inequalities

$$m_{13}\geqslant m_{12}-1\geqslant m_{23}-1\geqslant m_{22}-1\geqslant m_{33}-1,$$

which in terms of $l_{ij}=m_{ij}-i$ reads

$$l_{13}+1\geqslant l_{12}\geqslant l_{23}+1\geqslant l_{22}+1\geqslant l_{33}+2.$$

Therefore,

$$\Delta_1=\frac{(l_{12}-l_{23}-1)(l_{12}-l_{33}-1)+(l_{23}-l_{22}+1)(l_{22}-l_{33}-1)}{l_{12}-l_{22}}=l_{12}+l_{22}-l_{33}-l_{23}-2.$$

In a similar way, considering $s=2,3$, one shows that

$$\Delta_s=l_{12}+l_{22}-l_{13}-l_{23}-l_{33}+l_{s3}-2. \tag{B6}$$

Inserting (B6) in (B3), one obtains (5.44).

APPENDIX C: DERIVATION OF EQ. (5.60)

Denote by $C[(m)_3,s,p,q,j,r,f]$ the coefficients in (5.59),

$$C[(m)_3,s,p,q,j,r,f]=S(p,q)\left[\delta_{q3}-\alpha_{q3}^{r}\begin{pmatrix}[m]_3\\ [m]_2+\theta(p+q-4)^j\\ m_{11}+\theta(p+q-5)\end{pmatrix}\right]\left\langle\begin{matrix}0,0,-1 & [m]_3\\ 0,-\theta(p+q-4);[m]_2+\theta(p+q-4)^j\\ -\theta(p+q-5) & m_{11}+\theta(p+q-5)\end{matrix}\right|\left.\begin{matrix}[m]_3^{-i}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle$$

$$\left|\times\left\langle\begin{matrix} 1,0,0 & [m]_3 \\ 1-\delta_{3p},0;[m]_2+\theta(p+q-4)^j+\theta(2-q)^r \\ \delta_{1p} & m_{11}+\theta(p+q-5)+\delta_{1q}\end{matrix}\right|\begin{matrix}[m]_3^s \\ [m]_2+\theta(p+q-4)^j+\theta(2-q)^r+\theta(2-p)^f \\ m_{11}+\theta(p+q-5)+\delta_{1q}+\delta_{1p}\end{matrix}\right\rangle . \tag{C1}$$

Then the sum (5.59) can be written in the following way:

$$e_{03}\left|\begin{matrix}[m-1]_3^{-i} \\ [m-1]_2 \\ m_{11}-1\end{matrix}\right\rangle=\sum_{s=1}^{3}\left\{\sum_{\substack{p+q=3\\ j=r=f=1}}+\sum_{\substack{p=1,2\\ q=3\\ j=r=f=1}}+\sum_{\substack{q=1,2\\ p=3\\ j=r=f=1}}\right\}C\,[(m)_3,s,p,q,j,r,f\,]\left|\begin{matrix}[m-1]_3^s \\ m_{12}+1,m_{22}-1 \\ m_{11}\end{matrix}\right\rangle$$

$$+\sum_{s=1}^{3}\left\{\sum_{\substack{p+q=3\\ r=f=2\\ j=1}}+\sum_{\substack{p=1,2\\ q=3\\ f=j=2\\ r=1}}+\sum_{\substack{q=1,2\\ p=3\\ j=r=2\\ f=1}}\right\}C\,[(m)_3,s,p,q,j,r,f\,]\left|\begin{matrix}[m]_3^s \\ m_{12}-1,m_{22}+1 \\ m_{11}\end{matrix}\right\rangle$$

$$+\sum_{s=1}^{3}\left\{\sum_{\substack{p+q=3\\ r\neq f=1,2\\ j=1}}+\sum_{\substack{p=1,2\\ f\neq j=1,2\\ q=3\\ r=1}}+\sum_{\substack{q=1,2\\ r\neq j=1,2\\ p=3\\ f=1}}\right\}C\,[(m)_3,s,p,q,j,r,f\,]\left|\begin{matrix}[m-1]_3^s \\ [m]_2 \\ m_{11}\end{matrix}\right\rangle . \tag{C2}$$

Applying once more the relations (A2) and the considerations of Appendix A, we conclude that the coefficients in front of the vectors

$$\left|\begin{matrix}[m-1]_3^s \\ m_{12}+1,m_{22}-1 \\ m_{11}\end{matrix}\right\rangle \quad\text{and}\quad \left|\begin{matrix}[m-1]_3^s \\ m_{12}-1,m_{22}+1 \\ m_{11}\end{matrix}\right\rangle$$

in (C2) vanish. Represent the rest of the sum as

$$e_{03}\left|\begin{matrix}[m-1]_3^{-i} \\ [m-1]_2 \\ m_{11}-1\end{matrix}\right\rangle=\sum_{s=1}^{3}\left|\begin{matrix}[m-1]_3^s \\ [m]_2 \\ m_{11}\end{matrix}\right\rangle\left\{\sum_{\substack{p=1,2\\ f\neq j=1,2}}C\,[(m)_3,s,p,3,j,1,f\,]+\sum_{\substack{p\neq q=1,2\\ r\neq f=1,2}}C\,[(m)_3,s,p,q,1,r,f\,]\right.$$

$$\left.+\sum_{\substack{q=1,2\\ r\neq j=1,2}}C\,[(m)_3,s,3,q,j,r,1]\right\}\equiv\sum_{s=1}^{3}\left|\begin{matrix}[m-1]_3^s \\ [m]_2 \\ m_{11}\end{matrix}\right\rangle\{\Sigma_1+\Sigma_2+\Sigma_3\}\,. \tag{C3}$$

To compute

$$\Sigma_1=\sum_{\substack{p=1,2\\ f\neq j=1,2}}C\,[(m)_3,s,p,3,j,1,f\,], \tag{C4}$$

note that according to (5.42)

$$1-\alpha_{33}\begin{pmatrix}[m]_3 \\ [m]_2^j \\ m_{11}+\delta_{1k}\end{pmatrix}=l_{12}+l_{22}-l_{13}-l_{23}-l_{33}-1. \tag{C5}$$

for any $j,k=1,2$. Therefore,

$$\Sigma_1=(1+l_{12}+l_{22}-l_{13}-l_{23}-l_{33})\sum_{i=1}^{2}\sum_{k=1}^{2}\left\langle\begin{matrix}1,0,0\ [m]_3 \\ 1,0\ ;[m+1]_2^{-i} \\ \delta_{1k}\ \ m_{11}+1-\delta_{1k}\end{matrix}\right|\left.\begin{matrix}[m]_3^s \\ [m+1]_2 \\ m_{11}+1\end{matrix}\right\rangle\left\langle\begin{matrix}0,0,-1\ [m]_3 \\ 0,-1\ ;[m+1]_2^{-i} \\ -\delta_{2k}\ \ m_{11}+1-\delta_{1k}\end{matrix}\right|\left.\begin{matrix}[m]_3^{-i} \\ [m]_2 \\ m_{11}\end{matrix}\right\rangle . \tag{C6}$$

This sum has already been carried out in Appendix A, relation (A6). The result is

$$\Sigma_1=(l_{13}+l_{23}+l_{33}-l_{12}-l_{22}-1)\sum_{j=1}^{3}\epsilon_{sij}\left|\frac{\Pi_{k=1}^{2}(l_{k2}-l_{j3})}{\Pi_{k\neq j=1}^{3}(l_{k3}-l_{j3})}\right|^{1/2}. \tag{C7}$$

Inserting in $\Sigma_2+\Sigma_3$ the expressions for the CGC's (3.54)–(3.58) and (3.64)–(3.68), one obtains, after some transformations,

$$\Sigma_2+\Sigma_3=\frac{|\Pi_{k=1}^2(l_{k2}-l_{j3})|^{1/2}}{(l_{12}-l_{22})|\Pi_{k\neq i=1}^3(l_{k3}-l_{i3})\Pi_{k\neq s=1}^3(l_{k3}-l_{s3})|^{1/2}}$$
$$\times\{S(s,2)|(l_{s3}-l_{12})(l_{i3}-l_{12})(l_{i3}-l_{22})|-S(i,2)|(l_{i3}-l_{12})(l_{s3}-l_{12})(l_{s3}-l_{22})|$$
$$+S(i,1)|(l_{i3}-l_{22})(l_{s3}-l_{12})(l_{s3}-l_{22})|-S(s,1)|(l_{s3}-l_{22})(l_{i3}-l_{12})(l_{i3}-l_{22})|\}. \tag{C8}$$

Observe that $\Sigma_1+\Sigma_2+\Sigma_3$ is antisymmetric with respect to i and s. Therefore, the sum in (C3) is over all $s\neq i$. Summing together (C7) and (C8) for $i=1, s=2$; $i=2, s=3$; $i=3, s=1$, and taking into account the "betweenness" condition (3.12), we have

$$\Sigma_1+\Sigma_2+\Sigma_3|_{i=1,\,s=2}=-(l_{33}+1)\left|\frac{(l_{12}-l_{33})(l_{22}-l_{33})}{(l_{13}-l_{33})(l_{23}-l_{33})}\right|^{1/2}, \tag{C9}$$

$$\Sigma_1+\Sigma_2+\Sigma_3|_{i=2,\,s=3}=-(l_{13}+1)\left|\frac{(l_{12}-l_{13})(l_{22}-l_{13})}{(l_{23}-l_{13})(l_{33}-l_{13})}\right|^{1/2}, \tag{C10}$$

$$\Sigma_1+\Sigma_2+\Sigma_3|_{i=3,\,s=1}=-(l_{23}+1)\left|\frac{(l_{12}-l_{23})(l_{22}-l_{23})}{(l_{13}-l_{23})(l_{33}-l_{23})}\right|^{1/2}. \tag{C11}$$

In a unified form all three relations (C9)–(C11) read

$$\Sigma_1+\Sigma_2+\Sigma_3=-\sum_{j=1}^{3}\epsilon_{isj}(l_{j3}+1)\left|\frac{\Pi_{k=1}^2(l_{k2}-l_{j3})}{\Pi_{k\neq j=1}^3(l_{k3}-l_{j3})}\right|^{1/2}. \tag{C12}$$

Inserting (C12) in (C3), one obtains (5.60).

APPENDIX D: DERIVATION OF EQ. (5.66)

Consider the sum (5.65). As in Appendix A, we conclude that the vectors

$$\left|\begin{matrix}[m-1]_3^{-l}\\ m_{12},m_{22}-2\\ m_{11}-1\end{matrix}\right\rangle \quad\text{and}\quad \left|\begin{matrix}[m-1]_3^{-l}\\ m_{12}-2,m_{22}\\ m_{11}-1\end{matrix}\right\rangle$$

enter the sum with zero coefficients. The nonzero terms yield

$$e_{30}\left|\begin{matrix}[m-2]_3\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle=\sum_{l=1}^{3}\sum_{p<q=1}^{3}\sum_{r=1}^{3}\sum_{j=1}^{1+\theta(2-r)}\epsilon_{pqr}\left[\alpha_{r3}^{j}\begin{pmatrix}[m]_3\\ [m]_2\\ m_{11}\end{pmatrix}-2\delta_{r3}\right]$$
$$\times\left\langle\begin{matrix}0,0,-1 & [m]_3\\ 0,-\theta(2-r);[m]_2+\theta(2-r)^j\\ -\delta_{1r} & m_{11}+\delta_{1r}\end{matrix}\left|\begin{matrix}[m]_3^{-l}\\ [m]_2\\ m_{11}\end{matrix}\right\rangle\right|\left.\begin{matrix}[m-1]_3^{-l}\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle. \tag{D1}$$

Inserting here the expressions from (5.42) and the CGC's (3.64)–(3.68), after some calculations we obtain

$$e_{30}\left|\begin{matrix}[m-2]_3\\ [m-2]_2\\ m_{11}-2\end{matrix}\right\rangle=\sum_{l=1}^{3}\left|\frac{\Pi_{k=1}^2(l_{k2}-l_{l3})}{\Pi_{k\neq l=1}^3(l_{k3}-l_{l3})}\right|^{1/2}(l_{13}+l_{23}+l_{33}-l_{12}-l_{22}+1+\Delta_l)\left|\begin{matrix}[m-1]_3^{-l}\\ [m-1]_2\\ m_{11}-1\end{matrix}\right\rangle, \tag{D2}$$

where

$$\Delta_l=\frac{S(l,1)|\Pi_{k\neq l=1}^3(l_{k3}-l_{12})|+S(l,2)|\Pi_{k\neq l=1}^3(l_{k3}-l_{22})|}{l_{12}-l_{22}}. \tag{D3}$$

Using the "betweenness" condition (3.12) on easily derives from (D3)

$$\Delta_l=l_{12}+l_{22}+l_{l3}-l_{13}-l_{23}-l_{33}. \tag{D4}$$

Inserting (D4) in (D2), one derives Eq. (5.66).

[1]T. D. Palev, "Finite-dimensional representations of the Lie superalgebra sl(1,3) in a Gel'fand–Zetlin basis. II. Nontypical representations" (in preparation).

[2]V. G. Kac, Adv. Math. **26**, 8 (1977); Commun. Math. Phys. **53**, 31 (1977).

[3]V. G. Kac, Lecture Notes Math. **676**, 597 (1978).

[4]See also L. Ross, Trans. Am. Math. Soc. **120**, 17 (1965); G. Hochschild, Ill. J. Math. **20**, 107 (1976); D. Z. Djokovic and G. Hochschild, Ill. J. Math. **20**, 134 (1976); D. Z. Djokovic, J. Pure Appl. Alg. **9**, 25 (1976); F. A. Berezin, J. Nucl. Phys. **29**, 857 (1979), **30**, 605 (1979); T. D. Palev, J. Math. Phys. **21**, 1293 (1980); J.-P. Hurni and B. Morel, J. Math. Phys. **23**, 2236 (1982); **24**, 157 (1983).

[5]A. Pais and V. Rittenberg, J. Math. Phys. **16**, 2062 (1975); M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146, 155 (1977); M. Bednar and V. F. Šachl, J. Math. Phys. **20**, 367 (1979).

[6]T. D. Palev, J. Math. Phys. **22**, 2127 (1981), **21**, 1293 (1980); T. D. Palev and O. Ts. Stoytchev, Preprint JINR E5-82-54, 1982.

[7]M. Marcu, J. Math. Phys. **21**, 1277 (1980).
[8]I. M. Gel'fand and M. L. Zetlin, Doklay Akad. Nauk SSSR **71**, 825 (1950).
[9]We follow some terminology and notation, accepted in J. D. Louck, Am. J. Phys. **38**, 18 (1970).
[10]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1931), pp. 383, 391; see also Ref. 9, p. 10.
[11]A very useful prescription for decomposition of tensor representations into a direct sum of irreducible representations can be found in Ref. 9, p. 18.
[12]See, for instance, A. U. Klymik, *Matrix Elements of the Clebsch–Gordan Coefficients of Representations of Groups* (Naukova Dumka, Kiev, 1979), pp. 67, 163 (in Russian).
[13]G. Baird and L. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[14]T. D. Palev and O. Ts. Stoytchev, C. R. Acad. Bulg. Sci. **35**, 733 (1982).

Scattering theory for long-range systems at threshold

D. Bollé[a),b)] and F. Gesztesy[c),d)]
Zentrum für interdisziplinäre Forschung, Universität Bielefeld, D-4800 Bielefeld, Federal Republic of Germany

W. Schweiger
Institut für Theoretische Physik, Universität Graz, A-8010 Graz, Austria

(Received 12 December 1984; accepted for publication 15 February 1985)

Scattering theory is discussed for long-range systems around threshold, taking into account explicitly the possibility of zero-energy eigenstates of the underlying Schrödinger Hamiltonian. Low-energy expansions are derived for the scattering amplitude in the different cases. The Coulomb-modified scattering length and effective range parameter are defined directly in terms of the scattering amplitude. Explicit formulas for these quantities are provided. No spherical symmetry of the interaction is assumed.

I. INTRODUCTION

This paper continues the study, started in Ref. 1, of low-energy quantities like the scattering length and effective range parameter, in terms of the threshold behavior of the scattering amplitude, for Schrödinger Hamiltonians with general, nonspherically symmetric interactions. In Ref. 1, the short-range case has been treated extensively, allowing zero-energy resonances and zero-energy bound states of the Hamiltonian. However, for Coulomb plus short-range interactions, only a brief illustration has been given of how the short-range discussion can be extended, by presenting a definition and explicit formula for the Coulomb-modified scattering length in the generic case (i.e., the case without zero-energy resonances and zero-energy bound states).

The aim of this paper is to provide a complete treatment of the Coulomb-modified low-energy quantities for nonspherically symmetric interactions (including the possibility of zero-energy eigenstates). Such an extensive study is appropriate since, as is well established by now, these low-energy parameters play a fundamental role in the discussion of many (low-energy) phenomena in different branches of physics. For earlier discussions in this connection we refer to Refs. 2 and 3, to the Introduction of Ref. 4, and more recently to Refs. 5 and 6 for the definition and use of these parameters in two- (and n-) dimensional spherically symmetric systems.

It turns out that the details of the study presented here are very different from the short-range treatment.[1] In particular, as we will describe below, the threshold behavior (and consequently the case distinctions) of the long-range system is totally distinct from its short-range counterpart. Furthermore, in the long-range cases we obtain low-energy expansions, e.g., for the scattering amplitude, which are asymptotic in nature, whereas in the short-range cases analytic expansions have been derived.[1]

Let us finally describe the main results of our paper. In Sec. II we derive the low-energy asymptotic expansions of the pure Coulomb Green's function and wave function, using the results of Ref. 7 and Ref. 8, respectively. Section III discusses in detail the threshold behavior of repulsive and attractive Coulomb plus short-range (nonspherically symmetric) systems. This behavior is very different from the one in pure short-range systems.[1] Indeed, besides the generic case, the repulsive Coulomb system shows the phenomenon of getting automatically a threshold *bound state* when the short-range part of the potential has a critical coupling strength. For attractive Coulomb systems we discuss, again besides the generic case, the possibility of having a threshold resonance. Threshold bound states are excluded in this case.

In Sec. IV we expand the transition amplitude and scattering operator near threshold in all relevant cases. The low-energy expansions we obtain are asymptotic expansions in k^2, in contrast with the short-range cases, where analytic expansions in k are possible.[1]

In Sec. V we find the appropriate generalization of the definition of the Coulomb-modified scattering length and effective range parameter for nonspherically symmetric interactions in all relevant cases. This definition is directly related to the threshold behavior of the scattering amplitude. We also give and discuss explicit formulas for these low-energy quantities in terms of the interactions. Finally, the Appendix briefly sketches a proof of some technical results on the exceptional set of positive energy eigenstates, needed in the low-energy expansions of Sec. IV.

II. LOW-ENERGY EXPANSION OF PURE COULOMB QUANTITIES

Let the potential V be a real-valued measurable function belonging to the Rollnik class R, i.e.,

$$\int_{\mathbb{R}^6} d^3x\, d^3y\, |V(\mathbf{x})||\mathbf{x}-\mathbf{y}|^{-2}|V(\mathbf{y})| < \infty. \tag{2.1}$$

As in Ref. 1 we define the short-range Hamiltonian H_s in $L^2(\mathbb{R}^3)$ by the method of quadratic forms[9]

$$H_s = H_0 + V, \tag{2.2}$$

where H_0 denotes the kinetic energy operator

$$H_0 = -\Delta, \quad \text{on } \mathscr{D}(H_0) = H^{2,2}(\mathbb{R}^3), \tag{2.3}$$

[a)] Onderzoeksleider Nationaal Fonds voor Wetenschappelijk Onderzoek (N.F.W.O.), Belgium.

[b)] On leave of absence from Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium.

[c)] Alexander von Humboldt Research Fellow.

[d)] On leave of absence from Institut für Theoretische Physik, Universität Graz, Austria.

 0022-2488/85/071661-14$02.50

with resolvent

$$G_k = (H_0 - k^2)^{-1}, \quad \mathrm{Im}\, k > 0. \tag{2.4}$$

The Coulomb Hamiltonian H_c is defined by

$$H_c = H_0 + \gamma|\mathbf{x}|^{-1}, \quad \text{on } \mathscr{D}(H_c) = \mathscr{D}(H_0), \quad \gamma \in \mathbb{R}, \tag{2.5}$$

with

$$G_{\gamma,k} = (H_c - k^2)^{-1}, \quad \mathrm{Im}\, k > 0, \quad k \neq -i\gamma/2n, \quad n = 1,2,\ldots, \tag{2.6}$$

the Coulomb resolvent. Our general long-range Hamiltonian H is finally given as the form sum

$$H = H_c + V. \tag{2.7}$$

In the following, a splitting of the short-range interaction V according to

$$V = u \cdot v, \quad v(\mathbf{x}) = |V(\mathbf{x})|^{1/2}, \quad u(\mathbf{x}) = v(\mathbf{x}) \mathrm{sgn}\, V(\mathbf{x}) \tag{2.8}$$

turns out to be useful.

We first discuss the kernel of the Coulomb resolvent, $G_{\gamma,k}(\mathbf{x},\mathbf{y})$.

Lemma 2.1: (a) Let $\mathrm{Im}\, k > 0$, $k \neq -i\gamma/2n$, $n = 1,2\ldots$; then for all $\gamma \in \mathbb{R}$, $G_{\gamma,k}(\mathbf{x},\mathbf{y})$ is a Carleman kernel. In particular, $G_{\gamma,k}(\mathbf{x},\cdot) \in L^1(\mathbb{R}^3) \cap L^2(\mathbb{R}^3)$.

(b) Assume that $k \in \Pi_+ = \{z \in \mathbb{C} |\, -\pi/2 < \arg z < 3\pi/2\}$ for $\gamma > 0$ and $k \in \Pi_- = \{z \in \mathbb{C} |\, -3\pi/2 < \arg z < \pi/2\}$ for $\gamma < 0$. Then for all $\mathbf{x} \neq \mathbf{y}$ the function $G_{\gamma,k}(\mathbf{x}, \mathbf{y})$ has the following asymptotic expansion with respect to k^2 near $k^2 = 0$;

$$G_{\gamma,k}(\mathbf{x}, \mathbf{y}) \underset{k\to 0}{\sim} \sum_{n=0}^{\infty} (ik)^{2n} g_{\gamma,2n}(\mathbf{x}, \mathbf{y}),$$

$$\gamma \in \mathbb{R} \setminus \{0\}, \quad \mathbf{x} \neq \mathbf{y}, \quad k \in \Pi_\pm \quad \text{for } \gamma \gtrless 0. \tag{2.9}$$

The first two coefficients are explicitly given by

$$g_{\gamma,0}(\mathbf{x}, \mathbf{y}) = G_{\gamma,0}(\mathbf{x}, \mathbf{y})$$

$$= (4\pi|\mathbf{x}-\mathbf{y}|)^{-1} \begin{cases} [(2\gamma x_-)^{1/2} I_1((2\gamma x_-)^{1/2}) K_0((2\gamma x_+)^{1/2}) + (2\gamma x_+)^{1/2} I_0((2\gamma x_-)^{1/2}) K_1((2\gamma x_+)^{1/2})], \quad \gamma > 0, & (2.10) \\ -(i\pi/2)[(2|\gamma| x_-)^{1/2} J_1((2|\gamma| x_-)^{1/2}) H_0^{(1)}((2|\gamma| x_+)^{1/2}) - (2|\gamma| x_+)^{1/2} J_0((2|\gamma| x_-)^{1/2}) \\ \times H_1^{(1)}((2|\gamma| x_+)^{1/2})], \quad \gamma < 0 & (2.11) \end{cases}$$

[we note that $g_{0,0}(\mathbf{x}, \mathbf{y}) = (4\pi|\mathbf{x}-\mathbf{y}|)^{-1}$] and

$$g_{\gamma,2}(\mathbf{x}, \mathbf{y}) = -[2(|\mathbf{x}| + |\mathbf{y}|) I_0((2\gamma x_-)^{1/2}) K_0((2\gamma x_+)^{1/2}) + (x_+ x_-)^{1/2} I_1((2\gamma x_-)^{1/2}) K_1((2\gamma x_+)^{1/2})$$
$$+ (2x_+/\gamma)^{1/2} I_0((2\gamma x_-)^{1/2}) K_1((2\gamma x_+)^{1/2}) - (2x_-/\gamma)^{1/2} I_1((2\gamma x_-)^{1/2}) K_0((2\gamma x_+)^{1/2})]/12\pi, \quad \gamma > 0, \tag{2.12}$$

$$g_{\gamma,2}(\mathbf{x}, \mathbf{y}) = -i[2(|\mathbf{x}| + |\mathbf{y}|) J_0((2|\gamma| x_-)^{1/2}) H_0((2|\gamma| x_+)^{1/2}) + (x_+ x_-)^{1/2} J_1((2|\gamma| x_-)^{1/2}) H_1^{(1)}((2|\gamma| x_+)^{1/2})$$
$$- (2x_+/|\gamma|)^{1/2} J_0((2|\gamma| x_-)^{1/2}) H_1^{(1)}((2|\gamma| x_+)^{1/2}) - (2x_-/|\gamma|)^{1/2} J_1((2|\gamma| x_-)^{1/2}) H_0^{(1)}((2|\gamma| x_+)^{1/2})]/24, \quad \gamma < 0. \tag{2.13}$$

Here $J_\nu(z)(I_\nu(z))$, etc., denote the (modified) Bessel functions of order ν (Ref. 10) and

$$x_\pm = |\mathbf{x}| + |\mathbf{y}| \pm |\mathbf{x} - \mathbf{y}|. \tag{2.14}$$

Proof: Part (a) is a result of Ref. 11 [cf. estimate (4.1)]. To prove part (b) we first recall the explicitly known expression for $G_{\gamma,k}(\mathbf{x}, \mathbf{y})$ (Ref. 8)

$$G_{\gamma,k}(\mathbf{x}, \mathbf{y}) = (4\pi|\mathbf{x}-\mathbf{y}|)^{-1}\{2(x_+ - x_-)(x_+ x_-)^{-1} F_0^{(0)}(k,x_-/2) G_0^{(0)}(-k,x_+/2)$$
$$- 3^{-1}(k^2 + \gamma^2/4) F_1^{(0)}(k,x_-/2) G_0^{(0)}(-k,x_+/2) + 3F_0^{(0)}(k,x_-/2) G_1^{(0)}(-k,x_+/2)\}, \quad \mathbf{x} \neq \mathbf{y}, \tag{2.15}$$

where[12]

$$F_l^{(0)}(k,r) = r^{l+1} e^{ikr} {}_1F_1(l+1+i\gamma/2k; 2l+2; -2ikr), \quad l = 0,1,\ldots, \tag{2.16}$$

is real for $k \in \mathbb{R}$ and entire with respect to $k^2 \in \mathbb{C}$ and

$$G_l^{(0)}(-k,r) = \Gamma(2l+2)^{-1} \Gamma(l+1+i\gamma/2k)(2ie^{-i\pi}k)^{2l+1} r^{l+1} e^{ikr} U(l+1+i\gamma/2k; 2l+2; 2ie^{-i\pi}kr), \quad l = 0,1,\ldots. \tag{2.17}$$

Here ${}_1F_1(\alpha;\beta;z)$ $(U(\alpha;\beta;z))$ denotes the (ir)regular confluent hypergeometric function.[10] Introducing Lambert's irregular Coulomb wave function $\widetilde{G}_l^{(0)}(k,r)$ (Ref. 7)

$$\widetilde{G}_l^{(0)}(k,r) = G_l^{(0)}(-k,r) - 2^{2l}\Gamma(2l+2)^{-2}|\Gamma(1+i\gamma/2k)|^{-2}|\Gamma(l+1+i\gamma/2k)|^2 \gamma k^{2l} h(\gamma,k) F_l^{(0)}(k,r), \quad l = 0,1,\ldots, \tag{2.18}$$

where

$$h(\gamma,k) = \Psi(i\gamma/2k) - \ln(i|\gamma|/2k) - ik/\gamma \tag{2.19}$$

[$\Psi(z)$ is the digamma function[10]], one infers that $\widetilde{G}_l^{(0)}(k,r)$ (like $F_l^{(0)}$) is real for $k \in \mathbb{R}$ and entire with respect to $k^2 \in \mathbb{C}$. Insertion of Eq. (2.18) into (2.15) finally yields

$$G_{\gamma,k}(\mathbf{x},\mathbf{y}) = (4\pi|\mathbf{x}-\mathbf{y}|)^{-1}\{2(x_+ - x_-)(x_+x_-)^{-1}F_0^{(0)}(k,x_-/2)\widetilde{G}_0^{(0)}(k,x_+/2)$$
$$- 3^{-1}(k^2+\gamma^2/4)F_1^{(0)}(k,x_-/2)\widetilde{G}_0^{(0)}(k,x_+/2) + 3F_0^{(0)}(k,x_-/2)\widetilde{G}_1^{(0)}(k,x_+/2)\}$$
$$+ (4\pi|\mathbf{x}-\mathbf{y}|)^{-1}\gamma h(\gamma,k)\{2(x_+ - x_-)(x_+x_-)^{-1}F_0^{(0)}(k,x_-/2)F_0^{(0)}(k,x_+/2)$$
$$- 3^{-1}(k^2+\gamma^2/4)[F_1^{(0)}(k,x_-/2)F_0^{(0)}(k,x_+/2) - F_0^{(0)}(k,x_-/2)F_1^{(0)}(k,x_+/2)]\}$$
$$\equiv G_{\gamma,k}^{(0)}(\mathbf{x},\mathbf{y}) + \gamma h(\gamma,k)G_{\gamma,k}^{(1)}(\mathbf{x},\mathbf{y}), \quad \mathbf{x}\neq\mathbf{y}. \tag{2.20}$$

Using the asymptotic expansion[10]

$$h(\gamma,k) \underset{k\to 0}{\sim} -\sum_{n=1}^{\infty} B_{2n}\frac{(2k/i\gamma)^{2n}}{2n} + \begin{cases}0, & \gamma>0,\\ -i\pi, & \gamma<0\end{cases} \tag{2.21}$$

(B_{2n} = the Bernoulli numbers), which is valid near $k=0$ for $k\in\Pi_\pm$, $\gamma\gtrless 0$, and the fact that the curly brackets in Eq. (2.20) are entire with respect to $k^2\in\mathbb{C}$, we prove expansion (2.9). Equations (2.10)–(2.13) give the first coefficients in this expansion. □

Remark 2.2: (a) The computation of these coefficients (2.10)–(2.13) is straightforward but long and tedious, such that we have omitted further details here. Equations (2.10) and (2.11) correct an error in formulas (4.20) of Ref. 1 and (2.7) of Ref. 13.

(b) We have excluded the short-range case $\gamma=0$ in Lemma 2.1(b). This will also be done in almost all of what follows since this case has been dealt with completely in Ref. 1.

Given Lemma 2.1 we are able to state the following.

Lemma 2.3: Assume $e^{2a|\mathbf{x}|}V\in R$ for some $a>0$. Then for all $\gamma\in\mathbb{R}$ and $\operatorname{Im}k>-a$, $k\neq -i\gamma/2n$, $n=1,2,\ldots$, $uG_{\gamma,k}v$ is a Hilbert–Schmidt operator in $L^2(\mathbb{R}^3)$. Moreover, if $k\in\Pi_\pm$ for $\gamma\gtrless 0$, $uG_{\gamma,k}v$ has an asymptotic expansion in k^2 near $k^2=0$ valid in norm, viz.,

$$uG_{\gamma,k}v \underset{k\to 0}{\sim} \sum_{n=0}^{\infty}(ik)^{2n}r_{\gamma,2n}, \quad \gamma\in\mathbb{R}\setminus\{0\},$$
$$k\in\Pi_\pm \quad \text{for } \gamma\gtrless 0. \tag{2.22}$$

Here $r_{\gamma,2n}$, $n=0,1,\ldots$, are Hilbert–Schmidt operators in $L^2(\mathbb{R}^3)$ with kernels given by [cf. Eq. (2.9)]

$$r_{\gamma,2n}(\mathbf{x},\mathbf{y}) = u(\mathbf{x})g_{\gamma,2n}(\mathbf{x},\mathbf{y})v(\mathbf{y}). \tag{2.23}$$

Proof: Let $\operatorname{Im}k>-a$, $k\neq -i\gamma/2n$, $n=1,2,\ldots$. First we note that

$$\lim_{\mathbf{x}\to\mathbf{y}}(4\pi|\mathbf{x}-\mathbf{y}|)G_{\gamma,k}(\mathbf{x},\mathbf{y}) = 1, \tag{2.24}$$

which easily follows from (2.15) using Wronskian relations for the confluent hypergeometric functions.[10]

Second, the curly bracket in Eq. (2.15) is uniformly bounded as long as $x_\pm$ vary in compact domains. Therefore, it suffices to consider $x_\pm\to\infty$. For this region, a simple extension of the bound (B9) of Ref. 6 implies [replace (B7) of Ref. 6 by $\sigma(\epsilon,\gamma,k^2r) = \rho(\gamma,k^2,r) + \epsilon^2\kappa_0^5|(\partial/\partial k^2)P(\gamma,k^2,r)|^2 + \epsilon^2\kappa_0^3|(\partial^2/\partial k^2\,\partial r)P(\gamma,k^2,r)|^2$, $\epsilon>0$]

$$|H_l^{(0)}(k,r)|$$
$$\leqslant \text{const}(l,\gamma,\kappa_0,R_0,\epsilon)\exp\{(1+\epsilon)\kappa_0(r-R_0) + (|\gamma|/2\kappa_0)$$
$$\times\ln(r/R_0)\}, \quad \gamma\in\mathbb{R}, \quad \epsilon>0, \quad |k^2|\leqslant\kappa_0^2,$$
$$\kappa_0>0, \quad r\geqslant R_0>0, \tag{2.25}$$

where $H_l^{(0)}$ denotes $F_l^{(0)}$, $(\partial/\partial k^2)F_l^{(0)}$, $\widetilde{G}_l^{(0)}$, or $(\partial/\partial k^2)\widetilde{G}_l^{(0)}$. Thus Eq. (2.20) tells us that $uG_{\gamma,k}v\in\mathscr{B}_2(L^2(\mathbb{R}^3))$ for k in the open sphere of radius a centered at zero in $\mathbb{C}$. Standard Bessel function estimates[10] for $k\neq 0$ then yield $uG_{\gamma,k}v\in\mathscr{B}_2(L^2(\mathbb{R}^3))$ for $|\operatorname{Im}k|<a$. Combining this with the estimate (4.1) of Ref. 11 in fact shows $uG_{\gamma,k}v\in\mathscr{B}_2(L^2(\mathbb{R}^3))$ for all $\operatorname{Im}k>-a$, $k\neq -i\gamma/2n$, $n=1,2,\ldots$. Moreover, the explicit expansion of $F_l^{(0)}(k,x_-/2)$, $\widetilde{G}_l^{(0)}(k,x_+/2)$ around $k^2=0$ in terms of (modified) Bessel functions if $\gamma<0$ ($\gamma>0$) [cf., e.g., Eqs. (A1)–(A6) of Ref. 14] and standard Bessel function estimates[10] prove $r_{\gamma,2n}\in\mathscr{B}_2(L^2(\mathbb{R}^3))$ for all $n=0,1,\ldots$. Finally, the Hilbert–Schmidt operators $uG_{\gamma,k}^{(0)}v$ and $uG_{\gamma,k}^{(1)}v$ [cf. Eq. (2.20)] are seen to be norm analytic with respect to k^2 around $k^2=0$ by taking, e.g., matrix elements with $C_0^\infty(\mathbb{R}^3)$ functions. These facts together with the asymptotic expansion (2.21) complete the proof. □

Having found an appropriate expansion for the Coulomb resolvent $G_{\gamma,k}$ we still have to expand the Coulomb scattering wave functions $\Psi_c^\pm$ around $k=0$. These are defined by (cf., e.g., Refs. 1 and 13)

$$\Psi_c^-(k,\omega,\mathbf{x}) = e^{-\pi\gamma/4k}\Gamma(1+i\gamma/2k)e^{ik\omega\cdot\mathbf{x}}$$
$$\times {}_1F_1(-i\gamma/2k;1;ik(|\mathbf{x}|-\omega\cdot\mathbf{x})),$$
$$\Psi_c^+(k,\omega,\mathbf{x}) = \overline{\Psi_c^-(k,-\omega,\mathbf{x})}, \quad \omega\in S^2, \quad k\neq 0 \tag{2.26}$$

(S^2 is the unit sphere in $\mathbb{R}^3$. We then state the following:

Lemma 2.4: Let $e^{2a|\mathbf{x}|}V\in R$ for some $a>0$. Then

$$e^{\pi\gamma/4k}\Gamma\left(1+\frac{i\gamma}{2k}\right)^{-1}\begin{cases}u\Psi_c^-(k,\omega),\\ \\ v\Psi_c^+(k,\omega),\end{cases} \quad \gamma\in\mathbb{R}, \tag{2.27}$$

defines two analytic functions in $L^2(\mathbb{R}^3)$ with respect to k around $k=0$. In particular,

$$e^{\pi\gamma/4k}\Gamma(1+i\gamma/2k)^{-1}\Psi_c^-(k,\omega,\mathbf{x})$$
$$= I_0([2\gamma(|\mathbf{x}|-\omega\cdot\mathbf{x})]^{1/2}) + \{|\mathbf{x}|I_0([2\gamma(|\mathbf{x}|-\omega\cdot\mathbf{x})]^{1/2})$$
$$- [2(|\mathbf{x}|-\omega\cdot\mathbf{x})/\gamma]^{1/2}I_1([2\gamma(|\mathbf{x}|-\omega\cdot\mathbf{x})]^{1/2})$$
$$- [(|\mathbf{x}|-\omega\cdot\mathbf{x})/2]I_2([2\gamma(|\mathbf{x}|-\omega\cdot\mathbf{x})]^{1/2})\}ik$$
$$+ O(k^2), \quad \gamma>0, \tag{2.28}$$

$$e^{\pi\gamma/4k}\Gamma(1+i\gamma/2k)^{-1}\Psi_c^-(k,\omega,\mathbf{x})$$
$$= J_0([2|\gamma|(|\mathbf{x}|-\omega\cdot\mathbf{x})]^{1/2})$$
$$+ \{|\mathbf{x}|J_0([2|\gamma|(|\mathbf{x}|-\omega\cdot\mathbf{x})]^{1/2})$$
$$- [2(|\mathbf{x}|-\omega\cdot\mathbf{x})/|\gamma|]^{1/2}J_1([2|\gamma|(|\mathbf{x}|-\omega\cdot\mathbf{x})]^{1/2})$$
$$+ [(|\mathbf{x}|-\omega\cdot\mathbf{x})/2]J_2([2|\gamma|(|\mathbf{x}|-\omega\cdot\mathbf{x})]^{1/2})\}(ik)$$
$$+ O(k^2), \quad \gamma<0, \tag{2.29}$$

and similarly for Ψ_c^+ [using (2.26)].

Proof: Using the expansion[15]

$$
{}_1F_1(-1/\alpha;1;\alpha z) = J_0((4z)^{1/2}) + \sum_{\nu=1}^{\infty} \alpha^\nu \sum_{\mu=1}^{\nu} (-1)^\mu \Gamma(1+\nu+\mu)^{-1} \times c_{\nu\mu} z^{(\nu+\mu)/2} J_{\nu+\mu}((4z)^{1/2}), \tag{2.30}
$$

with

$$
c_{11}=1, \quad c_{21}=2, \quad c_{22}=3, \quad \text{etc.}, \tag{2.31}
$$

where Eq. (2.30) defines an entire function with respect to α and z, we immediately obtain Eq. (2.28) and (2.29). To prove analyticity we note that $e^{2a|x|}V \in R$ for some $a>0$ implies $e^{2a'|x|}V \in L^1(\mathbf{R}^3)$ for all $a'<a$ such that it suffices to get an appropriate bound for $(\partial/\partial k)_1F_1(-i\gamma/2k;1;ik\xi)$ as $\xi \to +\infty, \xi>0$. By Kummer's transformation and differential properties for ${}_1F_1(\alpha;\beta;z)$ (Ref. 10) one obtains

$$
\frac{\partial}{\partial\xi} {}_1F_1(-i\gamma/2k;1;ik\xi) = \gamma\xi^{-1} e^{ik\xi/2} F_0^{(0)}(k,\xi/2). \tag{2.32}
$$

Hence the estimate (2.25) implies

$$
\begin{aligned}
&|{}_1F_1(-i\gamma/2k;1;ik\xi)| \\
&\leqslant c'(\gamma,\kappa_0,R_0) + |\gamma| \int_{R_0}^{\xi} d\xi'(\xi')^{-1} |e^{ik\xi'/2} F_0^{(0)}(k,\xi'/2)| \\
&\leqslant c(\gamma,\kappa_0,R_0,\epsilon)\{1 + (\xi - R_0)|\gamma| R_0^{-1} e^{\kappa_0\xi/2} \exp[(1+\epsilon) \\
&\quad \times \kappa_0(\xi - R_0) + (|\gamma|/2\kappa_0)\ln(\xi/R_0)]\}, \\
&\qquad \epsilon>0, \quad \xi \geqslant R_0>0, \quad |k| \leqslant \kappa_0.
\end{aligned} \tag{2.33}
$$

Similar arguments work for $(\partial/\partial k)_1F_1$ by differentiating (2.32) with respect to k. Analyticity in the weak (and hence in the strong) sense of the vector-valued functions (2.29) now simply follows from Eq. (2.27) by taking matrix elements with $C_0^\infty(\mathbf{R}^3)$ functions using the above bounds and dominated convergence. □

III. ZERO-ENERGY STATES IN COULOMB SYSTEMS

In this section, we study the behavior of the long-range systems defined by (2.7) at the zero-energy threshold, allowing explicitly for zero-energy bound states and zero-energy resonances of H. This will give us the different cases that are possible in these systems.

To discuss bound states of H, we see that, recalling the resolvent equation

$$
(H-k^2)^{-1} = G_{\gamma,k} - G_{\gamma,k} v (u G_{\gamma,k} v + 1)^{-1} u G_{\gamma,k}, \quad k^2 \notin \sigma(H), \quad \operatorname{Im} k>0, k \neq -i\gamma/2n, \quad n=1,2,..., \tag{3.1}
$$

it is natural to consider the transition operator

$$
(uG_{\gamma,k}v+1)^{-1}, \quad \operatorname{Im} k>0, \quad k \neq -i\gamma/2n, \quad n=1,2,.... \tag{3.2}
$$

In particular, for $k \to 0$ one has to study the eigenvalue -1 of $uG_{\gamma,0}v$.

Let $V \in R$ and assume

$$
uG_{\gamma,0}v\phi_\gamma = -\phi_\gamma, \quad \text{for some } \phi_\gamma \in L^2(\mathbf{R}^3), \quad \gamma \in \mathbf{R}. \tag{3.3}
$$

Then we first state Lemma 3.1.

Lemma 3.1: Assume $V \in L^1(\mathbf{R}^3)$ for $\gamma \geqslant 0$ and in addition $(1+|\mathbf{x}|^{1/2})V \in L^1(\mathbf{R}^3)$ for $\gamma<0$. If one defines

$$
\psi_\gamma(\mathbf{x}) = (G_{\gamma,0} v \phi_\gamma)(\mathbf{x}), \quad \gamma \in \mathbf{R}, \tag{3.4}
$$

then for all $\gamma \in \mathbf{R}$, $\psi_\gamma \in L^2_{\text{loc}}(\mathbf{R}^3)$ and

$$
[-\Delta + \gamma|\mathbf{x}|^{-1} + V(\mathbf{x})]\psi_\gamma(\mathbf{x}) = 0, \tag{3.5}
$$

in the sense of distributions.

Proof: Let $\gamma \geqslant 0$. Since $K_0(r)$ and $rK_1(r)$ are monotonically decreasing in $r>0$ (Ref. 10), viz.,

$$
\frac{d}{dr} K_0(r) = -K_1(r) < 0, \quad r>0, \tag{3.6}
$$

$$
\frac{d}{dr}[rK_1(r)] = -rK_0(r) < 0, \quad r>0, \tag{3.7}
$$

we immediately get from (2.10)

$$
\begin{aligned}
g_{\gamma,0}(\mathbf{x},\mathbf{y}) &\leqslant (4\pi|\mathbf{x}-\mathbf{y}|)^{-1}(2\gamma x_-)^{1/2}[I_1((2\gamma x_-)^{1/2}) \\
&\quad \times K_0((2\gamma x_-)^{1/2}) + I_0((2\gamma x_-)^{1/2})K_1((2\gamma x_-)^{1/2})] \\
&= (4\pi|\mathbf{x}-\mathbf{y}|)^{-1} = g_{0,0}(\mathbf{x},\mathbf{y}), \quad \gamma \geqslant 0,
\end{aligned} \tag{3.8}
$$

where we have used the Wronskian relation between K_0 and I_0 (Ref. 10). Thus the first part of Proposition 2.1 in Ref. 4 directly applies and $\psi_\gamma \in L^2_{\text{loc}}(\mathbf{R}^3)$, $\gamma \geqslant 0$.

For $\gamma<0$ we note that standard Bessel function estimates[10] applied to Eq. (2.11) yield

$$
\begin{aligned}
|g_{\gamma,0}(\mathbf{x},\mathbf{y})| &\leqslant c'|\mathbf{x}-\mathbf{y}|^{-1}[1 + x_+^{1/4}] \\
&\leqslant c(1+|\mathbf{x}|^{1/4})|\mathbf{x}-\mathbf{y}|^{-1}(1+|\mathbf{y}|^{1/4}).
\end{aligned} \tag{3.9}
$$

So one can again follow the first part in the proof of Proposition 2.1 in Ref. 4 step by step. Since $V\psi_\gamma = -v\phi_\gamma \in L^1(\mathbf{R}^3)$, $V\psi_\gamma$ defines a distribution and Eq. (3.5) is proved as in the short-range case $\gamma=0$. □

A more detailed investigation for repulsive Coulomb systems leads to Lemma 3.2.

Lemma 3.2: Assume $\gamma>0$ and let $V \in R$, $(1+|\mathbf{x}|^{1+\epsilon}) \times V \in L^1(\mathbf{R}^3)$ for some $\epsilon>0$. Then ψ_γ, $\nabla\psi_\gamma \in L^2(\mathbf{R}^3)$ (the derivative is understood in the sense of distributions). In addition $\psi_\gamma \in \mathscr{D}(H)$ and

$$
H\psi_\gamma = 0, \tag{3.10}
$$

i.e., ψ_γ is a zero-energy eigenstate of H.

Proof: From Eq. (2.14) we have that

$$
\begin{aligned}
&x_+ \geqslant 2|\mathbf{y}|, \quad x_- \leqslant 2|\mathbf{x}|, \quad \text{if } |\mathbf{x}| \leqslant |\mathbf{y}|, \\
&x_+ \geqslant 2|\mathbf{x}|, \quad x_- \leqslant 2|\mathbf{y}|, \quad \text{if } |\mathbf{x}| \geqslant |\mathbf{y}|.
\end{aligned} \tag{3.11}
$$

Next, we note the following properties of the modified Bessel functions. First

$$
I_0(r) \geqslant I_1(r), \quad K_0(r) \leqslant K_1(r), \quad r>0 \tag{3.12}
$$

and, second, we recall Eq. (3.6) and

$$
\frac{d}{dr}[rI_1(r)] = rI_0(r) > 0, \quad r>0. \tag{3.13}
$$

Using this information in Eq. (2.10) one derives

$$
g_{\gamma,0}(\mathbf{x},\mathbf{y}) \leqslant (2\pi|\mathbf{x}-\mathbf{y}|)^{-1} \begin{cases} (4\gamma|\mathbf{y}|)^{1/2} I_0((4\gamma|\mathbf{x}|)^{1/2}) K_1((4\gamma|\mathbf{y}|)^{1/2}), & |\mathbf{x}| \leqslant |\mathbf{y}|, \\ (4\gamma|\mathbf{x}|)^{1/2} I_0((4\gamma|\mathbf{y}|)^{1/2}) K_1((4\gamma|\mathbf{x}|)^{1/2}), & |\mathbf{x}| \geqslant |\mathbf{y}|. \end{cases} \tag{3.14}
$$

By Lemma 3.1 it suffices to take $0<c<|\mathbf{x}|$ (c large enough). From Eqs. (3.4) and (3.14) we get

$$|\psi_\gamma(\mathbf{x})|\leqslant\int_{\mathbf{R}^3}d^3y\,|g_{\gamma,0}(\mathbf{x},\mathbf{y})|v(\mathbf{y})|\phi_\gamma(\mathbf{y})|\leqslant(2\pi|\mathbf{x}-\mathbf{y}|)^{-1}\int_{|\mathbf{y}|<|\mathbf{x}|}d^3y(4\gamma|\mathbf{x}|)^{1/2}I_0((4\gamma|\mathbf{y}|)^{1/2})K_1((4\gamma|\mathbf{x}|)^{1/2})v(\mathbf{y})|\phi_\gamma(\mathbf{y})|$$
$$+(2\pi|\mathbf{x}-\mathbf{y}|)^{-1}\int_{|\mathbf{y}|>|\mathbf{x}|}d^3y(4\gamma|\mathbf{y}|)^{1/2}I_0((4\gamma|\mathbf{x}|)^{1/2})K_1((4\gamma|\mathbf{y}|)^{1/2})v(\mathbf{y})|\phi_\gamma(\mathbf{y})|\equiv\psi_{1,\gamma}(\mathbf{x})+\psi_{2,\gamma}(\mathbf{x}). \tag{3.15}$$

Next we split up $\psi_{1,\gamma}$ into

$$\psi_{1,\gamma}(\mathbf{x})=(2\pi|\mathbf{x}|)^{-1}\int_{|\mathbf{y}|<|\mathbf{x}|}d^3y(4\gamma|\mathbf{x}|)^{1/2}I_0((4\gamma|\mathbf{y}|)^{1/2})K_1((4\gamma|\mathbf{x}|)^{1/2})v(\mathbf{y})|\phi_\gamma(\mathbf{y})|+(2\pi)^{-1}\int_{|\mathbf{y}|<|\mathbf{x}|}d^3y[|\mathbf{x}-\mathbf{y}|^{-1}$$
$$-|\mathbf{x}|^{-1}](4\gamma|\mathbf{x}|)^{1/2}I_0((4\gamma|\mathbf{y}|)^{1/2})K_1((4\gamma|\mathbf{x}|)^{1/2})v(\mathbf{y})|\phi_\gamma(\mathbf{y})|\equiv\psi^{(1)}_{1,\gamma}(\mathbf{x})+\psi^{2}_{1,\gamma}(\mathbf{x}). \tag{3.16}$$

From the fact that[10]

$$\frac{d}{dr}[r(1+br^{1+\epsilon})K_1(r)]=-rK_0(r)+b(2+\epsilon)r^{1+\epsilon}K_1(r)+br^{2+\epsilon}K_1'(r)$$
$$\leqslant-rK_0(r)+br^{2+\epsilon}[K_1(r)+K_1'(r)]<0,\quad\text{for } r\geqslant2+\epsilon,\quad\epsilon,b>0 \tag{3.17}$$

[since $K_1(r)+K_1'(r)<0$, $r>0$], we obtain

$$\psi^{(1)}_{1,\gamma}(\mathbf{x})=\{2\pi|\mathbf{x}|[1+b|\mathbf{x}|^{(1+\epsilon)/2}]\}^{-1}\int_{|\mathbf{y}|<|\mathbf{x}|}d^3y[1+b|\mathbf{x}|^{(1+\epsilon)/2}](4\gamma|\mathbf{x}|)^{1/2}I_0((4\gamma|\mathbf{y}|)^{1/2})K_1((4\gamma|\mathbf{x}|)^{1/2})v(\mathbf{y})|\phi_\gamma(\mathbf{y})|$$
$$\leqslant c'|\mathbf{x}|^{-(3+\epsilon)/2}\int_{\mathbf{R}^3}d^3y[1+b|\mathbf{y}|^{(1+\epsilon)/2}]v(\mathbf{y})|\phi_\gamma(\mathbf{y})|, \tag{3.18}$$

where we have used

$$rI_0(r)K_1(r)\leqslant\text{const}. \tag{3.19}$$

By Schwarz' inequality and the hypotheses on V the integral in (3.18) exists and hence $\psi^{(1)}_{1,\gamma}$ is L^2 near infinity. In order to treat $\psi^{(2)}_{1,\gamma}$ we start from

$$\int_{|\mathbf{x}|>c}d^3x\,|\psi^{(2)}_{1,\gamma}(\mathbf{x})|^2\leqslant\int_{|\mathbf{x}|>c}d^3x(4\pi^2)^{-1}\left\{\int_{|\mathbf{y}|<|\mathbf{x}|}d^3y\,||\mathbf{x}-\mathbf{y}|^{-1}-|\mathbf{x}|^{-1}|(4\gamma|\mathbf{y}|)^{1/2}I_0((4\gamma|\mathbf{y}|)^{1/2})K_1((4\gamma|\mathbf{y}|)^{1/2})v(\mathbf{y})|\phi_\gamma(\mathbf{y})|\right\}^2$$
$$\leqslant c''\int_{|\mathbf{x}|>c}d^3x\left\{\int_{\mathbf{R}^3}d^3y\,||\mathbf{x}-\mathbf{y}|^{-1}-|\mathbf{x}|^{-1}|v(\mathbf{y})|\phi_\gamma(\mathbf{y})|\right\}^2, \tag{3.20}$$

where we have employed (3.7) in the first inequality and the estimate (3.19) in the last inequality. Interchanging the order of integrations in Eq. (3.20) by Fubini's theorem and observing that

$$\int_{\mathbf{R}^3}d^3x\,||\mathbf{x}-\mathbf{y}|^{-1}-|\mathbf{x}|^{-1}|\,||\mathbf{x}-\mathbf{z}|^{-1}-|\mathbf{x}|^{-1}|\leqslant\text{const}|\mathbf{y}|^{1/2}|\mathbf{z}|^{1/2}, \tag{3.21}$$

finally proves that also $\psi^{(2)}_{1,\gamma}$ is L^2 near infinity. Next we split up $\psi_{2,\gamma}$ into

$$\psi_{2,\gamma}(\mathbf{x})=(2\pi|\mathbf{x}|)^{-1}\int_{|\mathbf{y}|>|\mathbf{x}|}d^3y(4\gamma|\mathbf{y}|)^{1/2}I_0((4\gamma|\mathbf{x}|)^{1/2})K_1((4\gamma|\mathbf{y}|)^{1/2})v(\mathbf{y})|\phi_\gamma(\mathbf{y})|+(2\pi)^{-1}\int_{|\mathbf{y}|>|\mathbf{x}|}d^3y[|\mathbf{x}-\mathbf{y}|^{-1}$$
$$-|\mathbf{x}|^{-1}](4\gamma|\mathbf{y}|)^{1/2}I_0((4\gamma|\mathbf{x}|)^{1/2})K_1((4\gamma|\mathbf{y}|)^{1/2})v(\mathbf{y})|\phi_\gamma(\mathbf{y})|\equiv\psi^{(1)}_{2,\gamma}(\mathbf{x})+\psi^{(2)}_{2,\gamma}(\mathbf{x}). \tag{3.22}$$

Employing the monotonic decrease of $r^{-\epsilon}K_1(r)$, viz.,

$$\frac{d}{dr}[r^{-\epsilon}K_1(r)]<0,\quad r>0,\quad\epsilon\geqslant0, \tag{3.23}$$

we derive

$$\psi^{(1)}_{2,\gamma}(\mathbf{x})=(2\pi|\mathbf{x}|)^{-1}\int_{|\mathbf{y}|>|\mathbf{x}|}d^3y\,|\mathbf{y}|^{(1+\epsilon)/2}|\mathbf{y}|^{-(1+\epsilon)/2}(4\gamma|\mathbf{y}|)^{1/2}I_0((4\gamma|\mathbf{x}|)^{1/2})K_1((4\gamma|\mathbf{y}|)^{1/2})v(\mathbf{y})|\phi_\gamma(\mathbf{y})|$$
$$\leqslant c'|\mathbf{x}|^{-1}\int_{|\mathbf{y}|>|\mathbf{x}|}d^3y\,|\mathbf{y}|^{(1+\epsilon)/2}|\mathbf{x}|^{-\epsilon/2}I_0((4\gamma|\mathbf{x}|)^{1/2})K_1((4\gamma|\mathbf{x}|)^{1/2})v(\mathbf{y})|\phi_\gamma(\mathbf{y})|$$
$$\leqslant c''|\mathbf{x}|^{-(3+\epsilon)/2}\int_{\mathbf{R}^3}d^3y\,|\mathbf{y}|^{(1+\epsilon)/2}v(\mathbf{y})|\phi_\gamma(\mathbf{y})|, \tag{3.24}$$

where the bound (3.19) has been used in the last inequality (3.24). Thus $\psi^{(1)}_{2,\gamma}$ is L^2 near infinity. The proof for $\psi^{(2)}_{2,\gamma}$ is similar, using inequality (3.20) and Eqs. (3.6) and (3.19). Next we sketch the proof of $\nabla\psi_\gamma \in L^2(\mathbb{R}^3)$. From Eqs. (2.10) and (3.4) we get

$$(\nabla\psi_\gamma)(\mathbf{x}) = (4\pi)^{-1}\int_{\mathbf{R}^3} d^3y\{|\mathbf{x}-\mathbf{y}|^{-3}(\mathbf{y}-\mathbf{x})A_\gamma(\mathbf{x},\mathbf{y}) + |\mathbf{x}-\mathbf{y}|^{-1}(\nabla_\mathbf{x}A_\gamma)(\mathbf{x},\mathbf{y})\} \equiv (\nabla\psi^{(1)}_\gamma)(\mathbf{x}) + (\nabla\psi^{(2)}_\gamma)(\mathbf{x}), \tag{3.25}$$

where

$$A_\gamma(\mathbf{x},\mathbf{y}) = [(2\gamma x_-)^{1/2}I_1((2\gamma x_-)^{1/2})K_0((2\gamma x_+)^{1/2}) + (2\gamma x_+)^{1/2}I_0((2\gamma x_-)^{1/2})K_1((2\gamma x_+)^{1/2})]v(\mathbf{y})\phi_\gamma(\mathbf{y}). \tag{3.26}$$

Since

$$|A_\gamma(\mathbf{x},\mathbf{y})| \leqslant v(\mathbf{y})|\phi_\gamma(\mathbf{y})|, \tag{3.27}$$

by the estimate (3.8), we obtain from Fubini's theorem and a special case of Riesz' composition formula

$$\int_{\mathbf{R}^3} d^3x\,|(\nabla\psi^{(1)}_\gamma)(\mathbf{x})|^2 \leqslant (4\pi)^{-2}\int_{\mathbf{R}^6} d^3y\,d^3z\,v(\mathbf{y})v(\mathbf{z})|\phi_\gamma(\mathbf{y})|\,|\phi_\gamma(\mathbf{z})|\int_{\mathbf{R}^3} d^3x\,|\mathbf{x}-\mathbf{y}|^{-2}|\mathbf{x}-\mathbf{z}|^{-2}$$
$$= c\int_{\mathbf{R}^6} d^3y\,d^3z\,v(\mathbf{y})|\mathbf{y}-\mathbf{z}|^{-1}v(\mathbf{z})|\phi_\gamma(\mathbf{y})|\,|\phi_\gamma(\mathbf{z})| \leqslant c'\|V\|_R\|\phi_\gamma\|_2^2 < \infty. \tag{3.28}$$

Thus $\nabla\psi^{(1)}_\gamma \in L^2(\mathbb{R}^3)$. Next we note

$$|\nabla_\mathbf{x}A_\gamma(\mathbf{x},\mathbf{y})| \leqslant 2\gamma\{I_0((2\gamma x_-)^{1/2})K_0((2\gamma x_+)^{1/2}) + (x_+/x_-)^{1/2}I_1((2\gamma x_-)^{1/2})K_1((2\gamma x_+)^{1/2})\}v(\mathbf{y})|\phi_\gamma(\mathbf{y})|$$
$$\leqslant c[|\ln(|\mathbf{y}|)| + 1]v(\mathbf{y})|\phi_\gamma(\mathbf{y})|, \tag{3.29}$$

where Eqs. (3.6), (3.7), (3.11),

$$\frac{d}{dr}I_0(r) = I_1(r) > 0, \quad r>0, \tag{3.30}$$

$$I_0(r)K_0(r) \leqslant c[1+|\ln(r)|], r>0, \tag{3.31}$$

and[16]

$$I_1(r)K_1(r) \leqslant 1/2, \quad r \geqslant 0 \tag{3.32}$$

have been used. Inequality (3.29) proves $\nabla\psi^{(2)}_\gamma \in L^2_{\mathrm{loc}}(\mathbb{R}^3)$. To show that $\nabla\psi^{(2)}_\gamma$ is L^2 near infinity we assume $|\mathbf{x}| \geqslant c > 0$ and insert $K_0(r) \leqslant K_1(r)$ into (3.29). From Eq. (3.11) and the monotonicity properties mentioned before we obtain

$$|\nabla\psi^{(2)}_\gamma(\mathbf{x})| \leqslant \gamma\int_{|\mathbf{y}|<|\mathbf{x}|} d^3y(2\pi|\mathbf{x}-\mathbf{y}|)^{-1}\{I_0((4\gamma|\mathbf{y}|)^{1/2})K_1((4\gamma|\mathbf{x}|)^{1/2}) + (|\mathbf{x}|/|\mathbf{y}|)^{1/2}I_1((4\gamma|\mathbf{y}|)^{1/2})$$
$$\times K_1((4\gamma|\mathbf{x}|)^{1/2})\}v(\mathbf{y})|\phi_\gamma(\mathbf{y})| + \gamma\int_{|\mathbf{y}|>|\mathbf{x}|} d^3y(2\pi|\mathbf{x}-\mathbf{y}|)^{-1}\{I_0((4\gamma|\mathbf{x}|)^{1/2})K_1((4\gamma|\mathbf{y}|)^{1/2})$$
$$+ (|\mathbf{y}|/|\mathbf{x}|)^{1/2}I_1((4\gamma|\mathbf{x}|)^{1/2})K_1((4\gamma|\mathbf{y}|)^{1/2})\}v(\mathbf{y})|\phi_\gamma(\mathbf{y})| \equiv (\nabla\psi^{(2)}_\gamma)_1(\mathbf{x}) + (\nabla\psi^{(2)}_\gamma)_2(\mathbf{x}). \tag{3.33}$$

From now on one can follow directly the proof of ψ_γ is L^2 near infinity using the monotonic increase of $r^{-1}I_1(r)$:

$$\frac{d}{dr}[r^{-1}I_1(r)] = r^{-1}I_2(r) > 0, \quad r>0. \tag{3.34}$$

The arguments that $\psi_\gamma \in \mathscr{D}(H)$ parallel those in Corollary II.8 of Ref. 9.

Remark 3.3: Lemma 3.2 shows that for repulsive Coulomb systems, only a threshold *bound state* is possible if the short-range potential V has a critical strength (cf. Refs. 17 and 18 for the notion of criticality). This phenomenon of getting automatically a threshold eigenstate of H has also been observed in two-electron systems.[19] In the special case of a spherically symmetric V it has been announced in Ref. 20. For the time decay of wave functions see Ref. 21.

For attractive Coulomb systems we note the following.

Lemma 3.4: Let $V \in R$ and assume that V has compact support. If $\gamma < 0$ then $\psi_\gamma \notin L^2(\mathbb{R}^3)$.

Proof: We first note the (rather crude) estimate

$$|J_{2l+1}((4|\gamma|r')^{1/2})H^{(1)}_{2l+1}((4|\gamma|r)^{1/2})|$$
$$\leqslant c\left\{\frac{(4|\gamma|^2rr')^{l+1/2}}{\Gamma(2l+1)^2} + \frac{(|\gamma|r')^{l+1/2}}{(|\gamma|r)^{3l+3/2}}\right\},$$
$$l = 0,1,2,\ldots, \quad r'<r, \tag{3.35}$$

which follows from[10]

$$|J_\nu(x)| \leqslant \Gamma(1+\nu)^{-1}(x/2)^\nu, \quad \nu \geqslant -1/2, \quad x \geqslant 0, \tag{3.36}$$

and a similar estimate on $|H^{(1)}_\nu(x)|$ based, e.g., on the integral representation 8.4.219 in Ref. 22. The estimate (3.35) proves the absolute convergence of the angular momentum expansion of the full Coulomb Green's function at zero energy, i.e.,

$$G_{\gamma,0}(\mathbf{x},\mathbf{y}) = i\pi(|\mathbf{x}|\,|\mathbf{y}|)^{-1/2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l} J_{2l+1}((4|\gamma|\,|\mathbf{y}|)^{1/2})$$
$$\times H^{(1)}_{2l+1}((4|\gamma|\,|\mathbf{x}|)^{1/2})\,\overline{Y_{l,m}(\omega_\mathbf{y})}\,Y_{l,m}(\omega_\mathbf{x}),$$
$$\gamma<0, \quad |\mathbf{y}|<|\mathbf{x}| \tag{3.37}$$

($Y_{l,m}$ denote the spherical harmonics). Insertion of Eq. (3.37) into Eq. (3.4), interchanging summation and integration, which is allowed by the Lebesgue dominated convergence theorem [use estimate (3.35) and the fact that supp V is compact], yields for $\mathbf{x}$ outside supp V

$$\psi_\gamma(\mathbf{x}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{l,m} |\mathbf{x}|^{-1/2} H^{(1)}_{2l+1}((4|\gamma|\,|\mathbf{x}|)^{1/2}) \times Y_{l,m}(\omega_\mathbf{x}), \quad |\mathbf{x}| \text{ large enough}, \tag{3.38}$$

$$c_{l,m} = i\pi \int_{\text{supp } V} d^3y\, |\mathbf{y}|^{-1/2} J_{2l+1}((4|\gamma|\,|\mathbf{y}|)^{1/2}) \times v(\mathbf{y})\phi_\gamma(\mathbf{y})\, \overline{Y_{l,m}(\omega_\mathbf{y})}.$$

Further, if $\chi_R(|\mathbf{x}|)$ denotes the characteristic function of the closed ball centered at the origin of radius $R>0$, we infer from Lemma 3.1 that $\chi_R \psi_\gamma \in L^2(\mathbb{R}^3)$ for all $R>0$. In particular,

$$\|\chi_R \psi_\gamma\|_2^2 = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} |c_{l,m}|^2 \int_0^R r^2\, dr\, r^{-1} \times |H^{(1)}_{2l+1}((4|\gamma| r)^{1/2})|^2 < \infty. \tag{3.39}$$

Since[10]

$$|H^{(1)}_{2l+1}((4|\gamma| r)^{1/2})|^2 \geqslant \pi^{-1}(|\gamma| r)^{-1/2}, \quad r>0, \tag{3.40}$$

$\|\chi_R\psi_\gamma\|_2$ has no limit as $R\to\infty$ unless all $c_{l,m}=0$. This proves $\psi_\gamma \notin L^2(\mathbb{R}^3)$. □

According to Lemmas 3.1, 3.2, and 3.4 we now introduce the following case distinction for Coulomb systems ($\gamma\neq 0$) assuming $V\in L^1(\mathbb{R}^3)\cap R$ for $\gamma>0$ and in addition $(1+|\mathbf{x}|^{1/2})V\in L^1(\mathbb{R}^3)$ for $\gamma<0$.

Case I: $uG_{\gamma,0}v$ has no eigenvalue -1 (i.e., no ψ_γ exists).

Case II: $\gamma>0$ and there exist N linearly independent functions $\psi_{\gamma,j}\in\mathscr{D}(H)$, $j=1,\dots,N$.

Case III: $\gamma<0$ and there exist N linearly independent functions $\psi_{\gamma,j}\in L^2_{\text{loc}}(\mathbb{R}^3)$, $j=1,\dots,N$ which are not in $L^2(\mathbb{R}^3)$.

Remark 3.5: (a) Clearly case I is the generic one. In particular, if $\gamma>0$ and $V\geqslant 0$ then only case I occurs.

(b) Since $uG_{\gamma,0}v$ is Hilbert–Schmidt for all $\gamma\in\mathbb{R}$ the number N in cases II or III is certainly finite. In case II, H has a zero-energy bound state of multiplicity N.

(c) In case III zero is a limit point of negative discrete eigenvalues of H (and 0 is not the endpoint of a spectral gap of H). This means that the techniques developed in Refs. 1, 4, 18, and 23–29 do not apply directly. So a careful investigation is needed, going far beyond the methods used in this paper. Because of this we restrict ourselves in the following to the cases I and II.

IV. EXPANSION OF THE TRANSITION AND SCATTERING OPERATOR

Knowing the threshold behavior for the Coulomb-type systems we are considering in detail, we now expand the transition operator (3.2) and the scattering amplitude.

To do this we first have to study the exceptional set $\mathscr{E}_\gamma$, $\gamma\in\mathbb{R}$ given by

$$\mathscr{E}_\gamma = \{k^2\geqslant 0 \mid uG_{\gamma,k}v\phi = -\phi, \text{ for some } \phi\in L^2(\mathbb{R}^3) \text{ and } k\geqslant 0\}, \quad \gamma\in\mathbb{R}. \tag{4.1}$$

As a first characterization of $\mathscr{E}_\gamma$ we state the following.

Lemma 4.1: If $V\in R$ then for all $\gamma\in\mathbb{R}$, $\mathscr{E}_\gamma$ is a compact subset of $[0,\infty)$ with Lebesgue measure zero. If in addition $e^{2a|\mathbf{x}|}V\in R$ for some $a>0$, then $\mathscr{E}_\gamma$ is discrete (i.e., 0 is the only possible accumulation point of $\mathscr{E}_\gamma$ in $[0,\infty)$) and hence the singular continuous spectrum of H is empty.

Proof: For completeness we essentially repeat the arguments of Ref. 1: As discussed in Refs. 30 and 31, $\mathscr{E}_\gamma$ is a closed set of Lebesgue measure zero which contains the singular continuous spectrum of H. Boundedness of $\mathscr{E}_\gamma$ then follows from a Klein–Zemach-type result.[9,31] If $e^{2a|\mathbf{x}|}V\in R$, one infers from Lemma 2.3 that $uG_{\gamma,k}v$ is Hilbert–Schmidt and analytic in the half-plane Im $k>-a$ cut along the nonpositive real line and along

$$\{k\in\mathbb{C} \mid \text{Re } k=0,\ -a\leqslant \text{Im } k\leqslant 0, \text{ for } \gamma>0 \text{ and } -a\leqslant \text{Im } k\leqslant -\gamma/2, \text{ for } \gamma<0\}.$$

An application of the analytic Fredholm theorem[32] then proves discreteness of $\mathscr{E}_\gamma$. □

A priori Lemma 4.1 does not rule out the fact that 0 is an accumulation point of $\mathscr{E}_\gamma$. In case I, however, $0\notin\mathscr{E}_\gamma$ such that the closedness of $\mathscr{E}_\gamma$ indeed implies that 0 is not an accumulation point. Hence $[0,k_0^2]\cap\mathscr{E}_\gamma=\varnothing$ for $0<k_0$ small enough. For case II we introduce the following.

Assumption A: Zero is not an accumulation point of $\mathscr{E}_\gamma$ in case II.

This assumption is supposed to be valid throughout the rest of the paper.

Remark 4.2: If V is spherically symmetric and

$$\int_0^R dr\, r|V(r)| + \int_R^\infty dr\, |V(r)| < \infty, \quad \text{for some } R>0,$$

then $\mathscr{E}_\gamma\cap(0,\infty)=\varnothing$ using ordinary differential equation (ODE) techniques.[33] Finally, assuming $e^{2a|\mathbf{x}|}V\in R$ for some $a>0$, one can follow Ref. 34, Chap. X (cf. also Ref. 35) to prove that only $0<k_0^2\in\mathscr{E}_\gamma$ is a positive eigenvalue of H. This is sketched in the Appendix. In a second step one uses well-known results on the absence of positive eigenvalues of H (see e.g., Refs. 32, 36, and 37) to show the validity of assumption A.

As a first result we expand the transition operator.

Lemma 4.3: Let $e^{2a|\mathbf{x}|}V\in R$ for some $a>0$ and assume case I or II. Then $(uG_{\gamma,0}v+1+\epsilon)^{-1}$, $\epsilon>0$ has a norm-convergent Laurent expansion around $\epsilon=0$, viz.,

$$(uG_{\gamma,0}v+1+\epsilon)^{-1} = \epsilon^{-1}P_\gamma + \sum_{m=0}^{\infty}(-\epsilon)^m T_\gamma^{m+1}, \quad \epsilon>0, \tag{4.2}$$

where P_γ is the projector onto the eigenspace of $uG_{\gamma,0}v$ to the eigenvalue -1, i.e.,

$$P_\gamma = 0, \quad \text{in case I}, \tag{4.3}$$

$$P_\gamma = \sum_{j=1}^{N} \frac{(\tilde\phi_{\gamma,j},\cdot)\phi_{\gamma,j}}{(\tilde\phi_{\gamma,j},\phi_{\gamma,j})}, \quad \text{in case II}, \tag{4.4}$$

with

$$\begin{aligned} &uG_{\gamma,0}v\phi_{\gamma,j} = -\phi_{\gamma,j},\\ &\tilde\phi_{\gamma,j} = (\text{sgn } V)\phi_{\gamma,j}, \quad \dim P_\gamma = N,\\ &(\tilde\phi_{\gamma,j},\phi_{\gamma,l}) = 0, \quad \textit{for } j\neq l,\\ &(\tilde\phi_{\gamma,j},\phi_{\gamma,j})\neq 0, \quad 1\leqslant j,l\leqslant N. \end{aligned} \tag{4.5}$$

The reduced resolvent T_γ is given by

$$T_\gamma = (uG_{\gamma,0}v + 1)^{-1}, \quad \text{in case I}, \tag{4.6}$$

$$T_\gamma = n - \lim_{\epsilon\to 0}(uG_{\gamma,0}v + 1 + \epsilon)^{-1}(1 - P_\gamma), \quad \text{in case II}. \tag{4.7}$$

Proof: One can follow the proof of Lemma 3.1 of Ref. 4 where the case $\gamma = 0$ has been treated in detail step by step. □

Lemma 4.4: Let $e^{2a|\mathbf{x}|}V\in R$ for some $a > 0$ and suppose assumption A is fulfilled. Moreover, assume the existence of the inverse of the $N\times N$ matrix $(\tilde{\phi}_{\gamma,j}, r_{\gamma,2}\phi_{\gamma,l})$, $1\leqslant j,l\leqslant N$ [denoted by $(\tilde{\phi}_\gamma, r_{\gamma,2}\phi_\gamma)_{jl}^{-1}$] in case II. Then, for $(0,k_0^2)\cap\mathscr{E}_\gamma = \varnothing$, $0 < k_0$ small enough, $(uG_{\gamma,k}v + 1)^{-1}$ has the following asymptotic expansions in cases I and II valid in norm:

$$(uG_{\gamma,k}v + 1)^{-1} \underset{k\to 0}{\sim} \sum_{n=-M}^{\infty} (ik)^{2n} t_{\gamma,2n},$$
$$0 < |k| < k_0 \quad \text{small enough},$$
$$k\in\Pi_\pm \quad \text{for } \gamma \gtrless 0, \tag{4.8}$$

where $M = 0$ in case I and $M = 1$ in case II. The first coefficients are given explicitly by

$$t_{\gamma,0} = (uG_{\gamma,0}v + 1)^{-1} = T_\gamma, \tag{4.9}$$

$$t_{\gamma,2} = -T_\gamma r_{\gamma,2} T_\gamma, \quad \text{in case I} \tag{4.10}$$

and

$$t_{\gamma,-2} = \sum_{j,l=1}^{N} (\tilde{\phi}_\gamma, r_{\gamma,2}\phi_\gamma)_{jl}^{-1}(\tilde{\phi}_{\gamma,l},\cdot)\phi_{\gamma,j}, \quad \text{in case II}. \tag{4.11}$$

Proof: It suffices to consider case II. By Lemma 4.3 we obtain

$$(uG_{\gamma,k}v + 1)^{-1} \underset{k\to 0}{\sim} \{1 - (uG_{\gamma,0}v + 1 + k^2)^{-1}k^2[1 + r_{\gamma,2} + O(k^2)]\}^{-1}(uG_{\gamma,0}v + 1 + k^2)^{-1}$$
$$\underset{k\to 0}{\sim} \{1 - P_\gamma[1 + r_{\gamma,2}] + O(k^2)\}^{-1}[k^{-2}P_\gamma + T_\gamma + O(k^2)]$$
$$\underset{k\to 0}{\sim} -k^{-2}\sum_{j,l=1}^{N} (\tilde{\phi}_\gamma, r_{\gamma,2}\phi_\gamma)_{jl}^{-1}(\tilde{\phi}_{\gamma,l},\cdot)\phi_{\gamma,j} + O(1), \quad 0 < |k| < k_0 \quad \text{small enough}, \quad k\in\Pi_\pm, \quad \text{for } \gamma\gtrless 0, \tag{4.12}$$

where[4]

$$\{1 - P_\gamma[1 + r_{\gamma,2}]\}^{-1}P_\gamma = -\sum_{j,l=1}^{N} (\tilde{\phi}_\gamma, r_{\gamma,2}\phi_\gamma)_{jl}^{-1}(\tilde{\phi}_{\gamma,l},\cdot)\phi_{\gamma,j} \tag{4.13}$$

has been used in the last step. Clearly all higher-order coefficients $t_{\gamma,2n}$, $n = 0,1,\ldots$, can be derived along these lines. □

Next we introduce the on-shell scattering matrix in $L^2(S^2)$ associated with H, denoted by $S(k)$, $k > 0$, $k^2\notin\mathscr{E}_\gamma$. Let

$$S^c(k) = \frac{\Gamma(\frac{1}{2} + (\mathbf{L}^2 + \frac{1}{4})^{1/2} + i\gamma/2k)}{\Gamma(\frac{1}{2} + (\mathbf{L}^2 + \frac{1}{4})^{1/2} - i\gamma/2k)}, \quad k > 0 \tag{4.14}$$

be the pure Coulomb on-shell scattering operator[38,39] where $\mathbf{L}^2$ represents the square of the angular momentum operator. Then $S(k)$ can be written as

$$S(k) = S^c(k) + T^{sc}(k), \quad k > 0, \quad k^2\notin\mathscr{E}_\gamma, \tag{4.15}$$

where $T^{sc}(k)$, $k > 0$, $k^2\notin\mathscr{E}_\gamma$ is a trace class operator in $L^2(S^2)$ (continuous in trace norm with respect to k for $k^2\notin\mathscr{E}_\gamma$) with kernel given by the Coulomb modified scattering amplitude $f^{sc}(k,\omega,\omega')$ (Ref. 13), viz.,

$$(T^{sc}(k)\phi)(\omega) = -(2\pi i)^{-1}k\int_{S^2} d\omega' f^{sc}(k,\omega,\omega')\phi(\omega'), \quad \phi\in L^2(S^2), \tag{4.16}$$

$$f^{sc}(k,\omega,\omega') = -(4\pi)^{-1}(v\Psi_c^+(k,\omega),(uG_{\gamma,k}v + 1)^{-1}u\Psi_c^-(k,\omega')), \quad k > 0, \quad k^2\notin\mathscr{E}_\gamma. \tag{4.17}$$

Since $S^c(k)$ is explicitly known [Eq. (4.14)] we only have to study the remainder term in $S(k)$, viz., $T^{sc}(k)$. The asymptotic expansion of its kernel f^{sc} near the threshold $k = 0$ is obtained from the following.

Theorem 4.5: Let $e^{2a|\mathbf{x}|}V\in R$ for some $a > 0$ and suppose assumption A holds. Assume that $(\tilde{\phi}_\gamma, r_{\gamma,2}\phi_\gamma)_{jl}^{-1}$ exists in case II. Then, for $(0,k_0^2)\cap\mathscr{E}_\gamma = \varnothing$, $0 < k_0$ small enough, we get the following asymptotic expansions in cases I and II:

$$-(4\pi)(\pi\gamma/k)^{-1}(e^{\pi\gamma/k} - 1)\frac{\Gamma(1 - i\gamma/2k)}{\Gamma(1 + i\gamma/2k)}f^{sc}(k,\omega,\omega') \underset{k\to 0}{\sim} \sum_{n=-M}^{\infty} (ik)^n f_{\gamma,n}^{sc}(\omega,\omega'), \quad 0 < |k| < k_0 \quad \text{small enough},$$
$$k\in\Pi_\pm, \quad \text{for } \gamma\gtrless 0, \tag{4.18}$$

where $M = 0$ in case I and $M = 2$ in case II. The first coefficients in case I read

$$f_{\gamma,0}^{sc}(\omega,\omega') = (vI_0([2\gamma(|\mathbf{x}| + \omega\cdot\mathbf{x})]^{1/2}), T_\gamma u I_0([2\gamma(|\mathbf{x}'| - \omega'\cdot\mathbf{x}')]^{1/2})), \quad \gamma > 0, \tag{4.19}$$

$$f_{\gamma,0}^{sc}(\omega,\omega') = (vJ_0([2|\gamma|(|\mathbf{x}| + \omega\cdot\mathbf{x})]^{1/2}), T_\gamma u J_0([2|\gamma|(|\mathbf{x}'| - \omega'\cdot\mathbf{x}')]^{1/2})), \quad \gamma < 0, \tag{4.20}$$

$$f_{\gamma,1}^{sc}(\omega,\omega') = (v\{|\mathbf{x}|I_0([2\gamma(|\mathbf{x}| + \omega\cdot\mathbf{x})]^{1/2}) - \gamma^{-1}[2\gamma(|\mathbf{x}| + \omega\cdot\mathbf{x})]^{1/2}I_1([2\gamma(|\mathbf{x}| + \omega\cdot\mathbf{x})]^{1/2})$$
$$- 2^{-1}[|\mathbf{x}| + \omega\cdot\mathbf{x}]I_2([2\gamma(|\mathbf{x}| + \omega\cdot\mathbf{x})]^{1/2})\}, T_\gamma u I_0([2\gamma(|\mathbf{x}'| - \omega'\cdot\mathbf{x}')]^{1/2}))$$
$$+ (vI_0([2\gamma(|\mathbf{x}| + \omega\cdot\mathbf{x})]^{1/2}), T_\gamma u\{|\mathbf{x}'|I_0([2\gamma(|\mathbf{x}'| - \omega'\cdot\mathbf{x}')]^{1/2})$$

$$-\gamma^{-1}[2\gamma(|\mathbf{x}'|-\omega'\cdot\mathbf{x}')]^{1/2}I_1([2\gamma(|\mathbf{x}'|-\omega'\cdot\mathbf{x}')]^{1/2})-2^{-1}[|\mathbf{x}'|-\omega'\cdot\mathbf{x}']I_2([2\gamma(|\mathbf{x}'|-\omega'\cdot\mathbf{x}')]^{1/2})\}),\quad \gamma>0, \tag{4.21}$$

$$\begin{aligned}f^{\mathrm{sc}}_{\gamma,1}(\omega,\omega') = {} & (v\{|\mathbf{x}|J_0([2|\gamma|(|\mathbf{x}|+\omega\cdot\mathbf{x})]^{1/2})-|\gamma|^{-1}[2|\gamma|(|\mathbf{x}|+\omega\cdot\mathbf{x})]^{1/2}J_1([2|\gamma|(|\mathbf{x}|+\omega\cdot\mathbf{x})]^{1/2})\\ & +2^{-1}[|\mathbf{x}|+\omega\cdot\mathbf{x}]J_2([2|\gamma|(|\mathbf{x}|+\omega\cdot\mathbf{x})]^{1/2})\},T_\gamma uJ_0([2|\gamma|(|\mathbf{x}'|-\omega'\cdot\mathbf{x}')]^{1/2}))\\ & +(vJ_0([2|\gamma|(|\mathbf{x}|+\omega\cdot\mathbf{x})]^{1/2}),T_\gamma u\{|\mathbf{x}'|J_0([2|\gamma|(|\mathbf{x}'|-\omega'\cdot\mathbf{x}')]^{1/2})-\gamma^{-1}[2|\gamma|(|\mathbf{x}'|-\omega'\cdot\mathbf{x}')]^{1/2}\\ & \times J_1([2|\gamma|(|\mathbf{x}'|-\omega'\cdot\mathbf{x}')]^{1/2})+2^{-1}[|\mathbf{x}'|-\omega'\cdot\mathbf{x}']J_2([2|\gamma|(|\mathbf{x}'|-\omega'\cdot\mathbf{x}')]^{1/2})\}),\quad \gamma<0.\end{aligned} \tag{4.22}$$

In case II we have

$$f^{\mathrm{sc}}_{\gamma,-2}(\omega,\omega')=\sum_{j,l=1}^{N}(\tilde{\phi}_\gamma,r_{\gamma,2}\phi_\gamma)^{-1}_{jl}(\tilde{\phi}_{\gamma,l},uI_0([2\gamma(|\mathbf{x}'|-\omega'\cdot\mathbf{x}')]^{1/2}))(vI_0([2\gamma(|\mathbf{x}|+\omega\cdot\mathbf{x})]^{1/2}),\phi_{\gamma,j}), \tag{4.23}$$

$$\begin{aligned}f^{\mathrm{sc}}_{\gamma,-1}(\omega,\omega') = {} & \sum_{j,l=1}^{N}(\tilde{\phi}_\gamma,r_{\gamma,2}\phi_\gamma)^{-1}_{jl}\{(\tilde{\phi}_{\gamma,l},u\{|\mathbf{x}'|I_0([2\gamma(|\mathbf{x}'|-\omega'\cdot\mathbf{x}')]^{1/2})-\gamma^{-1}[2\gamma(|\mathbf{x}'|-\omega'\cdot\mathbf{x}')]^{1/2}\\ & \times I_1([2\gamma(|\mathbf{x}'|-\omega'\cdot\mathbf{x}')]^{1/2})-2^{-1}[|\mathbf{x}'|-\omega'\cdot\mathbf{x}']I_2([2\gamma(|\mathbf{x}'|-\omega'\cdot\mathbf{x}')]^{1/2})\})(vI_0([2\gamma(|\mathbf{x}|+\omega\cdot\mathbf{x})]^{1/2}),\phi_{\gamma,j})\\ & +(\tilde{\phi}_{\gamma,l},uI_0([2\gamma(|\mathbf{x}'|-\omega'\cdot\mathbf{x}')]^{1/2}))(v\{|\mathbf{x}|I_0([2\gamma(|\mathbf{x}|+\omega\cdot\mathbf{x})]^{1/2})\\ & -\gamma^{-1}[2\gamma(|\mathbf{x}|+\omega\cdot\mathbf{x})]^{1/2}I_1([2\gamma(|\mathbf{x}|+\omega\cdot\mathbf{x})]^{1/2})-2^{-1}[|\mathbf{x}|+\omega\cdot\mathbf{x}]I_2([2\gamma(|\mathbf{x}|+\omega\cdot\mathbf{x})]^{1/2})\},\phi_{\gamma,j})\}.\end{aligned} \tag{4.24}$$

Proof: Expansion (4.18) follows from Lemmas 2.4 and 4.4 and Eq. (4.17). The coefficients (4.19)–(4.24) are directly obtained after inserting Eqs. (2.28), (2.29), and (4.9)–(4.11) into (4.17). □

Remark 4.6: Obviously, assuming that for all $|\gamma|<\epsilon$, ϵ small enough, we are in case I, the expressions for $f^{\mathrm{sc}}_{\gamma,n}(\omega,\omega')$ derived in Theorem 4.5 tend to the corresponding short-range quantities[1] as $\gamma\to 0$.

V. COULOMB-MODIFIED LOW-ENERGY PARAMETERS

We now want to find the generalization of the Coulomb-modified low-energy parameters to the nonspherically symmetric case and at the same time relate them to $f^{\mathrm{sc}}(k,\omega,\omega')$, like we have done[1] for the short-range case. Therefore, we first look at the (Coulomb-modified) effective range expansion.

If V is spherically symmetric and obeys

$$|V(r)|\leqslant cr^{-1}e^{-br},\quad \text{for some } b>0, \tag{5.1}$$

then f^{sc} has the absolutely convergent expansion[40,41]

$$f^{\mathrm{sc}}(k\omega,\omega')=4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{l}e^{2i\delta_l^c(k)}\frac{e^{2i\delta_l^{\mathrm{sc}}(k)}-1}{2ik}Y_{l,m}(\omega)\,\overline{Y_{l,m}(\omega')},\quad k>0, \tag{5.2}$$

where

$$\delta^c_l(k)=\arg\Gamma(l+1+i\gamma/2k),\quad l=0,1,\dots,k>0 \tag{5.3}$$

denote the pure Coulomb phase shifts and $\delta^{\mathrm{sc}}_l(k)$ the Coulomb-modified nuclear phase shifts. Furthermore, restricting ourselves to s waves ($l=0$), we obtain the Coulomb-modified effective range expansion

$$(\pi\gamma/k)(e^{\pi\gamma/k}-1)^{-1}[k\cot\delta^{\mathrm{sc}}_l(k)-ik]+\gamma h(\gamma,k)\underset{k\to 0_+}{=}-(a^{\mathrm{sc}}_0)^{-1}+2^{-1}r^{\mathrm{sc}}_0k^2+O(k^4). \tag{5.4}$$

The left-hand side of (5.4) is real-analytic with respect to k^2 around $k^2=0$ (Refs. 5 and 6). In Eq. (5.4) a^{sc}_0 and r^{sc}_0 denote the s-wave Coulomb-modified scattering length and effective range parameter, respectively, and $h(\gamma,k)$ has been introduced in Eq. (2.19).

Consequently, to generalize a^{sc}_0 and r^{sc}_0 to nonspherically symmetric interactions V we study the expression $F(\gamma,k)$ defined by

$$F(\gamma,k)=\left\{(\pi\gamma/k)^{-1}(e^{\pi\gamma/k}-1)e^{-2i\delta^c_0(k)}(4\pi)^{-2}\int_{S^2\times S^2}d\omega\,d\omega' f^{\mathrm{sc}}(k,\omega,\omega')\right\}^{-1}+\gamma h(\gamma,k),\quad k>0,\quad k^2\notin\mathscr{E}_\gamma, \tag{5.5}$$

which coincides with the left-hand side of Eq. (5.4), in case V is spherically symmetric. Inserting

$$u(\mathbf{x})\Psi^-_c(k,\omega',\mathbf{x})=\operatorname*{s-lim}_{L\to\infty}4\pi\sum_{l=0}^{L}\sum_{m=-l}^{l}(2ik)^l\Gamma(2l+2)^{-1}e^{-\pi\gamma/4k}\Gamma(l+1+i\gamma/2k)u(\mathbf{x})r^{-1}F^{(0)}_l(k,r)\,\overline{Y_{l,m}(\omega')}\,Y_{l,m}(\omega_{\mathbf{x}}), \tag{5.6}$$

together with Eq. (4.17) into Eq. (5.5), we get after doing the angular integrations

$$F(\gamma,k)=-[(4\pi)^{-1}(vr^{-1}F^{(0)}_0(k,r),(uG_{\gamma,k}v+1)^{-1}ur'^{-1}F^{(0)}_0(k,r'))]^{-1}+\gamma h(\gamma,k),\quad k^2\notin\mathscr{E}_\gamma, \tag{5.7}$$

where $F^{(0)}_l(k,r)$ is given by Eq. (2.16). Since $F^{(0)}_0(k,r)$ is entire with respect to k^2, Lemma 4.4 implies that $F(\gamma,k)$ has an asymptotic expansion of the type

$$F(\gamma,k) \underset{k\to 0}{\sim} -(\tilde{a}^{sc})^{-1} + 2^{-1}\tilde{r}^{sc}k^2 + O(k^4), \quad 0<k<k_0 \text{ small enough}, \quad \gamma\in\mathbb{R}\setminus\{0\}. \tag{5.8}$$

We recall that in case II, assumption A is supposed to hold. Next we calculate

$$\begin{aligned} F(\gamma,k) - \overline{F(\gamma,k)} = {} & 2\pi i\gamma(e^{\pi\gamma/k}-1)^{-1} - (ik/2\pi)(2ik)^{-1}(vr^{-1}F_0^{(0)}(k,r),(uG_{\gamma,k}v+1)^{-1}u(G_{\gamma,k}-G^*_{\gamma,k}) \\ & \times v(uG^*_{\gamma,k}v+1)^{-1}ur'^{-1}F_0^{(0)}(k,r')) \\ & \times(4\pi)^2|(vr^{-1}F_0^{(0)}(k,r),(uG_{\gamma,k}v+1)^{-1}ur'^{-1}F_0^{(0)}(k,r'))|^{-2}, \quad k^2\notin\mathscr{E}_\gamma, \end{aligned} \tag{5.9}$$

where

$$\gamma h(\gamma,k) - \overline{\gamma h(\gamma,h)} = 2\pi i\gamma(e^{\pi\gamma/k}-1)^{-1} \tag{5.10}$$

has been used. Employing the relation (cf. the Appendix)

$$(4\pi)^{-2}\int_{S^2} d\omega''\, \Psi_c^-(k,\omega'',\mathbf{x})\,\overline{\Psi_c^-(k,\omega'',\mathbf{y})} = (2ik)^{-1}\left[G_{\gamma,k}(\mathbf{x},\mathbf{y}) - \overline{G_{\gamma,k}(\mathbf{x},\mathbf{y})}\right] \tag{5.11}$$

in Eq. (5.9) the right-hand side of the latter becomes

$$\begin{aligned} & 2\pi i\gamma(e^{\pi\gamma/2k}-1)^{-1} \\ & \quad -(ik/2\pi)(4\pi)^{-2}\int_{S^2} d\omega'' \quad |(vr^{-1}F_0^{(0)}(k,r),(uG_{\gamma,k}v+1)^{-1}u\Psi_c^-(k,\omega''))|^2 \\ & \quad \times(4\pi)^2|(vr^{-1}F_0^{(0)}(k,r),(uG_{\gamma,k}v+1)^{-1}ur'^{-1}F_0^{(0)}(k,r'))|^{-2}, \quad k^2\notin\mathscr{E}_\gamma. \end{aligned} \tag{5.12}$$

Furthermore, inserting Eq. (5.6) into Eq. (5.12) and integrating over $d\omega''$, we arrive at

$$\begin{aligned} F(\gamma,k) - \overline{F(\gamma,k)} = {} & (-ik/2\pi)e^{-\pi\gamma/2k}\sum_{l=1}^{\infty}\sum_{m=-l}^{l} |(vr^{-1}F_0^{(0)}(k,r),(uG_{\gamma,k}v+1)^{-1}(2k)^l\Gamma(2l+2)^{-1}|\Gamma(l+1+i\gamma/2k)| \\ & \times ur'^{-1}F_l^{(0)}(k,r')Y_{l,m}(\omega_{\mathbf{x}'}))|^2(4\pi)^2|(vr^{-1}F_0^{(0)}(k,r),(uG_{\gamma,k}v+1)^{-1}ur'^{-1}F_0^{(0)}(k,r'))|^{-2}, \quad k^2\notin\mathscr{E}_\gamma. \end{aligned} \tag{5.13}$$

Finally we look at the small k behavior of these quantities. Inserting expansion (2.28) and (2.29) into expression (5.12) one obtains in case I

$$\begin{aligned} F(\gamma,k) - \overline{F(\gamma,k)} \underset{k\to 0}{\sim} {} & 2\pi i\gamma(e^{\pi\gamma/k}-1)^{-1} - (i\gamma/2)(e^{\pi\gamma/k}-1)^{-1} \\ & \times\left\{\int_{S^2} d\omega''\left|\left(v\begin{bmatrix}(\gamma r)^{-1/2}I_1(2(\gamma r)^{1/2}), & \gamma>0 \\ (|\gamma|r)^{-1/2}J_1(2(|\gamma|r)^{1/2}), & \gamma<0\end{bmatrix},(uG_{\gamma,0}v+1)^{-1}u\begin{bmatrix}I_0([2\gamma(|\mathbf{x}'|-\omega''\cdot\mathbf{x}')]^{1/2}), & \gamma>0 \\ J_0([2|\gamma|(|\mathbf{x}'|-\omega''\cdot\mathbf{x}')]^{1/2}), & \gamma<0\end{bmatrix}\right)\right|^2\right. \\ & \times\left.\left|\left(v\begin{bmatrix}(\gamma r)^{-1/2}I_1(2(\gamma r)^{1/2}), & \gamma>0 \\ (|\gamma|r)^{-1/2}J_1(2(|\gamma|r)^{1/2}), & \gamma<0\end{bmatrix},(uG_{\gamma,0}v+1)^{-1}u\begin{bmatrix}(\gamma r')^{-1/2}I_1(2(\gamma r')^{1/2}), & \gamma>0 \\ (|\gamma|r')^{-1/2}J_1(2(|\gamma|r')^{1/2}), & \gamma<0\end{bmatrix}\right)\right|^{-2} + O(k^2)\right\} \end{aligned} \tag{5.14}$$

$$\underset{k\to 0}{\sim} -(i\gamma/2)(e^{\pi\gamma/k}-1)^{-1}[c_\gamma + O(k^2)], \quad c_\gamma\geqslant 0, \tag{5.15}$$

$$\underset{k\to 0}{\sim}\begin{cases} O(e^{-\pi\gamma/k}), & \gamma>0 \\ O(1), & \gamma<0, \quad \text{in case I,}\end{cases} \tag{5.16}$$

by doing explicitly the $d\omega''$ integration with help of Ref. 42, pp. 87 and 88. Similarly in case II

$$F(\gamma,k) - \overline{F(\gamma,k)} \underset{k\to 0}{\sim} O(e^{-\pi\gamma/k}). \tag{5.17}$$

At this point it is instructive to compare with the short-range situation ($\gamma=0$). There the analog of Eq. (5.13) and its small k limit reads

$$\begin{aligned} F(0,k) - \overline{F(0,k)} = {} & (-ik/2\pi)\sum_{l=1}^{\infty}\sum_{m=-l}^{l}\left|\left(v\frac{\sin kr}{kr},(uG_kv+1)^{-1}uj_l(kr')Y_{l,m}(\omega_{\mathbf{x}'})\right)\right|^2 \\ & \times(4\pi)^2\left|\left(v\frac{\sin kr}{kr},(uG_kv+1)^{-1}u\frac{\sin kr'}{kr'}\right)\right|^{-2} = O(k^3), \quad k^2\notin\mathscr{E}_0 \end{aligned} \tag{5.18}$$

in all cases I–IV (cf. Ref. 1 for the case distinctions when $\gamma=0$).

From these results we see that $F(\gamma,k)$ is in general complex valued. The reason for this becomes clear from Eqs. (5.12) and (5.13): The first term in Eq. (5.12) [being identical to $\gamma h(\gamma,k) - \gamma\,\overline{h(\gamma,k)}$; cf. Eq. (2.19)] precisely cancels the s-wave part of $(2ik)^{-1}[G_{\gamma,k} - \overline{G_{\gamma,k}}]$. In other words, the term $\gamma h(\gamma,k)$ in Eq. (5.5) exactly subtracts the s-wave cut contribution of $G_{\gamma,k}$. But the long-range nature of the Coulomb potential $\gamma/|\mathbf{x}|$ obviously dominates $l(l+1)/r^2$ as $r\to\infty$ for all $l=0,1,2,\ldots$. Therefore, in order to get real low-energy parameters one is forced to first subtract the cut contribution of $G_{\gamma,k}$ for *all* $l=0,1,2,\ldots$, in Eq. (5.5). By calculations presented above this corresponds precisely to a replacement of $F(\gamma,k)$ by its real part. Consequently we obtain in general

$$\operatorname{Re} F(\gamma,k) \underset{k\to 0}{\sim} -(a^{sc})^{-1} + 2^{-1}r^{sc}k^2 + O(k^4), \quad 0<k<k_0 \text{ small enough}, \quad \gamma\in\mathbb{R}\setminus\{0\} \tag{5.19}$$

as the analog of the expansion (5.4). This leads to the following.

Definition 5.1: Let $e^{2a|x|}V\in R$ for some $a>0$ and suppose assumption A holds. Then the Coulomb-modified scattering length a^{sc} is defined as

$$(a^{sc})^{-1} = -\lim_{k\to 0_+} \operatorname{Re}\left\{\left[\left(\frac{\pi\gamma}{k}\right)^{-1}(e^{\pi\gamma/k}-1)e^{-2i\delta_0^c(k)}(4\pi)^{-2}\int_{S^2\times S^2} d\omega\, d\omega' f^{sc}(k,\omega,\omega')\right]^{-1}\right\}, \quad 0<k<k_0 \text{ small enough},$$

$$\gamma\neq 0 \quad \text{in cases I and II.} \tag{5.20}$$

The Coulomb-modified effective range parameter r^{sc} is then defined to be

$$r^{sc} = \frac{2}{3\gamma} + \lim_{k\to 0} 2k^{-2}\left\{\operatorname{Re}\left[\left(\frac{\pi\gamma}{k}\right)^{-1}(e^{\pi\gamma/k}-1)e^{-2i\delta_0^c(k)}(4\pi)^{-2}\int_{S^2\times S^2} d\omega\, d\omega' f^{sc}(k,\omega,\omega')\right]^{-1} + (a^{sc})^{-1}\right\},$$

$$0<k<k_0 \text{ small enough}, \quad \gamma\neq 0 \quad \text{in cases I and II.} \tag{5.21}$$

Definitions (5.20) and (5.21) are clear from Eqs. (5.5), (5.19), and

$$\operatorname{Re}[\gamma h(\gamma,k)] = k^2/3\gamma + O(k^4) \tag{5.22}$$

[cf. expansion (2.21)].

Remark 5.2: Obviously if V is spherically symmetric the right-hand side of Eq. (5.13) is identically zero and hence $F(\gamma,k)$ and the coefficients $\tilde{a}^{sc}$, $\tilde{r}^{sc}$ in (5.8) are real. In case I for $\gamma>0$ and in case II, the imaginary part of $F(\gamma,k)$ is of order $O(e^{-\pi\gamma/k})$ [cf. Eqs. (5.16) and (5.17)] such that $\tilde{a}^{sc}$ and $\tilde{r}^{sc}$ are again real. So in all these cases the real part in definitions (5.20) and (5.21) can be dropped. The same is true in the short-range case $\gamma=0$. Due to the fact that $\operatorname{Im}[F(0,k)]$ is at least of order $O(k^3)$ [cf. Eq. (5.18)], the low-energy parameters $\tilde{a}^s$ and $\tilde{r}^s$ (the analog of $\tilde{a}^{sc}$ and $\tilde{r}^{sc}$ if $\gamma=0$) are real and hence coincide with a^s and r^s (Ref. 1). Only for $\gamma<0$ in case I is the real part in definitions (5.20) and (5.21) needed and should be added to the corresponding expressions in Sec. IV of Ref. 1.

Given Definition 5.1 we finally provide explicit expressions for a^{sc} and r^{sc}, thereby extending the results of Ref. 1.

Theorem 5.3: Let $e^{2a|x|}V\in R$ for some $a>0$ and suppose assumption A holds. Then in case I we have

$$a^{sc} = (4\pi)^{-1}(v(\gamma r)^{-1/2}I_1((4\gamma r)^{1/2}),T_\gamma u(\gamma r')^{-1/2}I_1((4\gamma r')^{1/2})), \quad \gamma>0, \tag{5.23}$$

$$\begin{aligned} r^{sc} = 2/3\gamma + (2\pi)^{-1}(a^{sc})^{-2}[&(v(\gamma r)^{-1/2}I_1((4\gamma r)^{1/2}),T_\gamma r_{\gamma,2}T_\gamma u(\gamma r')^{-1/2}I_1((4\gamma r')^{1/2})\\ &-(vrI_2((4\gamma r)^{1/2}),T_\gamma u(\gamma r')^{-1/2}I_1((4\gamma r')^{1/2}))/3\gamma - (v(\gamma r)^{-1/2}I_1((4\gamma r)^{1/2}),T_\gamma ur'I_2((4\gamma r')^{1/2}))/3\gamma], \quad \gamma>0, \end{aligned} \tag{5.24}$$

$$(a^{sc})^{-1} = \operatorname{Re}\{[(4\pi)^{-1}(v(|\gamma|r)^{-1/2}J_1((4|\gamma|r)^{1/2}),T_\gamma u(|\gamma|r')^{-1/2}J_1((4|\gamma|r')^{1/2}))]^{-1}\}, \quad \gamma<0, \tag{5.25}$$

$$\begin{aligned} r^{sc} = 2/3\gamma - 8\pi&|(v(|\gamma|r)^{-1/2}J_1((4|\gamma|r)^{1/2}),T_\gamma u(|\gamma|r')^{-1/2}J_1((4|\gamma|r')^{1/2}))|^{-2}\\ &\times\operatorname{Re}\{(v(|\gamma|r)^{-1/2}J_1((4|\gamma|r)^{1/2}),T_\gamma r_{\gamma,2}T_\gamma u(|\gamma|r')^{-1/2}J_1((4|\gamma|r')^{1/2})) - (vrJ_2((4|\gamma|r)^{1/2}),T_\gamma u(|\gamma|r')^{-1/2}\\ &\times J_1((4|\gamma|r')^{1/2}))/3|\gamma| - (v(|\gamma|r)^{-1/2}J_1((4|\gamma|r)^{1/2}),T_\gamma ur'J_2((4|\gamma|r')^{1/2}))/3|\gamma|\}, \quad \gamma<0. \end{aligned} \tag{5.26}$$

In case II, if $(v(\gamma r)^{-1/2}I_1((4\gamma r)^{1/2}),\phi_{\gamma,j_0})\neq 0$ for some $1\leqslant j_0\leqslant N$, we get

$$(a^{sc})^{-1} = 0, \tag{5.27}$$

$$r^{sc} = 2/3\gamma + 8\pi\left[\sum_{j,l=1}^{N}(\tilde{\phi}_\gamma,r_{\gamma,2}\phi_\gamma)_{jl}^{-1}(\tilde{\phi}_{\gamma,l},u(\gamma r')^{-1/2}I_1((4\gamma r')^{1/2}))(v(\gamma r)^{-1/2}I_1((4\gamma r)^{1/2}),\phi_{\gamma,j})\right]^{-1}, \tag{5.28}$$

assuming the existence of $(\tilde{\phi},r_{\gamma,2}\phi)_{jl}^{-1}$ (cf. Lemma 4.4). If $(v(\gamma r)^{-1/2}I_1((4\gamma r)^{1/2}),\ \phi_{\gamma,j})=0$ for all $1\leqslant j\leqslant N$ we obtain that a^{sc} is again given by Eq. (5.23) [with T_γ defined in Eq. (4.7)] and r^{sc} is given by

$$\begin{aligned} r^{sc} = 2/3\gamma + (2\pi)^{-1}(a^{sc})^{-2}\{&(v(\gamma r)^{-1/2}I_1((4\gamma r)^{1/2}),T_\gamma r_{\gamma,2}T_\gamma u(r'\gamma)^{-1/2}I_1((4\gamma r')^{1/2}))\\ &-(v(\gamma r)^{-1/2}I_1((4\gamma r)^{1/2}),T_\gamma a_{\gamma,-2}T_\gamma u(\gamma r')^{-1/2}I_1((4\gamma r')^{1/2})) - [(vrI_2((4\gamma r)^{1/2}),T_\gamma u(\gamma r')^{-1/2}I_1((4\gamma r')^{1/2}))\\ &+(v(\gamma r)^{-1/2}I_1((4\gamma r)^{1/2}),T_\gamma ur'I_2((4\gamma r')^{1/2}))]/3\gamma + [(vrI_2((4\gamma r)^{1/2}),a_{\gamma,-2}r_{\gamma,2}T_\gamma u(\gamma r')^{-1/2}I_1((4\gamma r')^{1/2}))\\ &+(v(\gamma r)^{-1/2}I_1((4\gamma r)^{1/2}),T_\gamma r_{\gamma,2}a_{\gamma,-2}ur'I_2((4\gamma r')^{1/2}))]/3\gamma - (vrI_2((4\gamma r)^{1/2}),a_{\gamma,-2}ur'I_2((4\gamma r')^{1/2}))/9\gamma^2\}, \end{aligned} \tag{5.29}$$

with

$$a_{\gamma,-2} = \sum_{j,l=1}^{N}(\tilde{\phi}_\gamma,r_{\gamma,2}\phi_\gamma)_{jl}^{-1}(\tilde{\phi}_{\gamma,l},\cdot)\phi_{\gamma,j}, \tag{5.30}$$

assuming again the existence of $(\tilde{\phi},r_{\gamma,2}\phi)_{jl}^{-1}$.

Proof: Expressions (5.23)–(5.29) follow after inserting expansion (4.18) into definitions (5.20)and (5.21), integrating over $d\omega\, d\omega'$. □

Remark 5.4: (a) If V is spherically symmetric and $\phi_{\gamma,j}$, $1\leqslant j\leqslant N$ belong to angular momentum subspaces indexed by $l\geqslant 1$ then obviously $(v(\gamma r)^{-1/2}I_1((4\gamma r)^{1/2}),\ \phi_{\gamma,j})=0$ for all $1\leqslant j\leqslant N$ and the last part of Theorem 5.3 applies.

(b) By inspection one can show that a^{sc} and r^{sc} are continuous at $\gamma=0$ (as conjectured in Ref. 1). Furthermore, they converge to the corresponding short-range quantities a^s and r^s if we assume, e.g., that we are in case I for all $|\gamma|<\epsilon$, ϵ small enough. In particular, one derives

$$(a^{sc})^{-1}=(a^s)^{-1}+\gamma[\ln\gamma+2C-1]-\gamma(a^s)^{-2}(8\pi)^{-1}\times\{(v|\mathbf{x}|,T_0u)+(v,T_0u|\mathbf{x}'|)-(v,T_0u\ln(x_+/2)v\times T_0u)/2\pi\}+O(\gamma^2\ln\gamma),\quad \gamma>0, \tag{5.31}$$

where C denotes Euler's constant[22] and $u\ln(x_+/2)v$ denotes the Hilbert–Schmidt operator with kernel

$$u(\mathbf{x})\ln[(|\mathbf{x}|+|\mathbf{y}|+|\mathbf{x}-\mathbf{y}|)/2]v(\mathbf{y}).$$

Such a formula has also been written down for finite-size particle scattering in the spherically symmetric case in Refs. 5 and 6.

For discussions on charge symmetry of nuclear forces in connection with expansion (5.31) we refer to Refs. 29, 43, and 44 and the literature cited therein.

ACKNOWLEDGMENTS

We are indebted to Professor L. Streit and Professor H. Zingl for many helpful discussions. Two of us (D. B. and F. G.) would like to thank Professor L. Streit for his warm hospitality at the Zentrum für interdisziplinäre Forschung der Universität Bielefeld, Federal Republic of Germany during Project No. 2: Mathematics + Physics. F. G. and W. S. thank the Institute for Theoretical Physics of the University of Leuven, Belgium, for hospitality during the first stages of this work.

F. G. gratefully acknowledges financial support by the Alexander von Humboldt Foundation.

APPENDIX: THE EXCEPTIONAL SET $\mathscr{E}_\gamma$

Following the treatment in Ref. 34 we sketch a possible way to verify that every nonzero point $k_0^2\in\mathscr{E}_\gamma$ belongs to the point spectrum of H. Throughout this Appendix we assume $e^{2a|\mathbf{x}|}V\in R$ for some $a>0$ (this could be relaxed in the following) and

$$uG_{\gamma,k_0}v\phi_{\gamma,k_0}=-\phi_{\gamma,k_0},\quad \phi_{\gamma,k_0}\in L^2(\mathbb{R}^3),\quad k_0^2\in\mathscr{E}_\gamma,\quad k_0>0. \tag{A1}$$

We first introduce the spectral representation associated with $H_cP_{ac}(H_c)$, $P_{ac}(H_c)$ being the projection onto the absolutely continuous subspace corresponding to H_c

$$U_c:\begin{cases}P_{ac}(H_c)L^2(\mathbb{R}^3)\to L^2((0,\infty),dk;L^2(S^2)),\\ g(\mathbf{x})\to \underset{R\to\infty}{\text{s-lim}}(2\pi)^{-3/2}k\int_{|\mathbf{x}|<R}d^3x\ \overline{\Psi_c^-(k,\omega,\mathbf{x})}g(\mathbf{x})\end{cases}$$

$$k>0,\quad \omega\in S^2. \tag{A2}$$

We then state the following.

Lemma A.1: Let ϕ_{γ,k_0} be as above; then

$$(U_cv\phi_{\gamma,k_0})(k_0,\omega)=0,\ \text{in } L^2(S^2). \tag{A3}$$

Moreover, introducing

$$(M_v^c(k)g)(\omega)=(U_cvg)(k,\omega),\quad g\in L^2(\mathbb{R}^3),\quad k>0,\quad \omega\in S^2, \tag{A4}$$

then $M_v^c(k)\in\mathscr{B}_2(L^2(\mathbb{R}^3),L^2(S^2))$ and $M_v^c(k)$ is continuous in the Hilbert-Schmidt norm with respect to $k>0$.

Proof: From Eq. (2.32) and from Ref. 10

$${}_1F_1(\alpha;\beta;z)=[\Gamma(\alpha)\Gamma(\beta-\alpha)]^{-1}\Gamma(\beta)\int_0^1 dt\,e^{zt}t^{\alpha-1}\times(1-t)^{\beta-\alpha-1},\quad \operatorname{Re}\beta>\operatorname{Re}\alpha>0,$$

one obtains for all $a>0$

$$|e^{-a|\mathbf{x}|}\Psi_c^-(k,\omega,\mathbf{x})|\leqslant c(k),\quad k>0.$$

Thus the kernel $M_v^c(k,\omega,\mathbf{x})$ of $M_v^c(k)$ obeys

$$M_v^c(k,\omega,\mathbf{x})=(2\pi)^{-3/2}k\ \overline{\Psi_c^-(k,\omega,\mathbf{x})}v(\mathbf{x})\in L^2(S^2\times\mathbb{R}^3),\quad k>0$$

and

$$\|M_v^c(k)-M_v^c(k')\|_2^2=(2\pi)^{-3}\int_{S^2\times\mathbb{R}^3}d\omega\,d^3x\,|k\Psi_c^-(k,\omega,\mathbf{x})-k'\Psi_c^-(k',\omega,\mathbf{x})|^2|V(\mathbf{x})|\underset{k'\to k}{\longrightarrow}0,$$

by dominated convergence. Next we compute

$$\|(U_cv\phi_{\gamma,k_0})(k_0)\|^2_{L^2(S^2)}=(M_v^c(k_0)^*M_v^c(k_0)\phi_{\gamma,k_0},\phi_{\gamma,k_0})=(ik_0/\pi)([vG_{\gamma,k_0}v-vG^*_{\gamma,k_0}v]\phi_{\gamma,k_0},\phi_{\gamma,k_0}). \tag{A5}$$

Here we have used Eq. (5.11). A simple proof of Eq. (5.11) can be obtained from the partial wave decomposition (5.6) for Ψ_c^- and from

$$G_{\gamma,k}(\mathbf{x},\mathbf{y})=(|\mathbf{x}|\,|\mathbf{y}|)^{-1}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}g_l^{(0)}(-k,|\mathbf{x}|,|\mathbf{y}|)\times Y_{l,m}(\omega_{\mathbf{x}})\overline{Y_{l,m}(\omega_{\mathbf{y}})},$$

$$g_l^{(0)}(-k,r,r')=\begin{cases}G_l^{(0)}(-k,r)F_l^{(0)}(k,r'), & r'\leqslant r,\\ G_l^{(0)}(-k,r')F_l^{(0)}(k,r), & r'\geqslant r.\end{cases} \tag{A6}$$

Finally, Eq. (A5) implies

$$\|(U_cv\phi_{\gamma,k_0})(k_0)\|^2_{L^2(S^2)}=-(2k_0/\pi)\operatorname{Im}(vG_{\gamma,k_0}v\phi_{\gamma,k_0},\phi_{\gamma,k_0})=(2k_0/\pi)\operatorname{Im}(\phi_{\gamma,k_0},(\operatorname{sgn}V)\phi_{\gamma,k_0})=0.\qquad\square$$

Lemma A.2: Let ϕ_{γ,k_0} be as above and define

$$\hat{\psi}_c(k) = (k^2 - k_0^2)^{-1} M_v^c(k)\phi_{\gamma,k_0} \in L^2(S^2),$$
$$k > 0, \quad k \neq k_0. \tag{A7}$$

Then

$$\|\hat{\psi}_c(k)\|_{L^2(S^2)} \in L^2((0,\infty); dk)$$

and hence

$$\psi_{\gamma,k_0} = U_c^{-1}\hat{\psi}_c \in L^2(\mathbb{R}^3). \tag{A8}$$

Proof: One can follow the proof of Lemma 10.15 of Ref. 34 step by step. □

From the fact that

$$U_c^{-1}: \begin{cases} L^2((0,\infty); dk; L^2(S^2)) \rightarrow P_{ac}(H_c) L^2(\mathbb{R}^3), \\ h(k,\omega) \rightarrow \underset{R\to\infty}{\text{s-lim}} (2\pi)^{-3/2} \int_0^R dk \int_{S^2} d\omega\, \Psi_c^-(k,\omega,\mathbf{x}) \\ \qquad \times kh(k,\omega), \end{cases} \tag{A9}$$

one infers that

$$u\psi_{\gamma,k_0} = \underset{\epsilon\to 0_+}{\text{s-lim}}\, u(H_c - k^2 - i\epsilon)^{-1} v\phi_{\gamma,k_0} = -\phi_{\gamma,k_0} \tag{A10}$$

and

$$\psi_{\gamma,k_0} = G_{\gamma,k_0} v\phi_{\gamma,k_0}, \tag{A11}$$

by using the eigenfunction expansion of H_c in terms of ψ_c^-. Employing the explicit expression of $G_{\gamma,k_0}(\mathbf{x},\mathbf{y})$ given by Eqs. (2.15)–(2.17) one can show after a tedious calculation that $\nabla\psi_{\gamma,k_0} \in L^2_{\text{loc}}(\mathbb{R}^3)$. Since

$$(-\Delta + \gamma|\mathbf{x}|^{-1} + V - k_0^2)\psi_{\gamma,k_0} = 0, \tag{A12}$$

in the sense of distributions, it remains to prove $\nabla\psi_{\gamma,k} \in L^2(\mathbb{R}^3)$ [i.e., $\psi_{\gamma,k_0} \in H^{2,1}(\mathbb{R}^3)$] in order to conclude (Ref. 9, Corollary 2.8)

$$\psi_{\gamma,k_0} \in \mathscr{D}(H) \quad \text{and}\ H\psi_{\gamma,k_0} = k_0^2 \psi_{\gamma,k_0}. \tag{A13}$$

If H is given by an operator sum [i.e., V is bounded with respect to H_c with relative bound smaller than one and hence $\mathscr{D}(H) = \mathscr{D}(H_0)$] then (A13) results by mimicking the proof of Proposition 10.17 of Ref. 34. If $V \in R$ and, e.g., supp V is compact, one can argue as follows: Equation (A3) is equivalent to

$$\int_{\mathbb{R}^3} d^3y\, \overline{\Psi_c^-(k_0,\omega,\mathbf{y})} v(\mathbf{y})\phi_{\gamma,k_0}(\mathbf{y}) = 0,$$
$$\text{for a.e. } \omega \in S^2, \quad k_0 > 0, \tag{A14}$$

and hence by expansion (5.6)

$$\int_{\mathbb{R}^3} d^3y\, |\mathbf{y}|^{-1} F_l^{(0)}(k_0,|\mathbf{y}|)\, \overline{Y_{l,m}(\omega_{\mathbf{y}})}\, v(\mathbf{y})\phi_{\gamma,k_0}(\mathbf{y}) = 0,$$
$$k_0 > 0, \quad l = 0,1,2,\ldots. \tag{A15}$$

On the other hand, Eqs. (A6) and (A11) imply for $\mathbf{x}$ outside supp V

$$\psi_{\gamma,k_0}(\mathbf{x}) = \underset{L\to\infty}{\text{s-lim}}\, |\mathbf{x}|^{-1} \sum_{l=0}^{L} \sum_{m=-l}^{l} G_l^{(0)}(-k_0,|\mathbf{x}|) Y_{l,m}(\omega_{\mathbf{x}})$$
$$\times \int_{\mathbb{R}^3} d^3y\, |\mathbf{y}|^{-1} F_l^{(0)}(k_0,|\mathbf{y}|)\, \overline{Y_{l,m}(\omega_{\mathbf{y}})}\, v(\mathbf{y})\phi_{\gamma,k_0}(\mathbf{y})$$
$$= 0, \quad \text{for } |\mathbf{x}| \text{ large enough.} \tag{A16}$$

Thus supp ψ_{γ,k_0} is compact and hence $\nabla\psi_{\gamma,k_0} \in L^2(\mathbb{R}^3)$, implying the validity of (A13).

[1] S. Albeverio, D. Bollé, F. Gesztesy, R. Høegh-Krohn, and L. Streit, Ann. Phys. (NY) **148**, 308 (1983).
[2] R. G. Newton, *Scattering Theory of Waves and Particles* (Springer, New York, 1982); C. J. Joachain, *Quantum Collision Theory* (North-Holland, Amsterdam, 1975); R. Taylor, *Scattering Theory* (Wiley, New York, 1972).
[3] A. M. Badalyan, L. P. Kok, M. I. Polikarpov, and Yu. A. Simonov, Phys. Rep. **82**, 31 (1982).
[4] S. Albeverio, F. Gesztesy, and R. Høegh-Krohn, Ann. Inst. H. Poincaré A **37**, 1 (1982).
[5] D. Bollé and F. Gesztesy, Phys. Rev. Lett. **52**, 1469 (1984).
[6] D. Bollé and F. Gesztesy, Phys. Rev. A **30**, 1279 (1984).
[7] E. Lambert, Helv. Phys. Acta **42**, 667 (1968).
[8] L. Hostler, J. Math. Phys. **8**, 642 (1967).
[9] B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, NJ, 1971).
[10] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).
[11] J. C. Guillot, Indiana Univ. Math. J. **25**, 1105 (1976).
[12] F. Gesztesy and B. Thaller, J. Phys. A **14**, 639 (1981).
[13] F. Gesztesy, J. Math. Phys. **23**, 74 (1982).
[14] J. W. De Maag, L. P. Kok, and H. van Haeringen, J. Math. Phys. **25**, 684 (1984).
[15] H. Schmidt, Math Z. **43**, 539 (1937/1938).
[16] S. Klarsfeld, Nuovo Cimento A **48**, 1059 (1967).
[17] B. Simon, J. Funct. Anal. **25**, 338 (1977).
[18] M. Klaus and B. Simon, Ann. Phys. **130**, 251 (1980).
[19] M. Hoffmann-Ostenhof, T. Hoffman-Ostenhof, and B. Simon, J. Phys. A **16**, 1125 (1983).
[20] M. Klaus and B. Simon (in preparation).
[21] D. R. Yafaev, Sov. Phys. Dokl. **26**, 582 (1981).
[22] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).
[23] R. G. Newton, J. Math. Phys. **18**, 1348 (1977).
[24] A. Jensen and T. Kato, Duke Math. J. **46**, 583 (1979).
[25] A. Jensen, Duke. Math. J. **47**, 57 (1980).
[26] M. Murata, J. Funct. Anal. **49**, 10 (1982).
[27] M. Klaus, Helv. Phys. Acta **55**, 49 (1982).
[28] D. Bollé and S. F. J. Wilk, J. Math. Phys. **24**, 1555 (1983).
[29] S. Albeverio, F. Gesztesy, R. Høegh-Krohn, and L. Streit, Ann. Inst. H. Poincaré A **38**, 263 (1983).
[30] S. T. Kuroda, J. Anal. Math. **20**, 57 (1967).
[31] M. Reed and B. Simon, *Methods of Modern Mathematical Physics III, Scattering Theory* (Academic, New York, 1979).
[32] M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV, Analysis of Operators* (Academic, New York, 1978).
[33] J. Weidmann, Math. Z. **98**, 268 (1967).
[34] W. O. Amrein, J. M. Jauch, and K. B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, Reading, MA, 1977).
[35] J. Ginibre and M. Moulin, Ann. Inst. H. Poincaré A **21**, 97 (1974).
[36] M.S.P. Eastham and H. Kalf, "Schrödinger-Type Operators with Continuous Spectra," in *Research Notes in Mathematics*, Vol. 65 (Pitman, London, 1982).
[37] R. Froese, I. Herbst, M. Hoffmann-Ostenhof, and T. Hoffman-Ostenhof, J. Anal. Math. **41**, 272 (1982).

[38]H. Grosse, H. R. Grümm, H. Narnhofer, and W. Thirring, Acta Phys. Austriaca **40**, 97 (1974).
[39]W. Thirring, *A Course in Mathematical Physics 3, Quantum Mechanics of Atoms and Molecules* (Springer, New York, 1981).
[40]S. Klarsfeld, Nuovo Cimento A **43**, 1077 (1966).
[41]M. D. Semon and J. R. Taylor, Nuovo Cimento A **26**, 48 (1975).
[42]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, Berlin 1966), 3rd ed.
[43]S. Albeverio, L. S. Ferreira, F. Gesztesy, R. Høegh-Krohn, and L. Streit, Phys. Rev. C **29**, 680 (1984).
[44]F. Gesztesy, G. Karner, and L. Streit, " Charge particles with a short-range force: Perturbation theory with respect to the range and to additional effects", ZiF preprint Nr. 115, University of Bielefeld, 1984.

Transition matrix of point interactions as the scaling limit of integrable potentials on the real line

L. Pittner and M. Valjavec
Institut für Theoretische Physik, Universität Graz, Universitätplaz 5, A-8010 Graz, Austria

(Received 2 May 1984; accepted for publication 15 February 1985)

On the real line, the transition matrix corresponding to the nonrelativistic one-particle Hamilton operator for a finite number of zero-range interaction points is the scaling limit of the transition matrix for corresponding integrable potentials.

I. INTRODUCTION

The mathematically rigorous study of the Schrödinger equation with the potential describing so-called point or contact interactions is useful for models in various areas of physics, e.g., nuclear physics, solid state physics, elementary particle physics, and corresponding few-body and many-body problems, with the realistic potentials inserted only solvable numerically, can be evaluated explicitly in the model limit of point interactions.[1–13]

On the heuristic level,[14–19] these point interactions are described by "δ functions," and many of the emerging results are correct in the sense that they can be derived rigorously.[20–27] The most general starting point for such a rigorous definition of contact potentials is the "minimal Schrödinger operator"

$$\mathring{H}_{\mathscr{A}} := -\Delta + V, \quad V \in L^2_{\mathrm{loc}}(\mathscr{R}^n - \mathscr{A}),$$
$$\operatorname{dom} \mathring{H}_{\mathscr{A}} := \mathscr{C}_0^\infty(\mathscr{R}^n - \mathscr{A}),$$

where $\mathscr{A}$ denotes the closed exceptional set of Lebesgue measure zero. If the set $\mathscr{A}$ is bounded, then under certain additional assumptions the wave operators exist and are complete in the sense of Kato (Kuroda), i.e.,

$$\operatorname{ran} \Omega_\pm (H_{\mathscr{A}}, H_0) = \mathscr{H}_{\mathrm{ac}}(H_{\mathscr{A}}),$$

where $H_{\mathscr{A}}$ is any self-adjoint extension of $\mathring{H}_{\mathscr{A}}$, $\mathscr{H}_{\mathrm{ac}}(H_{\mathscr{A}})$ denotes the absolutely continuous subspace of $\mathscr{H} := L^2(\mathscr{R}^n)$ with respect to $H_{\mathscr{A}}$, and H_0 is the free Hamilton operator; the singularly continuous spectrum of $H_{\mathscr{A}}$ is empty.[28,29]

In the special case considered throughout this work, where N contact points on the real line $\mathscr{R}$ are considered, the minimal operator

$$\mathring{H}_{\mathscr{A}} := -\frac{d^2}{dx^2}, \quad \operatorname{dom} \mathring{H}_{\mathscr{A}} := \mathscr{C}_0^\infty(\mathscr{R} - \mathscr{A}),$$
$$\mathscr{A} := \{a_1, \dots, a_N\}, \tag{1.1}$$

can be closed to[30]

$$\overline{\mathring{H}}_{\mathscr{A}} = -\frac{d^2}{dx^2}, \quad \operatorname{dom} \overline{\mathring{H}}_{\mathscr{A}}$$
$$= \{f \in H^{2,2}(\mathscr{R});\ f(a_i) = 0,\ i = 1, \dots, N\}. \tag{1.2}$$

Its adjoint is the "maximal operator"

$$\mathring{H}^\dagger_{\mathscr{A}} = -\frac{d^2}{dx^2}, \quad \operatorname{dom} \mathring{H}^\dagger_{\mathscr{A}} = H^{2,2}(\mathscr{R} - \mathscr{A}) \cap H^{1,2}(\mathscr{R}); \tag{1.3}$$

here $H^{2,2}(\Omega)$ denotes the corresponding Sobolev space over an open subset Ω of $\mathscr{R}$ (see Ref. 31). All the self-adjoint extensions of $\mathring{H}_{\mathscr{A}}$ can then be parametrized by the boundary conditions[27,32]

$$H_\alpha := -\frac{d^2}{dx^2},$$
$$\operatorname{dom} H_\alpha := \{f \in H^{1,2}(\mathscr{R}) \cap H^{2,2}(\mathscr{R} - \mathscr{A});$$
$$f'(a_i + 0) - f'(a_i - 0) = \alpha_i f(a_i),$$
$$\alpha_i \in \mathscr{R},\ i = 1, \dots, N\}. \tag{1.4}$$

These Hamilton operators H_α are just the self-adjoint operators associated with the quadratic forms[27,32]

$$Q_\alpha(f, g) := \langle f' | g' \rangle + \sum_{i=1}^N \alpha_i \overline{f(a_i)} g(a_i),$$
$$\operatorname{dom} Q_\alpha := H^{1,2}(\mathscr{R}). \tag{1.5}$$

The difference of the resolvents of H_α and H_0 is an integral operator with the separable kernel

$$k_\alpha(x, x') := -\sum_{i=1}^N \sum_{j=1}^N g_k(x - a_i) \Gamma^k_{ij} g_k(x' - a_j),$$
$$\Gamma^k_{ij} := [T^{-1}]_{ij}, \quad T_{ij} := \delta_{ij}/\alpha_i + g_k(a_i - a_j),$$
$$\operatorname{Im} k > 0; \tag{1.6}$$

here g_k denotes the kernel of the free resolvent G_k,

$$g_k(x, x') := (i/2k) e^{ik|x - x'|},$$
$$G_k := (H_0 - k^2)^{-1}, \quad \operatorname{Im} k > 0. \tag{1.7}$$

This Hamilton operator H_α has at most N eigenvalues which are negative; its singularly continuous spectrum is empty; its absolutely continuous spectrum covers the positive real line.[32]

The aim of this work is to obtain the scattering theory for this N-point interaction via an appropriate limiting procedure from the scattering formalism for a rather wide class of potentials; here the unitary group of dilations

$$U_\epsilon f(x) := \epsilon^{-1/2} f(x/\epsilon), \quad x \in \mathscr{R}, \quad f \in \mathscr{H} := L^2(\mathscr{R}), \tag{1.8}$$

offers an elegant possibility.[32] The corresponding limit in three dimensions for integrable Rollnik potentials has been performed elsewhere in the sense of strong resolvent convergence.[33]

Here we start from the following scaled potentials:

$$H_\epsilon := H_0 \dot{+} \sum_{i=1}^{N} \epsilon^{-2}\lambda_i(\epsilon) V_i\left(\frac{x-a_i}{\epsilon}\right),$$
$$\epsilon > 0, \quad V_i \in L^1(\mathscr{R}), \tag{1.9}$$

where the real-valued functions λ_i are chosen to be holomorphic in a neighborhood of zero, with the value $\lambda_i(0) = 0$; the above sum of kinetic energy and potential is defined in the sense of quadratic forms.[34] For $\epsilon \to 0$, this Hamilton operator converges in the sense of norm resolvent to the contact interactions (1.4), with the "coupling constants"[32]

$$\alpha_i := \lambda_i'(0) \int_{\mathscr{R}} V_i(x) dx, \quad i = 1,...,N. \tag{1.10}$$

Correspondingly, we shall prove convergence of the scaled transition matrix to the transmission and reflection coefficients of the Hamilton operator (1.4) with N zero-range interaction centers.

In the appendices, we define the transition matrices for both the scaled potentials (1.9) and the point interactions (1.4) within the framework of time-independent scattering theory.

II. TRANSMISSION AND REFLECTION COEFFICIENTS FOR A FINITE NUMBER OF ZERO-RANGE INTERACTIONS

The one-dimensional quantum scattering problem on the positive real axis[35] or on the real line $\mathscr{R}$, respectively,[36–41] is usually described by transmission and reflection coefficients. The time-independent derivation of these coefficients shall be sketched in the appendices. Here we start from the scattering wave functions of the Hamilton operator H_α (see Ref. 19),

$$\Psi_k^\pm(x) := e^{\pm ikx} - \frac{i}{2|k|} \times \sum_{i=1}^{N}\sum_{j=1}^{N} \Gamma_{ij}^{|k|} e^{\pm ika_j} e^{i|k|\,|x-a_i|}, \quad x, k \in \mathscr{R}, \tag{2.1}$$

which solve the formal Lippmann–Schwinger equation

$$\Psi_k^\pm(x) = e^{\pm ikx} - \frac{i}{2|k|}\int_{\mathscr{R}} dy\, e^{i|k|\,|x-y|} \times \sum_{i=1}^{N} \alpha_i \delta(y-a_i) \Psi_k^\pm(y). \tag{2.2}$$

These waves describe one particle with momentum k incident from the left- $(+)$ or right- $(-)$ hand side and being transmitted through or reflected by the point interactions.

The corresponding transmission and reflection coefficients are then defined by the limits

$$t^\pm(k) := \lim_{x\to\pm\infty} e^{\mp ikx}\Psi_k^\pm(x),$$

$$r^\pm(k) := \lim_{x\to\mp\infty} e^{\pm ikx}(\Psi_k^\pm(x) - e^{\pm ikx}), \quad k \geqslant 0. \tag{2.3}$$

Inserting the scattering eigenfunctions (2.1) one obtains

$$t^+(k) = t^-(k) = 1 + \frac{1}{2ik}\sum_{i=1}^{N}\sum_{j=1}^{N} e^{-ika_i}\Gamma_{ij}^k e^{+ika_j}$$
$$=: t(k),$$
$$r^\pm(k) = \frac{1}{2ik}\sum_{i=1}^{N}\sum_{j=1}^{N} e^{\pm ika_i}\Gamma_{ij}^k e^{\pm ika_j}, \quad k \geqslant 0. \tag{2.4}$$

In the special case of one zero-range interaction center, the eigenfunctions and scattering coefficients are reduced to

$$\Psi_k^\pm(x) = e^{\pm ikx} - \frac{i\alpha}{2|k| + i\alpha} e^{\pm ika} e^{i|k|\,|x-a|}, \quad x, k \in \mathscr{R};$$
$$t(k) = \frac{2k}{2k + i\alpha}, \quad r^\pm(k) = \frac{-i\alpha}{2k+i\alpha} e^{\pm 2ika}, \quad k \geqslant 0. \tag{2.5}$$

Thus the unitary scattering matrix, which is defined by

$$S(k) := \begin{bmatrix} t^+(k) & r^-(k) \\ r^+(k) & t^-(k)\end{bmatrix}, \quad k \geqslant 0, \tag{2.6}$$

takes the simple form

$$S(k) = \frac{1}{2k+i\alpha}\begin{bmatrix} 2k & -i\alpha \\ -i\alpha & 2k\end{bmatrix} \underset{k\to 0}{\to} \begin{bmatrix} 0 & -1 \\ -1 & 0\end{bmatrix}. \tag{2.7}$$

III. SCALING LIMIT OF THE TRANSITION MATRIX FOR INTEGRABLE POTENTIALS

Here we use the following implicit definition of the transition operator:

$$(H_\epsilon - k^2)^{-1} = G_k - G_k T_k(\epsilon) G_k, \quad \operatorname{Im} k > 0. \tag{3.1}$$

The time-independent derivation of this stationary expression shall be presented in Appendix A.

The scaling limit $(\epsilon \to 0)$ of this Schrödinger operator has been studied in one and three dimensions.[32,33,42–44] Especially, the convergence of H_ϵ to H_α in the norm-resolvent sense was proved,[32] with the corresponding "coupling constants" α_i defined by Eq. (1.10).

We present the proof of convergence for the transition matrix in two steps: at first, we restrict the point interaction to one center; then we generalize this limit to the case of N centers.

A. One interaction center

For the Hamilton operator

$$H_\epsilon := H_0 \dot{+} [\lambda(\epsilon)/\epsilon^2] V((x-a)/\epsilon), \tag{3.2}$$

the transition operator takes the following "symmetrized" form[32,34,45,46]:

$$T_k(\epsilon) = \frac{\lambda(\epsilon)}{\epsilon^2} v_\epsilon \left(I + \frac{\lambda(\epsilon)}{\epsilon^2} u_\epsilon G_k v_\epsilon\right)^{-1} u_\epsilon, \quad \operatorname{Im} k > 0,$$
$$v_\epsilon := v((\_ - a)/\epsilon), \quad u_\epsilon := u((\_ - a)/\epsilon),$$
$$u := v \operatorname{sgn} v, \quad v := |V|^{1/2}. \tag{3.3}$$

The corresponding transmission and reflection coefficients are defined by its matrix elements

$$t_k^\pm(\epsilon) := 1 + (1/2ik)\langle e^{\pm ik\_} | T_k(\epsilon) e^{\pm ik\_}\rangle,$$
$$r_k^\pm(\epsilon) := (1/2ik)\langle e^{\mp ik\_} | T_k(\epsilon) e^{\pm ik\_}\rangle,$$
$$\operatorname{Re} k \geqslant 0, \quad \operatorname{Im} k \to 0^+. \tag{3.4}$$

Now the unitary scaling transformation U_ϵ defined by Eq. (1.8), which acts as

$$U_\epsilon V(\_)U_\epsilon^{-1} = V(\_/\epsilon), \quad U_\epsilon H_0 U_\epsilon^{-1} = \epsilon^2 H_0,$$
$$U_\epsilon G_k U_\epsilon^{-1} = \epsilon^{-2} G_{k/\epsilon}, \tag{3.5}$$

and an appropriate translation of the center to the origin yields the expressions

$$t_k^\pm(\epsilon) = 1 + [\lambda(\epsilon)/2i\epsilon k]$$
$$\times \langle e^{\pm i\epsilon k\_} v(\_)|(I + \lambda(\epsilon)u(\_)G_{\epsilon k}v(\_))^{-1}u(\_)e^{\pm i\epsilon k\_}\rangle,$$
$$r_k^\pm(\epsilon) = [\lambda(\epsilon)/2i\epsilon k]e^{\pm 2ika}$$
$$\times \langle e^{\mp i\epsilon k\_} v(\_)|(I + \lambda(\epsilon)u(\_)G_{\epsilon k}v(\_))^{-1}u(\_)e^{\pm i\epsilon k\_}\rangle; \tag{3.6}$$

the Hilbert–Schmidt operator[32]

$$B_k(\epsilon) := \lambda(\epsilon)u(\_)G_{\epsilon k}v(\_),$$
$$G_{\epsilon k} = (H_0 - \epsilon^2 k^2)^{-1}, \quad \operatorname{Im} k > 0, \tag{3.7}$$

acts as an integral operator with the kernel

$$\frac{i\lambda(\epsilon)}{2\epsilon k}u(x)e^{i\epsilon k|x-x'|}v(x') \underset{\epsilon\to 0}{\to} \frac{i\lambda'(0)}{2k}u(x)v(x'). \tag{3.8}$$

This pointwise convergence of the kernel, with the aid of dominated convergence, implies the convergence of $B_k(\epsilon)$ in the sense of the Hilbert–Schmidt norm; the second resolvent equation then leads to the limit in the sense of the operator norm

$$\lim_{\epsilon\to 0}\|(I + B_k(\epsilon))^{-1} - (I + (i\alpha/2k)P)^{-1}\| = 0,$$
$$P := |u\rangle\langle v|/\langle v|u\rangle, \tag{3.9}$$

where we assumed $\langle v|u\rangle = \int_{\mathscr{R}} dx\, V(x) \neq 0$. Thus we finally obtain the desired limits

$$\lim_{\epsilon\to 0} t_k^\pm(\epsilon) = t(k), \quad \lim_{\epsilon\to 0} r_k^\pm(\epsilon) = r^\pm(k), \quad k \geqslant 0. \tag{3.10}$$

The simple expressions of these scattering coefficients are quoted in Eq. (2.5).

In the special case $\langle v|u\rangle = \int_{\mathscr{R}} dx\, V(x) = 0$, the limit

$$\lim_{\epsilon\to 0}\|(I + B_k(\epsilon))^{-1} - I + (i/2k)\lambda'(0)|u\rangle\langle v|\,\| = 0 \tag{3.11}$$

leads to the result

$$\lim_{\epsilon\to 0} t_k^\pm(\epsilon) = 1, \quad \lim_{\epsilon\to 0} r_k^\pm(\epsilon) = 0, \quad k \geqslant 0, \tag{3.12}$$

which corresponds to the case of no point interaction, i.e., $\Gamma_{ij}^k = 0$, $k \geqslant 0$.

B. N interaction centers

Here we start with the stationary expression of the transition operator emerging from its implicit definition (3.1), i.e.,[32]

$$T_k(\epsilon) = \sum_{i=1}^{N}\sum_{j=1}^{N} T_k^{ij}(\epsilon),$$
$$T_k^{ij}(\epsilon) := [\lambda_j(\epsilon)/\epsilon^2]v_{i\epsilon}[A_k^{-1}(\epsilon)]_{ij}u_{j\epsilon},$$
$$[A_k(\epsilon)]_{ij} := \delta_{ij} + [\lambda_i(\epsilon)/\epsilon^2]u_{i\epsilon}G_k v_{j\epsilon}, \quad \operatorname{Im} k > 0,$$
$$v_{j\epsilon} := v_j((\_ - a_j)/\epsilon), \quad u_{j\epsilon} := u_j((\_ - a_j)/\epsilon),$$
$$u_j := v_j \operatorname{sgn} v_j, \quad v_j := |V_j|^{1/2}. \tag{3.13}$$

The kernel $h_{\epsilon,k}^{ij}$ of the Hilbert-Schmidt operator $u_{i\epsilon}G_k v_{j\epsilon}$ explicitly reads

$$h_{\epsilon,k}^{ij}(x,x') := u_i\left(\frac{x - a_i}{\epsilon}\right)\frac{i}{2k}e^{ik|x-x'|}v_j\left(\frac{x'-a_j}{\epsilon}\right),$$
$$\operatorname{Im} k > 0. \tag{3.14}$$

Now one can again translate each center to the origin and insert the scaling transformation U_ϵ; one obtains the following limits:

$$\lim_{\epsilon\to 0}\langle e^{\sigma ik\_}|T_k^{ij}(\epsilon)e^{\sigma' ik\_}\rangle$$
$$= \lambda_j'(0)e^{ik(\sigma' a_j - \sigma a_i)}\langle v_i|[A_{k,0}^{-1}]_{ij}u_j\rangle, \tag{3.15}$$
$$[A_{k,0}]_{ij} := \delta_{ij} + \lambda_i(0)g_k(a_i - a_j)|u_i\rangle\langle v_j|, \quad \sigma, \sigma' = \pm 1.$$

Straightforward evaluation, in the limit $\operatorname{Im} k \to 0^+$, yields

$$\lim_{\epsilon\to 0}\langle e^{\sigma ik\_}|T_k^{ij}(\epsilon)e^{\sigma' ik\_}\rangle = e^{-\sigma ika_i}\Gamma_{ij}^k e^{+\sigma' ika_j}, \quad k \geqslant 0. \tag{3.16}$$

Thus we finally obtain the following theorem.

Theorem: Let the scaled potentials V_i fulfill the integrability assumption quoted in the definition (1.9) of H_ϵ. In the scaling limit $\epsilon\to 0$, the scattering amplitude of the Schrödinger operator H_ϵ tends to the corresponding transition matrix of point interactions, described by the Hamilton operator H_α, in the sense that

$$\lim_{\epsilon\to 0} t_k^\pm(\epsilon) = t(k), \quad \lim_{\epsilon\to 0} r_k^\pm(\epsilon) = r^\pm(k), \quad k \geqslant 0, \tag{3.17}$$

if the "coupling constants" α_i are chosen according to definition (1.10).

The result of norm-resolvent convergence in the scaling limit has been generalized to the case of an infinitely countable, discrete set of interaction points.[32] This limiting procedure may be useful for the evaluation of periodic interactions with additional local impurities, as in crystals, for instance.[40]

APPENDIX A: TIME-INDEPENDENT DEFINITION OF THE TRANSITION AMPLITUDE FOR A WIDE CLASS OF POTENTIALS

In three dimensions, for integrable Rollnik potentials, the so-called eigenfunction expansion has been presented in detail in monographs,[34,47] using quadratic form techniques; also in one dimension, the time-independent derivation of the scattering amplitude and its connection with the scattering waves can be found in the literature.[36,48,49] Accordingly, for the eigenfunction expansion of the Schrödinger operator $H := -\Delta \dot{+} V$ on the real line, defined in the sense of quadratic forms, we assume

$$\int_{\mathscr{R}} dx|V(x)| < \infty. \tag{A1}$$

The scattering eigenfunctions $\Phi_k^\pm(x)$ obey the Lippmann–Schwinger equations

$$\Phi_k^{\pm}(x) = e^{\pm ikx} - \frac{i}{2|k|}\int_{\mathscr{R}} dy\, e^{i|k|\,|x-y|}V(y)\Phi_k^{\pm}(y),$$
$$k \in \mathscr{R} - \Sigma, \quad k \neq 0; \tag{A2}$$

here Σ is some closed set of Lebesgue-measure zero. (In fact, Σ is either empty or only contains the origin.[48,49] These eigenfunctions can then be used to express the corresponding wave operators $\Omega_\pm$ by their matrix elements[48]

$$\langle f|\Omega_\pm g\rangle = \int_{\mathscr{R}} dk\, \overline{\hat{f}_\pm(k)}\tilde{g}(k),$$
$$\tilde{g}(k) := (2\pi)^{-1/2}\int_{\mathscr{R}} dx\, e^{-ikx}g(x),$$
$$\hat{f}_\pm(k) := (2\pi)^{-1/2}\int_{\mathscr{R}} dx\, \overline{\Phi_k^{\pm}(x)}f(x)dx,$$
$$f,g \in \mathscr{H} := L^2(\mathscr{R}), \quad k \in \mathscr{R} - \{0\}. \tag{A3}$$

The matrix element of the scattering operator $S := \Omega_-^\dagger \Omega_+$ can then be evaluated as

$$\langle f|(S-I)g\rangle = \langle f|(\Omega_-^\dagger - \Omega_+^\dagger)\Omega_+ g\rangle = \langle(\Omega_- - \Omega_+)f|\Omega_+ g\rangle = \lim_{\epsilon\to 0}(-i)\int_{-\infty}^{+\infty} dt\, e^{-\epsilon|t|}$$
$$\times \langle |V|^{1/2}e^{-itH_0}f|\operatorname{sgn} V\cdot|V|^{1/2}e^{-itH}\Omega_+ g\rangle = \lim_{\epsilon\to 0}(-i)\int_{-\infty}^{+\infty} dt\, e^{-\epsilon|t|}\int_{\mathscr{R}} dk\int_{\mathscr{R}} dk'\, e^{it(k'^2-k^2)}$$
$$\times \overline{\tilde{f}(k')}\,\tilde{g}(k)\cdot(2\pi)^{-1}\int_{\mathscr{R}} dx\, \Phi_k^+(x)V(x)e^{-ik'x}, \tag{A4}$$

between states f, g such that $f, \Omega_+ g \in \operatorname{dom} H_0^{1/2}$; here the supports of $\tilde{f}$ and $\tilde{g}$ are assumed to be disjoint from $\{0\}$. Thus one obtains the time-dependent derivation of the scattering amplitude

$$t(k',k) := (2\pi)^{-1}\int_{\mathscr{R}} dx\, e^{-ik'x}V(x)\Phi_k^+(x),$$
$$k \in \mathscr{R} - \{0\},\ k' \in \mathscr{R}, \tag{A5}$$

on the energy shell, e.g.,

$$\langle f|(S-I)g\rangle = -2\pi i\int_{\mathscr{R}} dk\int_{\mathscr{R}} dk'$$
$$\times \overline{\tilde{f}(k')}\tilde{g}(k)\delta(k^2 - k'^2)t(k',k), \tag{A6}$$

with the supports of $\tilde{f}$ and $\tilde{g}$ being disjoint from $\{0\}$, by assumption.

Then the implicit definition

$$(H - k^2)^{-1} = G_k - G_k T_k G_k, \quad \operatorname{Im} k > 0, \tag{A7}$$

of the transition operator T_k, which implies its stationary expression[32,44,46,47]

$$T_k = v(I + uG_k v)^{-1}u,$$
$$\operatorname{Im} k > 0, \quad u := v\operatorname{sgn} v, \quad v := |V|^{1/2}, \tag{A8}$$

together with the Lippmann–Schwinger equations (A2), in the limit $\operatorname{Im} k\to 0^+$, leads to the on-shell transition matrix

$$\langle e^{\sigma ik\cdot}|T_k e^{\sigma' ik\cdot}\rangle = 2\pi t(\sigma k, \sigma' k), \quad \sigma, \sigma' = \pm 1, \quad k > 0. \tag{A9}$$

APPENDIX B: TIME-INDEPENDENT DERIVATION OF THE TRANSMISSION AND REFLECTION COEFFICIENTS FOR N ZERO-RANGE INTERACTION CENTERS

In order to derive the transmission and reflection coefficients (23.3) from the wave operators (which exist and are complete),[29] we again use the scattering eigenfunctions (2.1) as kernels of the integral transformations

$$\hat{f}_\pm(k) := (2\pi)^{-1/2}\int_{\mathscr{R}} dx\, \overline{\Psi_k^{\pm}(x)}f(x)dx,$$
$$k \in \mathscr{R}, \quad f \in \mathscr{H}. \tag{B1}$$

Again we start from the wave operators $\Omega_\pm$; between states f, $g \in \operatorname{dom} H_0^{1/2}$ we calculate

$$\langle f|\Omega_\pm g\rangle = \langle f|g\rangle - i\int_{-\infty}^{0} dt\, e^{\epsilon t}$$
$$\times \sum_{i=1}^{N} \alpha_i\, \overline{(e^{-itH_\alpha}f)(a_i)}\cdot(e^{-itH_0}g)(a_i)$$
$$= \langle f|g\rangle - i\int_{-\infty}^{0} dt\, e^{\epsilon t}(2\pi)^{-1}\int_{-\infty}^{+\infty} dk$$
$$\times e^{-itk^2}\tilde{g}(k)e^{ika_i}\int_0^\infty dk'\, e^{itk'^2}$$
$$\times\ \overline{(\Psi_{k'}^{+}(a_i)}\ \overline{\hat{f}_+(k')}$$
$$+\ \overline{\Psi_{k'}^{-}(a_i)}\ \overline{\hat{f}_-(k')}). \tag{B2}$$

Then the eigenvalue equations

$$((H_\alpha - k^2)^{-1}\hat{f})_\pm(k') = (k'^2 - k^2)^{-1}\hat{f}_\pm(k'),$$
$$k' \in \mathscr{R}, \quad \operatorname{Im} k > 0, \tag{B3}$$

together with the eigenfunction expansion

$$f(x) = (2\pi)^{-1/2}\int_0^\infty dk\,(\hat{f}_+(k)\Psi_k^+(x) + \hat{f}_-(k)\Psi_k^-(x))$$
$$= (2\pi)^{-1/2}\int_{\mathscr{R}} dk\,\hat{f}_\pm(k)\Psi_k^{\pm}(x), \tag{B4}$$

which are valid for all $f \in \mathscr{H}$, lead to the result

$$\langle f|\Omega_\pm g\rangle = \int_{\mathscr{R}} dk\, \overline{\hat{f}_\pm(k)}\tilde{g}(k), \quad f,g \in \mathscr{H}. \tag{B5}$$

The matrix element of the scattering operator can then be calculated using the integral kernels of the wave operators $\Omega_\pm$, i.e.,

$$(\Omega_\pm^\dagger f)^\sim(k) = (2\pi)^{-1/2}\int_{\mathscr{R}} dx\, \overline{\Psi_k^{\pm}(x)}f(x), \quad k \in \mathscr{R};$$
$$(\Omega_\pm f)(x) = (2\pi)^{-1/2}\int_{\mathscr{R}} dk\,\tilde{f}(k)\Psi_k^{\pm}(x), \quad x \in \mathscr{R}. \tag{B6}$$

One obtains the integral representation

$$\langle f|(S-I)g\rangle = \lim_{\epsilon\to 0}(-i)\int_{-\infty}^{+\infty} dt\, e^{-\epsilon|t|}\sum_{i=1}^{N}\alpha_i\,\overline{(e^{-itH_0}f)(a_i)}(e^{-itH}\Omega_+g)(a_i)$$

$$= -2\pi i\int_{\mathscr{R}} dk\int_{\mathscr{R}} dk'\ \overline{\tilde{f}(k')}\tilde{g}(k)\delta(k^2-k'^2)t(k',k),\quad f,\Omega_+g\in \operatorname{dom} H_0^{1/2}, \tag{B7}$$

with the kernel

$$2\pi t(k',k) = \sum_{i=1}^{N}\alpha_i e^{-ik'a_i}\Psi_k^+(a_i),\quad k,k'\in\mathscr{R}. \tag{B8}$$

Comparing this result with the transmission and reflection coefficients (2.3) one ends up with the desired result

$$t^+(k) = t^-(k) = 1 + (2\pi/2ik)t(\pm k,\ \pm k),\quad r^\pm(k) = (2\pi/2ik)t(\mp k,\ \pm k),\quad k\geqslant 0. \tag{B9}$$

ACKNOWLEDGMENT

The authors are indebted to F. Gesztesy for many stimulating discussions.

[1]S. Albeverio, Helv. Phys. Acta **40**, 135 (1967).
[2]S. Albeverio, R. Høegh-Krohn, W. Kirsch, and F. Martinelli, Adv. Appl. Math. **3**, 435 (1982).
[3]F. Gesztesy and W. Kirsch, "One-dimensional Schrödinger operators with interactions singular on a discrete set," preprint ZIF, Bielefeld, 1983.
[4]A. Grossmann, R. Høegh-Krohn, and M. Mebkhout, J. Math. Phys. **21**, 2376 (1980).
[5]E. M. Haacke and L. L. Foldy, Phys. Rev. C **23**, 1320, 1330 (1981).
[6]R. Høegh-Krohn, H. Holden, and F. Martinelli, Lett. Math. Phys. **7**, 221 (1983).
[7]E. H. Lieb and M. de Llano, J. Math. Phys. **19**, 860 (1978).
[8]C. K. Majomdar, J. Math. Phys. **13**, 705 (1972).
[9]J. B. McGuire, J. Math. Phys. **6**, 432 (1965); **7**, 123 (1966).
[10]J. B. McGuire and C. A. Hurst, J. Math. Phys. **13**, 1595 (1972).
[11]D. Testard, "An explicitly solvable model of periodic point interation in one dimension," preprint ZIF, Bielefeld, 1983.
[12]L. E. Thomas, in *Lecture Notes in Physics 130* (Springer, Berlin, 1980), p. 159.
[13]C. N. Yang, Phys. Rev. **168**, 1920 (1968).
[14]L. R. Dodd, J. Math. Phys. **11**, 207 (1969).
[15]G. Flamand, *Cargese Lectures in Theoretical Physics* (Gordon and Breach, New York, 1967), p. 247.
[16]M. G. Fuda, Phys. Rev. C **7**, 1365 (1973).
[17]A. K. Jain, S. K. Deb, and C. S. Shastry, Am. J. Phys. **46**, 147 (1978).
[18]V. I. Karloukovski, Rep. Math. Phys. **10**, 87 (1976).
[19]J. F. Reading and J. L. Sigel, Phys. Rev. B **5**, 556 (1972).
[20]S. Albeverio, R. Høegh-Krohn, and L. Streit, J. Math. Phys. **18**, 907 (1977).
[21]F. A. Berezin and L. D. Faddeev, Sov. Math. Dokl. Ak. Nauk. **2**, 372 (1961).
[22]J. D. Dollard, J. Math. Phys. **19**, 806 (1978).
[23]M. D. Faddeev and B. S. Pavlov, Teor. Mat. Fiz. **55**, 257 (1983).
[24]R. A. Minlos and L. D. Faddeev, Sov. Phys. JETP **14**, 1315 (1961).
[25]R. A. Minlos and L. D. Faddeev, Sov. Phys. Dokl. **6**, 1072 (1962).
[26]L. E. Thomas, J. Math. Phys. **20**, 1848 (1979).
[27]J. Zorbas, J. Math. Phys. **21**, 840 (1980).
[28]M. Combescure and J. Ginibre, J. Funct. Anal. **29**, 54 (1978).
[29]F. Gesztesy and L. Pittner, Rep. Math. Phys. **19**, 143 (1984).
[30]M. A. Naimark, *Linear Differential Operators. Part II* (Frederick Ungar, New York, 1968).
[31]R. A. Adams, *Sobolev Spaces* (Academic, New York, 1975).
[32]S. Albeverio, F. Gesztesy, R. Høegh-Krohn, and W. Kirsch, "On point interactions in one dimension," to appear in J. Op. Theory.
[33]S. Albeverio and R. Høegh-Krohn, J. Op. Theory **6**, 313 (1981).
[34]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U.P., Princeton, NJ, 1971).
[35]L. D. Faddeev and B. Seckler, J. Math. Phys. **4**, 72 (1963).
[36]K. Chadan and P. C. Sabatier, *Inverse Problems in Quantum Scattering Theory* (Springer, New York, 1977).
[37]Ö. Dammert, J. Math. Phys. **24**, 2163 (1983).
[38]E. H. Lieb and D. C. Mattis, *Mathematical Physics in One Dimension* (Academic, New York, 1966).
[39]R. G. Newton, J. Math. Phys. **21**, 493 (1980).
[40]R. G. Newton, J. Math. Phys. **24**, 2152 (1983).
[41]A. Peres, J. Math. Phys. **24**, 1110 (1983).
[42]S. Albeverio, F. Gesztesy, R. Høegh-Krohn, and L. Streit, Ann. Inst. H. Poincaré A **37**, 1 (1982); **38**, 263 (1983).
[43]H. Holden, R. Høegh-Krohn, and S. Johannesen, Adv. Appl. Math. **4**, 402 (1983); "The short-range expansion in solid state," to appear in Ann. Inst. H. Poincaré A.
[44]H. Holden, R. Høegh-Krohn, and M. Mebkhout, "The short-range expansion for multiple well scattering theory," preprint ZIF, Bielefeld, 1984.
[45]W. Thirring, *Quantum Mechanics of Atoms and Molecules* (Springer, New York, 1981).
[46]D. Bollé, F. Gesztesy, and S. F. J. Wilk, "A complete treatment of low-energy scattering in one dimension," to appear in J. Op. Theory.
[47]M. Reed and B. Simon, *Methods of Modern Mathematical Physics. III. Scattering Theory* (Academic, New York, 1979).
[48]M. Schechter, *Operator Methods in Quantum Mechanics* (North-Holland, New York, 1981).
[49]P. Deift, E. Trubowitz, Commun. Pure Appl. Math. **32**, 121 (1979).

Quantization of systems with many degrees of freedom by the method of collective coordinates: Quantum mechanics around a classical periodic orbit

H. J. W. Müller-Kirsten and A. Wiedemann
Department of Physics, University of Kaiserslautern, 6750 Kaiserlautern, West Germany

(Received 27 September 1984; accepted for publication 22 February 1985)

A fairly complete perturbation theoretical solution of the Schrödinger equation around a classical periodic orbit is given as an example demonstrating the quantization of a system with many degrees of freedom by the method of collective coordinates. In particular, explicit expressions are derived for the creation and annihilation operators of quantum fluctuations orthogonal to the classical path, as well as the corresponding eigenfunctions, in both the adiabatic and nonadiabatic domains. It is also shown that corresponding expressions become proportional (and thus can be linked) in domains where they overlap. Finally the splitting of asymptotically degenerate states, and hence the band structure of the resulting spectrum, is calculated.

I. INTRODUCTION

The quantization of nonlinear field theories possessing stable finite energy (or action) classical solutions such as solitons, instantons, and monopoles has been the subject of numerous investigations in recent years. There are essentially two approaches: the path-integral method and the Hamiltonian method. The latter, which will be discussed here, is basically that of the Schrödinger equation with the Hamiltonian expressed in terms of collective variables that parametrize the classical solution (or are symmetry parameters as in extended theories) and field variables which describe the fluctuations around it. In general this transformation leads to a larger number of variables so that appropriate constraints have to be imposed in order to remove superfluous degrees of freedom. In field theory the classical solution does not possess some particular symmetry of the Hamiltonian, and consequently one or more zero energy eigenstates arise which lead to an ill-defined Green's function. The essential point of the procedure is therefore the subdivision of the space spanned by the states of the unperturbed Hamiltonian into two mutually orthogonal subsets such that the perturbation expansion involves only states of the subset of transverse states, and the zero modes are contained in the complementary set. Choosing the constraints in this way the Green's functions are well defined and the fluctuations around the classical solution can be quantized in accordance with the familiar canonical procedure. It is evident that although the procedure is clear, its application to specific models can be complicated. In the following we consider therefore in some detail the application of the procedure to quantum mechanics around a classical periodic orbit, since this permits explicit calculation and is of fundamental interest. The problem as such was first investigated by Gervais and Sakita,[1] who also advocated an appropriate WKB wave function approach for application to systems with many degrees of freedom. Although we shall follow in part the idea of these authors, our objective is to treat the method as a straightforward generalization of well-known one-dimensional perturbation theory. Our ultimate aim here is the calculation of the band structure of the resulting spectrum,[2] so that the eigenfunctions of the Hamiltonian have to be found in both the adiabatic and nonadiabatic domains, as well as their proportionality in the region where they overlap. The procedure we adopt for this purpose is similar to that used for the calculation of the level splitting of the double-well potential,[3,4] and was originally applied to the Schrödinger equation for the periodic potential[5] and other periodic equations.[6] The considerations presented below are distinctly different, however, and can serve as a prototype for the application of the method to more complicated model theories.

In order to have a basis for our subsequent considerations, Sec. II is devoted to a recapitulation of the main steps involved in transforming the Schrödinger equation to collective and fluctuation variables.[7] Section III deals with the validity of the perturbation expansion and Gribov ambiguities.[8] In Secs. IV and V we construct the fluctuation creation and annihilation operators in both the adiabatic and nonadiabatic domains (in the latter case we follow in part the suggestions of Ref. 1 and the derivation of Ref. 9). In Sec. VI we demonstrate that both types of quantization operators merge into one another in the domain where they overlap. In Sec. VII we calculate the WKB wave functional of the ground state and with the help of the latter the Bohr–Sommerfeld quantization condition for the multidimensional case under consideration. In Secs. VIII and IX we derive the unperturbed adiabatic and nonadiabatic wave functionals in the domain where they overlap; in Sec. X we demonstrate their proportionality in the overlap region. Finally in Sec. XI we calculate the level splitting and hence the band structure of the resulting spectrum.

II. THE SCHRÖDINGER EQUATION IN TERMS OF COLLECTIVE AND FLUCTUATION COORDINATES

In order to have a basis for our subsequent considerations we recapitulate first from Refs. 1 and 7 the main steps involved in reexpressing the Schrödinger equation in terms of collective and fluctuation coordinates.

We consider the Lagrangian

$$L(\mathbf{R}) = \tfrac{1}{2}\dot{\mathbf{R}}^2 - V(\mathbf{R}), \tag{1}$$

where $\mathbf{R}$ is the N-dimensional Euclidean vector with conjugate momentum

$$\mathbf{P} = \frac{\partial L}{\partial \dot{\mathbf{R}}} = \dot{\mathbf{R}}. \tag{2}$$

 0022-2488/85/071680-15$02.50

We assume that $\mathbf{R}$ is given approximately by the classical path $\mathbf{r}(f(q))$ defined by Newton's equation

$$\mathbf{r}_{ff} = -\left(\frac{\partial V(\mathbf{R})}{\partial \mathbf{R}}\right)_{\mathbf{r}}. \tag{3}$$

Here $f(q)$ is a function which fixes the parametrization of the curve $\mathbf{r}$, and the parameter q is called the collective variable. Of course, it is only in the context of classical mechanics that f plays the role of time. Introducing fluctuation variables η_a which describe the deviations away from $\mathbf{r}$ we can write

$$\mathbf{R} = \mathbf{r}(f(q)) + \sum_{a=2} \mathbf{n}_a(f(q))\eta_a, \tag{4}$$

where $\{\mathbf{n}_a\}$ is a set of orthonormal unit vectors which are orthogonal to the classical path at q, i.e.,

$$\mathbf{n}_a \cdot \mathbf{n}_b = \delta_{ab}, \quad \mathbf{n}_a \cdot \mathbf{r}_f = 0, \tag{5}$$

where $a,b,... = 2,...,N$ and $\mathbf{r}_f \equiv d\mathbf{r}(f)/df$.

Setting

$$\mathbf{n}_1 = \mathbf{r}_f/(\mathbf{r}_f^2)^{1/2}, \tag{6}$$

we see that $\mathbf{n}_1, \mathbf{n}_a$, $a = 2,...,N$ form a moving local reference frame. Then

$$R_i = \sum_{j=1}^{N} (\mathbf{n}_j)_i Q_j, \tag{7}$$

where

$$Q_1 = \mathbf{r}\cdot\mathbf{r}_f/(\mathbf{r}_f^2)^{1/2}, \quad Q_a = \eta_a + \mathbf{r}\cdot\mathbf{n}_a, \tag{8}$$

and R_i, $i = 1,...,N$ are the components of $\mathbf{R}$ with respect to some fixed frame of reference. Replacing $\mathbf{R}$ of $L(\mathbf{R})$ by (7) we see that $L(q,Q_j)$ is a function of $N+1$ variables with conjugate momenta

$$p = \frac{\partial L}{\partial \dot{q}} = \frac{\partial L}{\partial \dot{R}_i}\frac{\partial \dot{R}_i}{\partial \dot{q}},$$

$$p_j = \frac{\partial L}{\partial \dot{Q}_j} = \frac{\partial L}{\partial \dot{R}_i}\frac{\partial \dot{R}_i}{\partial \dot{Q}_j} \tag{9}$$

(summations understood). Now

$$\frac{\partial \dot{R}_i}{\partial \dot{Q}_j} = \frac{\partial R_i}{\partial Q_j} = (\mathbf{n}_j)_i \equiv M_{ij}. \tag{10}$$

Defining W as the inverse of M we have

$$W_{ij} = (\mathbf{n}_i)_j, \quad M_{ij}W_{jk} = \delta_{ik}. \tag{11}$$

The latter of these is, in fact, the completeness relation

$$\frac{(\mathbf{r}_f)_i(\mathbf{r}_f)_k}{\mathbf{r}_f^2} + \sum_{a=2}^{N} (\mathbf{n}_a)_i\,(\mathbf{n}_a)_k = \delta_{ik}. \tag{12}$$

Eliminating $\partial L/\partial \dot{R}_i$ from (9) we obtain the Dirac ϕ or primary constraint

$$\phi \equiv p - p_j W_{ji} T_i = 0, \tag{13}$$

where

$$T_i(Q,q) = \frac{\partial \dot{R}_i}{\partial \dot{q}} = \frac{\partial}{\partial q} R_i. \tag{14}$$

The constraint (13) allows the removal of the superfluous variable of $L(q,Q_j)$.

Our next step is to express P_i in terms of p, p_a. We have

$$P_i = \frac{\partial L}{\partial \dot{R}_i} = p_j W_{ji} = p_1 W_{1i} + p_a W_{ai}. \tag{15}$$

Inserting p_1 from (13) this becomes

$$P_i = W_{1i}(p - p_a W_{aj} T_j)/W_{1k} T_k + p_a W_{ai}.$$

Inserting (14) for T_j we obtain

$$P_i = \frac{(\mathbf{r}_f)_i}{(\mathbf{r}_f^2 - \mathbf{r}_{ff}\cdot\boldsymbol{\eta})}\left(\frac{1}{f'}p - \sum_{a,b=2}^{N} \Gamma^a_b \eta^b p_a\right) + \sum_{a=2}^{N} (\mathbf{n}_a)_i\, p_a, \tag{16}$$

where

$$\Gamma^a_b = \mathbf{n}_a \cdot \mathbf{n}_{bf}, \quad \mathbf{n}_{bf} = \frac{d\mathbf{n}_b(f)}{df}. \tag{17}$$

The Hamiltonian is given by the Legendre transform

$$H(R_i,P_i) = \mathbf{P}\cdot\dot{\mathbf{R}} - L(\mathbf{R}) = \tfrac{1}{2}\mathbf{P}^2 + V(\mathbf{R}). \tag{18}$$

In terms of the new coordinates

$$L(R_i(q,Q)) \equiv L(q,Q_i)$$

and

$$H'(q,Q_i,p,p_i) = p\dot{q} + p_i\dot{Q}_i - L(q,Q_i). \tag{19}$$

It can be verified (with $p_j = \dot{R}_i M_{ij}$, $p_j^2 = \dot{R}_i^2$) that

$$H' = \tfrac{1}{2}p_j^2 + V(R_i(q,Q)) + \dot{q}(p - p_j W_{jk} T_k) = H,$$

where $\dot{q}$ plays the role of a Lagrange multiplier.

In (16) we placed the momenta on the far right. Then with

$$[p,q] = -i, \quad [p_a,\eta_b] = -i\delta_{ab},$$

we have

$$[P_i,R_j] = \frac{-i}{(\mathbf{r}_f^2 - \mathbf{r}_{ff}\cdot\boldsymbol{\eta})}\Big\{(\mathbf{r}_f)_i\,(\mathbf{r}_f)_j + (\mathbf{r}_f)_i \sum_a (\mathbf{n}_{af})_j \eta_a - (\mathbf{r}_f)_i \sum_{a,b} \Gamma^a_b\, \eta_b (\mathbf{n}_a)_j\Big\} - i\sum_{a=2}^{N} (\mathbf{n}_a)_i (\mathbf{n}_a)_j.$$

Using $\mathbf{r}_f\cdot\mathbf{n}_a = 0$, $\mathbf{r}_{ff}\cdot\mathbf{n}_a = -\mathbf{r}_f\cdot\mathbf{n}_{af}$, and multiplying from both sides by functions of the form

$$\Psi = \sum_{i,j} (\mathbf{n}_i)_j \phi^i_j(\eta), \tag{20}$$

we find

$$[P_i,R_j] = -i\delta_{ij}.$$

Thus it is correct to place the momenta on the far right in P_i, and the transformation from R_i to q and η_a is canonical. Of course, this also implies that $\mathbf{P}$ is a self-adjoint operator in the space of functions of the form (20). It is important to realize that (20) represents an expansion of Ψ in terms of solutions $v_i(f)$ of the equation of small fluctuations with respect to the fixed frame of reference, i.e.,

$$\ddot{v}_i = -V^{(2)}_{ij}(\mathbf{r}_f)\, v_j.$$

Thus

$$\Psi = \sum_i v_i(f)\phi_i(\eta) = \sum_{i,a} (\mathbf{n}_a)_i \eta_a \phi_i(\eta),$$

as explained in the Appendix.

Hence

$$H = \frac{\mathbf{r}_f^2}{2(\mathbf{r}_f^2 - \mathbf{r}_{ff}\cdot\boldsymbol{\eta})^2}\left(\frac{1}{f'}p - \sum_{a,b=2}^{N} \Gamma^a_b \eta_b p_a\right)^2 + \frac{1}{2}\sum_{a=2}^{N} p_a^2 + V(\mathbf{r}+\boldsymbol{\eta}). \tag{21}$$

Using (16) and

$$dR_i = \left\{(\mathbf{r}_f)_i + \sum_{a=2}^{N} (\mathbf{n}_{af})_i \eta_a\right\} df + \sum_{a=2}^{N} (\mathbf{n}_a)_i \, d\eta_a$$

as well as $\mathbf{r}_f\cdot\mathbf{n}_a = 0$, etc. and derivatives, it can be verified that

$$\sum_{i=1}^{N} P_i \, dR_i \equiv p\, dq + \sum_{a=2}^{N} p_a \, d\eta_a.$$

The generating function F of the canonical transformation is therefore given by

$$\begin{aligned} dF &= H'(q,\eta_a; p, p_a)\, dt - p\, dq - \sum_{a=2}^{N} p_a \, d\eta_a \\ &\quad - H(R_i; P_i)\, dt + \sum_{i=1}^{N} P_i \, dR_i \\ &= [H'(q,\eta_a; p, p_a) - H(R_i; P_i)]\, dt. \end{aligned}$$

As noted the ansatz (20) for Ψ amounts to expanding Ψ in terms of small fluctuations $\Sigma_a (\mathbf{n}_a)_i \eta^a = v_i$ (see the Appendix) with respect to the fixed frame of reference. In Sec. V we shall be interested in a similar expansion in terms of small fluctuations with respect to the instantaneous reference frame.

Next we introduce a parameter g by setting

$$\boldsymbol{\sigma} = g\mathbf{r}, \quad \mathbf{r} \equiv \mathbf{r}(f(q)), \tag{22}$$

so that

$$\mathbf{r} \sim o(1/g), \quad \boldsymbol{\eta} \sim o(g^0)$$

and

$$V(\mathbf{r}) \equiv (1/g^2)\, \widetilde{V}(g\mathbf{r}) \sim o(1/g^2). \tag{23}$$

It will be seen that g plays the role of a semiclassical expansion parameter. Now

$$V(\mathbf{r}+\boldsymbol{\eta}) = V(\mathbf{r}) + \left(\frac{\partial V(\mathbf{R})}{\partial R_i}\right)_{\mathbf{r}} (\boldsymbol{\eta})_i + \frac{1}{2}\frac{\partial^2 V(\mathbf{R})}{\partial R_i \, \partial R_j}(\boldsymbol{\eta})_i(\boldsymbol{\eta})_j + \cdots, \tag{24}$$

where

$$\begin{aligned} \sum_{i=1}^{N} \frac{\partial V}{\partial R_i}(\boldsymbol{\eta})_i &= \sum_{i=1}^{N} \left\{\frac{(\mathbf{r}_f)_i}{(\mathbf{r}_f^2 - \mathbf{r}_{ff}\cdot\boldsymbol{\eta})}\left(\frac{1}{f'}\frac{\partial V(\mathbf{r})}{\partial q} - \eta_b \Gamma^a_b \frac{\partial V(\mathbf{r})}{\partial \eta_a}\right) + (\mathbf{n}_a)_i \frac{\partial V(\mathbf{r})}{\partial \eta_a}\right\} (\mathbf{n}_c)_i \eta_c \\ &= \frac{\partial V(\mathbf{r})}{\partial \eta_a}\eta_a \sim o\left(\frac{1}{g}\right). \end{aligned} \tag{25}$$

Thus the terms of (24) are of order g^i with $i = -2, -1, 0, \ldots$, respectively. We can now expand H in rising powers of g. Then

$$\begin{aligned} H &= \frac{p^2}{2f'^2\mathbf{r}_f^2} + V(\mathbf{r}) + \frac{p^2}{f'^2(\mathbf{r}_f^2)^2}\mathbf{r}_{ff}\cdot\boldsymbol{\eta} \\ &\quad + \frac{\partial V(\mathbf{r})}{\partial \eta_a}\eta_a + \frac{1}{2}p_a^2 + \frac{1}{2}\frac{\partial^2 V(\mathbf{r})}{\partial \eta_a \, \partial \eta_b}\eta_a \eta_b \\ &\quad - \frac{p\Gamma^b_a \eta_a p_b}{f'\mathbf{r}_f^2} + \frac{3}{2}\frac{p^2}{f'^2}\frac{(\mathbf{r}_{ff}\cdot\boldsymbol{\eta})^2}{(\mathbf{r}_f^2)^3} + o(g^5). \end{aligned} \tag{26}$$

Assuming that p is of order g^0 these terms are of order g^i with $i = 2, -2, 3, -1, 0, 0, 2$, and 4, respectively.

We now set

$$H\Psi = E\Psi$$

and

$$E = E_0 + E_1,$$

where $E_0 = O(1/g^2)$. E_0 can be called the classical energy, i.e., the total energy associated with the system described by the classical solution $\mathbf{r}(f)$; and E_1 represents the quantum correction of the classical energy.

III. UNIQUENESS OF THE EXPANSION AND GRIBOV AMBIGUITIES

It should be observed that the transformation (4) to the collective coordinate q and fluctuation variables η_a is "*a priori*" not unique; i.e., without the constraints (5) and for a given vector $\mathbf{R}$ any point on the classical path would be a possible origin of the moving frame of reference. This nonuniqueness is removed by the specification of the constraints (5) which for any given value of $\mathbf{R}$ admit only one particular point along the classical path (which then serves as the origin of the moving frame of reference) provided the fluctuations η_a are not too large (and thus avoid Gribov ambiguities[8]). This problem is similar to that in electrodynamics where the nonuniqueness of gauge-equivalent potentials $A_\mu(x)$ is removed by the imposition of a gauge constraint. Thus in the present context (5) plays the role of a gauge constraint.

Now $\mathbf{r}_f(f)$, the tangent to the classical path, is also a so-called zero mode, i.e., a solution of the linearized form of the classical equation of motion with eigenvalue zero. This can be seen by differentiating (3) with respect to f:

$$\ddot{\mathbf{r}}_f = -\frac{d}{df}\nabla V(\mathbf{r}) = -\left(\frac{\partial^2 V(\mathbf{R})}{\partial \mathbf{R}\, \partial \mathbf{R}}\right)_{\mathbf{r}} \mathbf{r}_f.$$

The constraints (5) therefore demand that the fluctuations η_a be normal to this zero mode (constraints of this type are called unitary gauge conditions), and this condition then ensures that the Green's function which is required for the complete solution of the Schrödinger equation is well defined (i.e., does not lead to infrared divergences).

We remarked above that the fluctuations η_a have to be sufficiently small in order that the transformation of $\mathbf{R}$ to the new variables be meaningful. If this transformation is unique then in principle

$$\mathbf{r}_f(f)\cdot\boldsymbol{\eta} = 0 \tag{27}$$

can be solved for f. This value of $f(q)$ then determines the value of q at which the moving frame of reference is located. It is conceivable, however, that in general equation (27) admits a number of different solutions, and then the perturbation procedure becomes ambiguous. The duplicates of the solutions q of (27) which arise in this way correspond to the Gribov ambiguities[8] discussed in the context of non-Abelian gauge theories. In the latter these duplicates result from a multitude of intersections of the orbit of the gauge field A_μ (formed by the application of any element g of the gauge group G to an element of the functional space) and the hyperplane defined by the gauge fixing condition. Figure 1 illustrates the possibility of two such intersections (at q,q') in the present case, and it is evident that the consistency of the perturbation procedure around the classical path requires the additional restriction that the fluctuations be small.

IV. QUANTIZATION: THE ADIABATIC APPROXIMATION

In the adiabatic or oscillatorlike approximation we set

$$H = V(\mathbf{r}) + H_f + o(g^2)$$

and

$$H_f = \frac{\partial V(\mathbf{r})}{\partial \eta_a}\eta_a + \frac{1}{2}p_a^2 + \frac{1}{2}\eta_a M_{ab}\eta_b, \tag{28}$$

where

$$M_{ab} \equiv \frac{\partial^2 V(\mathbf{r})}{\partial \eta_a\, \partial \eta_b}. \tag{29}$$

The word adiabatic implies the slow variation of an external parameter of the system around an equilibrium position, i.e., a local minimum of V at which $\partial V(\mathbf{r})/\partial \eta_a = 0$ and M is positive definite, i.e.,

$$My_c = \tfrac{1}{4}h_c^4 y_c, \quad h_c^4 > 0, \quad c = 2,\ldots,N. \tag{30}$$

Any y_c corresponding to $h_c^4 = 0$ is called a zero mode. Here we assume that each $h_c^4 \neq 0$. If any $h_c^4 = 0$ the corresponding momentum p_c is conserved; in this case the appropriate variable η_c has to be separated out before the WKB procedure of the nonadiabatic case is applied.

Setting

$$\eta_a = A_{ab}^T y_b, \quad A_{ab}^T V_{bc}^{(2)} A_{cd} = \tfrac{1}{4}h_a^4 \delta_{ad}, \quad AA^T = 1 \tag{31}$$

and

$$z_c = h_c y_c,$$

we can define operators

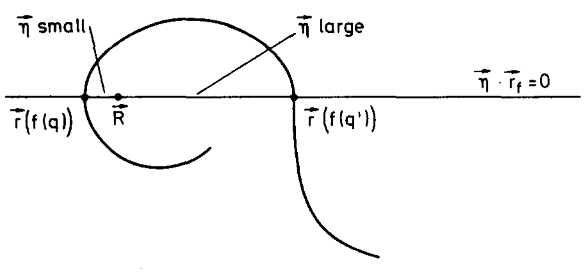


FIG. 1. Intersection of the classical path with the hyperplane $\boldsymbol{\eta}\cdot\mathbf{r}_f = 0$ at two different values of q, one corresponding to a small fluctuation $\boldsymbol{\eta}$, and the other to a large fluctuation $\boldsymbol{\eta}$.

$$A_c = -i\left(\frac{d}{dz_c} + \frac{1}{2}z_c\right),$$
$$A_c^+ = -i\left(\frac{d}{dz_c} - \frac{1}{2}z_c\right) \tag{32}$$

satisfying

$$[A_a, A_b] = 0, \quad [A_a^+, A_b^+] = 0, \quad [A_a, A_b^+] = \delta_{ab}, \tag{33}$$

so that we obtain the regular, diagonal representation of H_f, i.e.,

$$H_f = \frac{1}{2}\sum_{c=2}^{N} h_c^2\left(A_c^+ A_c + \frac{1}{2}\right). \tag{34}$$

To the same degree of approximation the eigenvalue E of H is

$$E = V(\mathbf{r}) + \frac{1}{2}\sum_{c=2}^{N} h_c^2\left(n_c + \frac{1}{2}\right), \tag{35}$$

where $n_c = 0,1,2,\ldots$. Thus there are $N-1$ zero-point energy contributions. This number is independent of the domain of validity of the adiabatic approximation of the wave function and results from treating one degree of freedom classically. This point will become clearer later when the quantization of the collective variable is considered in detail. Of course, diagonalization of H_f is also possible near (i.e., not only around) a minimum, i.e., for $\partial V(\mathbf{r})/\partial \eta_a \neq 0$, but then the ordering in rising powers of g is destroyed.

The semiclassical ground state $|S(\mathbf{r})\rangle$ is defined by

$$A_c|S(\mathbf{r})\rangle = 0, \quad c = 2,\ldots,N,$$

i.e., apart from a normalization factor, the wave function is

$$S(z) = \prod_c e^{-(1/4)z_c^2}.$$

It follows that

$$\langle S(\mathbf{r})|\mathbf{R}|S(\mathbf{r})\rangle = \langle S(\mathbf{r})|\mathbf{r} + \boldsymbol{\eta}|S(\mathbf{r})\rangle = \mathbf{r}(f), \tag{36}$$

after reexpressing each component η_a in terms of A_a and A_a^+.

Before higher-order terms of H can be treated perturbatively, we have to deal with the first term of (26) and the corresponding quantization of the collective coordinate q. We defer this point to Sec. VII where the wave functionals are derived explicitly. Thus, once this has been done, we can reexpress each factor η_a, p_a, q, or p in $H_i \equiv H - V(\mathbf{r}) - H_f$ in terms of corresponding quantization creation and annihilation operators, and the perturbation theory can, in principle, be carried out to any desired order in the adiabatic domain.

V. QUANTIZATION: THE NONADIABATIC APPROXIMATION

Here we begin by recapitulating steps from Refs. 1 and 9. As before we have $H\Psi = E\Psi$ with $E = E_0 + E_1$, $E_0 = o(1/g^2)$ but this time one substitutes

$$\Psi(q,\boldsymbol{\eta}) = \frac{e^{i\epsilon S_0(q)}\chi(q,\boldsymbol{\eta})}{(dS_0/dq)^{1/2}}, \tag{37}$$

where $\epsilon^2 = \pm 1$ depending on whether $E \gtrless V$, i.e., in the classically allowed or forbidden region. Assuming that $S_0 = o(1/g^2)$ and equating in $H\Psi = E\Psi$ [with H replaced by (26)] terms of the same order in g, we obtain

$$\left(\frac{dS_0(q)}{dq}\right)^2 = 2f'^2 \mathbf{r}_f^2 |E_0 - V(\mathbf{r})|,$$

i.e., (38)

$$S_0(q) = \int^q dq' \, \{2f'^2 \mathbf{r}_f^2 |E_0 - V(\mathbf{r})|\}^{1/2}$$

and

$$\mathscr{H}\chi(q,\boldsymbol{\eta}) = E_1\chi(q,\boldsymbol{\eta}), \tag{39}$$

where

$$\mathscr{H} = -\frac{1}{2}\frac{\partial^2}{\partial\eta_a^2} - i\epsilon\left(\frac{1}{f'}\frac{\partial}{\partial q} - \Gamma^b_a \eta^a \frac{\partial}{\partial\eta_b}\right) + \frac{1}{2} W_{ab}\eta_a\eta_b \tag{40}$$

and[7]

$$W_{ab} = \left(\frac{\partial^2 V}{\partial\eta_a\,\partial\eta_b}\right)_{\mathbf{r}} + \frac{3\epsilon^2}{\mathbf{r}_f^2}(\mathbf{r}_{ff}\cdot\mathbf{n}_a)(\mathbf{r}_{ff}\cdot\mathbf{n}_b). \tag{41}$$

For the diagonalization of $\mathscr{H}$ we construct a suitable set of orthonormal basic functions with the help of the classical Hamiltonian corresponding to (40), i.e.,

$$\mathscr{H}_0 = \tfrac{1}{2}\zeta_a^2 + \epsilon(p/f' - \Gamma^b_a \xi^a \zeta_b) + \tfrac{1}{2} W_{ab}\xi_a\xi_b. \tag{42}$$

Setting (see Refs. 1 and 7)

$$D_{ab} \equiv \delta_{ab}\frac{\partial}{\partial f} + \Gamma^a_b \tag{43}$$

and applying Hamilton's equations to (42) we have

$$\epsilon\,\dot{\xi}_a = \frac{\partial\mathscr{H}_0}{\partial\zeta_a}, \quad \dot{\zeta}_a = -\frac{\partial\mathscr{H}_0}{\partial\epsilon\,\xi_a}$$

and thus

$$\zeta_a = \epsilon\, D_{ab}\xi_b, \qquad \epsilon D_{ab}\zeta_b = -W_{ab}\xi_b. \tag{44}$$

Hence

$$D_{ab}D_{bc}\xi_c + \epsilon^2 W_{ab}\xi_b = 0. \tag{45}$$

This is the equation of small fluctuations with respect to a frame of reference which travels with the point q along the classical path. In the Appendix we give an alternative derivation of this equation from first principles, as this is the method usually referred to in the literature.[1,2]

We now set

$$\mathbf{v} = \begin{pmatrix}\xi_a\\ \zeta_a\end{pmatrix} \tag{46}$$

and

$$M = \frac{1}{i}\begin{pmatrix}\epsilon D_{ab} & -\delta_{ab}\\ W_{ab} & \epsilon D_{ab}\end{pmatrix}, \tag{47}$$

so that

$$M\mathbf{v} = 0. \tag{48}$$

We assume that the classical path is periodic with period T, i.e.,

$$\mathbf{r}(f) = \mathbf{r}(f+T).$$

Then we can search for solutions $\mathbf{v}$ of (48) which are such that

$$\mathbf{v}(f) = e^{-i\nu f/\epsilon}\mathbf{u}(f), \quad \mathbf{u} = \begin{pmatrix}\mu_a\\ \rho_a\end{pmatrix}, \tag{49}$$

with

$$\mathbf{u}(f+T) = \mathbf{u}(f). \tag{50}$$

Then

$$M\mathbf{u} = \nu\mathbf{u}. \tag{51}$$

In the space of periodic functions of period T we define the scalar product of two solutions $\mathbf{u}_1,\mathbf{u}_2$ by

$$(\mathbf{u}_2,\mathbf{u}_1) = \int_0^T \mathbf{u}_2^+ \sigma_2 \mathbf{u}_1 \, df. \tag{52}$$

It can be shown that the periodicity of $\mathbf{u}_1,\mathbf{u}_2$ implies that M is Hermitian, i.e.,

$$(M\mathbf{u}_2,\mathbf{u}_1) = (\mathbf{u}_2,M\mathbf{u}_1),$$

so that the eigenvalues ν are real. By complex conjugation of (51) we see that if ν_i is an eigenvalue with eigenfunction $\mathbf{u}^{(i)}$, then $-\nu_i$ is an eigenvalue with eigenfunction $\mathbf{u}^{(i)*}$. It follows that $\{u_b^{(i)}/i = 2,\ldots,N\}$ is a set of $N-1$ linearly independent eigenfunctions. We now use the $\mu_b^{(i)}$ of

$$\mathbf{u}^{(i)} = \begin{pmatrix}\mu_b^{(i)}\\ \rho_b^{(i)}\end{pmatrix}$$

as a set of orthonormal basic vectors to span the space orthogonal to $\mathbf{r}_f$ [the tangent to the path $\mathbf{r}$ at $f = f(q)$]. We write

$$\mu_{ai} \equiv \mu_a^{(i)} \quad (a,i = 2,\ldots,N)$$

and define μ^{-1} as the inverse of the matrix μ. Then the elements μ_{ai} form $N-1$ vectors with respect to index i obeying the completeness relation

$$\sum_i \mu_a^{(i)}\mu_b^{-1(i)} = \delta_{ab} \tag{53}$$

or, with

$$\xi_a^{(i)} = e^{-i\nu_i f/\epsilon}\mu_a^{(i)} \tag{54}$$

the relation

$$\sum_i \xi_a^{(i)}\xi_b^{-1(i)} = \delta_{ab}. \tag{55}$$

Considering $(\xi_a^{(i)})$ as a nonsingular $(\nu \neq 0)$ square matrix we have also

$$\sum_a \xi_a^{-1(i)}\xi_a^{(j)} = \delta_{ij}. \tag{56}$$

The existence of the inverse ξ^{-1} of ξ is ensured by the linear independence of the vectors $\boldsymbol{\xi}^{(j)}$. From (51) one can show with the help of (54) that $\xi_a^{(i)}$ is a solution of

$$D_{ab}D_{bc}\xi_c^{(i)} + \epsilon^2 W_{ab}\xi_b^{(i)} = 0 \tag{57}$$

or $\mu_a^{(i)}$ is a solution of

$$(D_{ab} - (i/\epsilon)\nu_i\delta_{ab})(D_{bc} - (i/\epsilon)\nu_i\delta_{bc})\mu_c^{(i)} + \epsilon^2 W_{ab}\mu_b^{(i)} = 0. \tag{58}$$

Now

$$\frac{\partial}{\partial\eta_b}(D_{bc}\eta_c) - (\eta_a D_{ab})\frac{\partial}{\partial\eta_b} = (N-1)\frac{\partial}{\partial f} + 2\Gamma^b_a\eta_a\frac{\partial}{\partial\eta_b}$$

and so

$$\frac{1}{2}\left(\frac{\partial}{\partial\eta_b}+i\epsilon\eta_a D_{ab}\right)\left(\frac{\partial}{\partial\eta_b}-i\epsilon D_{bc}\eta_c\right)$$

$$=\frac{1}{2}\frac{\partial^2}{\partial\eta_b^2}+\frac{\epsilon^2}{2}\eta_a D_{ab}D_{bc}\eta_c$$

$$-i\epsilon\left\{\frac{N-1}{2}\frac{\partial}{\partial f}+\Gamma^b_{\ a}\eta_a\frac{\partial}{\partial\eta_b}\right\}. \qquad (59)$$

As shown in Ref. 9, one now inserts Kronecker deltas and hence (55) between the brackets on the left and between the factors D_{bc} and η_c on the right, and considers the action of the second term on the right of (59) on a solution $\chi(f,\eta)$ expanded in the form

$$\eta_d\chi(f,\eta)=\sum_j\xi_d^{(j)}(f)\phi_j(\eta) \qquad (60)$$

in terms of solutions $\xi^{(j)}$ of the equation of small fluctuations with respect to the instantaneous frame of reference. This expansion is analogous to the familiar partial wave expansion in quantum mechanics and expresses the separation of variables associated with the problem of simultaneously diagonalizing two or more Hermitian operators, i.e., here H and M. We then obtain

$$\sum_{i=2}^{N}\frac{1}{2}\left(\frac{\partial}{\partial\eta_b}+i\epsilon\eta_a D_{ab}\right)\xi_b^{(i)}\xi_c^{-1(i)}\left(\frac{\partial}{\partial\eta_c}-i\epsilon D_{cd}\eta_d\right)$$

$$=-\mathscr{H}-\frac{i\epsilon}{2}(N+1)\frac{\partial}{\partial f}. \qquad (61)$$

We now define

$$A_i=-\frac{e^{-i\nu_i f/\epsilon}}{(\nu_i)^{1/2}}\xi_b^{-1(i)}\left(\frac{\partial}{\partial\eta_b}-i\epsilon D_{ba}\eta_a\right),$$

$$A_i^{+}=\frac{e^{+i\nu_i f/\epsilon}}{(\nu_i)^{1/2}}\left(\frac{\partial}{\partial\eta_b}+i\epsilon\eta_a D_{ab}\right)\xi_b^{(i)}. \qquad (62)$$

Then

$$\mathscr{H}=\frac{1}{2}\sum_{i=2}^{N}\nu_i A_i^{+}A_i-\frac{i\epsilon}{2}(N+1)\frac{\partial}{\partial f}. \qquad (63)$$

One can verify that

$$[A_i,A_j]=0,\quad [A_i,A_j^{+}]=\delta_{ij},\quad [A_i^{+},A_j^{+}]=0$$

$$(i,j=2,...,N) \qquad (64)$$

provided $\xi_a^{(i)}$ is normalized such that

$$\sum_{b=2}^{N}\{\xi_b^{(i)}(D\xi^{(j)})_b^{+}-\xi_b^{-1(j)}(D\xi^{(i)})_b\}=\frac{i}{\epsilon}\nu_i\delta_{ij} \qquad (65)$$

or equivalently

$$\sum_{i=2}^{N}\{\xi_a^{(i)}(D\xi^{(i)})_b^{+}-(D\xi^{(i)})_a\xi_b^{+(i)}\}=\frac{i}{\epsilon}\nu_i\,\xi_a^{+(i)}\xi_b^{(i)}.$$

Using (54), this can be written

$$\sum_{i=2}^{N}\{(D\mu^{(i)})_b^{+}\mu_a^{(i)}-\mu_b^{+(i)}(D\mu^{(i)})_a\}=-\frac{i}{\epsilon}\nu_i\,\mu_b^{+(i)}\mu_a^{(i)}. \qquad (66)$$

For the correct identification of quantum numbers in agreement with those of Sec. IV it is important to observe that the quasiparticle number operator $A_i^{+}A_i$ in (63) can assume only odd integral eigenvalues $2n_i+1, n_i=0,1,2,...$. This identification follows from comparison with the corresponding treatment of the simple harmonic oscillator [cf., e.g., Eqs. (21)–(26) of Ref. 7]. Hence quantization of the fluctuations η_a implies

$$\mathscr{H}=\sum_{i=2}^{N}\nu_i\left(n_i+\frac{1}{2}\right)-\frac{i\epsilon}{2}(N+1)\frac{\partial}{\partial f}.$$

This agrees with (34) of the adiabatic case provided $\nu_i=\frac{1}{2}h_i^2$, as will be seen to be the case.

The operators (62) can be reexpressed in another form with the help of (60). Using in addition (65) and (66) we have

$$\xi_b^{-1(i)}D_{ba}\eta_a\chi$$

$$=\sum_j\xi_b^{-1(i)}(D\xi^{(j)})_b\phi_j(\eta)$$

$$=-\frac{i}{\epsilon}\nu_i\sum_j\delta_{ij}\phi_j(\eta)+\sum_j(D\xi^{(i)})_b^{+}\xi_b^{(j)}\phi_j(\eta)$$

$$=-\frac{i}{\epsilon}\nu_i\sum_a\xi_a^{-1(i)}\eta_a\chi+(D\xi^{(i)})_b^{+}\eta_b\chi.$$

Inserting this into A_i we obtain

$$A_i=-\frac{e^{-i\nu_i f/\epsilon}}{(\nu_i)^{1/2}}\times\left(\xi_b^{-1(i)}\frac{\partial}{\partial\eta_b}-i\epsilon(D\xi^{(i)})_b^{+}\eta_b-\nu_i\xi_a^{-1(i)}\eta_a\right).$$

Recalling that

$$\xi_a^{(i)}(f)=e^{-i\nu_i f/\epsilon}\mu_a^{(i)}(f),$$

we have

$$A_i=-\frac{1}{(\nu_i)^{1/2}}\left(\mu_b^{-1(i)}\frac{\partial}{\partial\eta_b}-i\epsilon(D\mu^{(i)})_b^{+}\eta_b\right) \qquad (67a)$$

and correspondingly

$$A_i^{+}=\frac{1}{(\nu_i)^{1/2}}\left(\mu_b^{(i)}\frac{\partial}{\partial\eta_b}-i\epsilon(D\mu^{(i)})_b\eta^b\right). \qquad (67b)$$

Two points should be noted with regard to these expressions: the factor $(D\mu)$ is no longer an operator but a c number, and the periodicity of these creation and annihilation operators follows from the periodicity of the function $\mu_b^{(i)}$ (assuming that $\mathbf{n}_a$ is periodic).

In view of (67) we can write (63)

$$\mathscr{H}=\frac{1}{2}\sum_{i=2}^{N}\nu_i A_i^{+}A_i+\frac{\epsilon}{2}(N+1)\frac{1}{f'}p, \qquad (68)$$

where we have replaced $\partial/\partial f$ by the appropriate momentum operator. The operator p in (68) has to be quantized along with its conjugate variable q. We write the corresponding creation and annihilation operators a_1^{+}, a_1. Then p and q can be expressed in terms of a_1, a_1^{+}, and these commute with A_i, A_i^{+}, since there are altogether N independent degrees of freedom.

We observe that (68) does not yet exhibit explicitly the zero-point energy generally associated with each quantized oscillator degree of freedom. We can arrange this, however, in the following way. We consider

$$-\left[\mu_a^{(i)}\frac{\partial}{\partial\eta_a}-i\epsilon(D\mu^{(i)})_a\eta_a-\nu_i\mu_a^{(i)}\eta_a\right]$$
$$\times\left[\mu_b^{+(i)}\frac{\partial}{\partial\eta_b}-i\epsilon(D\mu^{(i)})_b^+\eta_b+\nu_i\mu_b^{+(i)}\eta_b\right]$$
$$=\nu_iA_i^+A_i+\nu_i^2\mu_a^{(i)}\mu_b^{+(i)}\eta_a\eta_b$$
$$+\nu_i\mu_a^{(i)}\mu_b^{+(i)}\left(\eta_a\frac{\partial}{\partial\eta_b}-\frac{\partial}{\partial\eta_a}\eta_b\right)$$
$$+i\epsilon\nu_i\eta_a\eta_b\left\{(D\mu^{(i)})_a\mu_b^{+(i)}-\mu_a^{(i)}(D\mu^{(i)})_b^+\right\}.$$

Using (66) this becomes

$$\nu_iA_i^+A_i+\nu_i^2\mu_a^{(i)}\mu_b^{+(i)}\eta_a\eta_b-\nu_i\mu_a^{(i)}\mu_a^{+(i)}$$
$$-\nu_i^2\eta_a\eta_b\mu_a^{+(i)}\mu_b^{(i)}$$
$$=\nu_iA_i^+A_i-\nu_i\mu_a^{+(i)}\mu_a^{(i)}$$
$$=\sum_i\nu_i[A_i^+A_i-1].$$

Thus, defining

$$\widetilde{A}_i=\frac{1}{(\nu_i)^{1/2}}\left[\mu_a^{+(i)}\frac{\partial}{\partial\eta_a}-i\epsilon(D\mu^{(i)})_a^+\eta_a+\nu_i\mu_a^{+(i)}\eta_a\right]$$
$$=\frac{e^{-i\nu_if/\epsilon}}{(\nu_i)^{1/2}}\left[\xi_a^{+(i)}\frac{\partial}{\partial\eta_a}-i\epsilon(D\xi^{(i)})_a^+\eta_a\right],\tag{69a}$$
$$\widetilde{A}_i^+=-\frac{1}{(\nu_i)^{1/2}}\left[\mu_a^{(i)}\frac{\partial}{\partial\eta_a}-i\epsilon(D\mu^{(i)})_a\eta_a-\nu_i\mu_a^{(i)}\eta_a\right]$$
$$=-\frac{e^{+i\nu_if/\epsilon}}{(\nu_i)^{1/2}}\left[\xi_a^{(i)}\frac{\partial}{\partial\eta_a}-i\epsilon(D\xi^{(i)})_a\eta_a\right],\tag{69b}$$

we have

$$\frac{1}{2}\sum_i\nu_iA_i^+A_i=\sum_i\nu_i\left\{\frac{1}{2}\widetilde{A}_i^+\widetilde{A}_i+\frac{1}{2}\right\}$$

and (70a)

$$\mathscr{H}=\sum_i\nu_i\left[\frac{1}{2}\widetilde{A}_i^+\widetilde{A}_i+\frac{1}{2}\right]-\frac{i\epsilon}{2}(N+1)\frac{\partial}{\partial f}.$$

The operator $\widetilde{A}_i$ is similar to a corresponding operator conjectured previously by Gervais and Sakita.[1]

We have observed earlier that quantization implies $A_i^+A_i\to 2n_i+1$, $n_i=0,1,2,\ldots$. It follows therefore from (70a) that the eigenvalues of $\widetilde{A}_i^+\widetilde{A}_i$ are even integers $2n_i$. Hence again

$$\mathscr{H}=\sum_i\nu_i\left(n_i+\frac{1}{2}\right)-\frac{i\epsilon}{2}(N+1)\frac{\partial}{\partial f}.\tag{70b}$$

It may be noted that $\widetilde{A}_i$, $\widetilde{A}_i^+$ have the same general form as the corresponding operators in general multidimensional quantum mechanics.[7] They satisfy the commutation relations (64) on account of (65).

It can be seen that $\mathscr{H}$ and the $2(N-1)$ linearly independent operators $\widetilde{A}_i$, $\widetilde{A}_i^+$ form the Cartan basis of a Lie algebra. This follows from various properties of the $\widetilde{A}_i$. In particular one has

$$[\mathscr{H},\widetilde{A}_i]=-\nu_i/2\widetilde{A}_i,$$
$$[\mathscr{H},\widetilde{A}_i^+]=+(\nu_i/2)\widetilde{A}_i^+.$$

The components ν_i of the vector $\mathbf{v}=(\nu_2,\ldots,\nu_N)$ are then the roots of the Cartan basis. We have seen above, that if $\mathbf{v}/2$ is a root, then $-\mathbf{v}/2$ is also a root and the total number of roots is even.

Finally we observe that in the nonadiabatic case the ground state wave function $\chi_0(f,\eta)$ is defined by

$$\widetilde{A}_i\,\chi_0(f,\eta)=0,\quad i=2,\ldots,N,\tag{71}$$

up to an arbitrary function of f. The perturbation method then proceeds in the usual way. We do not elaborate on the wave functions here since these will be dealt with in detail in Secs. VIII–X.

VI. THE MATCHING OF ADIABATIC AND NONADIABATIC APPROXIMATIONS IN THEIR OVERLAP REGION

Above we have seen that the adiabatic expansion is valid in a small domain of $o(1/h_c)$ around a minimum of the potential, and that the nonadiabatic expansion is valid in the region beyond this small domain. We now demonstrate that both types of expansion merge into each other in the transition region where they overlap. We consider only the matching of the quantization operators here; the matching of wave functions will be considered in Sec. X.

We recall that $\mathbf{n}_a\cdot\mathbf{r}_f=0$ so that $\mathbf{n}_{af}\cdot\mathbf{r}_f+\mathbf{n}_a\cdot\mathbf{r}_{ff}=0$. Since $\mathbf{r}_{ff}=-(\partial V/\partial\mathbf{R})_\mathbf{r}\to 0$ as we approach the minimum at $\mathbf{r}_0$, we see that (with $|f_0|=\infty$)

$$\lim_{f\to f_0}\mathbf{n}_{a_f}(f)=0.\tag{72}$$

Hence

$$W_{ab}=\frac{\partial^2V(\mathbf{r})}{\partial\eta_a\,\partial\eta_b}+\frac{3\epsilon^2}{\mathbf{r}_f^2}(\mathbf{r}_{ff}\cdot\mathbf{n}_a)(\mathbf{r}_{ff}\cdot\mathbf{n}_b)$$
$$\to\frac{\partial^2\mathrm{V}(\mathbf{r}_0)}{\partial\eta_a\,\partial\eta_b}=M_{ab}$$

and

$$\Gamma_b^a=\mathbf{n}_a\cdot\mathbf{n}_{b_f}\to 0\quad\text{as}\quad\mathbf{r}\to\mathbf{r}_0.$$

Thus (57) becomes (as $\mathbf{r}\to\mathbf{r}_0$)

$$\frac{\partial^2}{\partial f^2}\xi_a^{(i)}+\epsilon^2M_{ab}\,\xi_b^{(i)}=0.\tag{73}$$

Using

$$\xi_a^{(i)}=e^{-i\nu_if/\epsilon}\mu_a^{(i)}$$

and[7]

$$\nu_i\sim(V_{ii}^{(2)})^{1/2}=\tfrac{1}{2}h_i^2,$$

one can show that

$$\begin{array}{ccc}\widetilde{A}_i,\widetilde{A}_i^+ & \overset{f\to f_0}{\longrightarrow} & A_i,A_i^+\\ \text{nonadiabatic} & & \text{adiabatic}\end{array}\tag{74}$$

apart from irrelevant overall phase factors, where $\widetilde{A}_i^+,\widetilde{A}_i$ are the creation–annihilation operators (69a), (69b) of the nonadiabatic case, whereas A_i^+,A_i are the corresponding operators in the adiabatic approximation (32). Explicitly in the limit $f\to f_0$

$$\widetilde{A}_i\to\frac{e^{-i\nu_if/\epsilon}}{\nu_i^{1/2}}\xi_c^{-1(i)}\left(\frac{\partial}{\partial\eta_c}+\nu_c\eta_c\right).$$

With $z_c = 2(\nu_c)^{1/2}\eta_c$ and using the normalization condition (65)

$$\xi_c^{(i)}\xi_c^{1+(i)} \xrightarrow{f\to f_0} \tfrac{1}{2},$$

we have

$$\widetilde{A}_i \to -\left(\frac{\partial}{\partial z_c} + \frac{1}{2} z_c\right) e^{i\nu_i f/\epsilon}\delta_{ci} . \tag{75}$$

VII. THE GROUND-STATE WAVE FUNCTIONAL AND THE BOHR–SOMMERFELD QUANTIZATION CONDITION

Our aim here is to derive the explicit ground-state WKB wave functional χ_0 defined by the condition

$$\widetilde{A}_i(f)\chi_0(f,\boldsymbol{\eta}) = 0, \quad i = 2,\ldots,N,$$

where $\widetilde{A}_i$ is given by (69a). Inserting (69a), we obtain a differential equation from which χ_0 may be determined. Thus, using (55), we have

$$\frac{\partial}{\partial \eta_a}\chi_0(f,\boldsymbol{\eta}) + \Omega_{ab}(f)\eta_b\chi_0(f,\boldsymbol{\eta}) = 0,$$

where

$$\Omega_{ab}(f) = -i\epsilon \sum_{i=2}^{N} \xi_a^{(i)}(D\xi^{(i)})_b^+ . \tag{76}$$

Hence

$$\chi_0(f,\boldsymbol{\eta}) = d(f)\exp\{-\tfrac{1}{2}\Omega_{ab}(f)\,\eta^a\eta^b\}, \tag{77}$$

where $d(f)$ remains to be determined.

With the help of (55), i.e.,

$$\xi(f)\xi^{-1}(f) = \xi(f)\xi^+(f) = 1 \tag{78}$$

and so

$$\xi\frac{\partial\xi^+}{\partial f} + \frac{\partial\xi}{\partial f}\xi^+ = 0,$$

it follows that

$$\Omega(f) = \Omega^+(f),$$

since [using $\Gamma^+ = -\Gamma$ and (78)]

$$\begin{aligned} i(\Omega - \Omega^+) &= \xi(D\xi)^+ + (D\xi)\xi^+ \\ &= +\xi\frac{\partial\xi^+}{\partial f} + \frac{\partial\xi}{\partial f}\xi^+ + \xi(\Gamma\xi)^+ + (\Gamma\xi)\xi^+ \\ &= 0. \end{aligned}$$

In order to determine the f-dependent factor $d(f)$ of $\chi_0(\mathrm{f},\boldsymbol{\eta})$ we substitute (77) into (39). Performing the differentiations with respect to η one obtains

$$\begin{aligned} &\left\{-i\epsilon\frac{\partial}{\partial f} - E_1 + \frac{1}{2}\sum_{c=2}^{N}\Omega_{cc}(f)\right\}d(f) \\ &\quad -\tfrac{1}{2}d(f)\left\{\Omega_{ac}\Omega_{cb} - i\epsilon\left(\frac{\partial}{\partial f}\Omega_{ab} - 2\Gamma^c_a\Omega_{cb}\right)\right. \\ &\qquad\qquad \left. - W_{ab}\right\}\eta^a\eta^b = 0. \end{aligned}$$

With some cumbersome manipulations [using (76), (78), and (58)] the expression in the second bracket can be shown to vanish. The solution of the resulting first-order differential equation is

$$d(f) = \exp\left(\frac{iE_1 f}{\epsilon}\right)\exp\left(-\frac{i}{2}\int\sum_{c=2}^{N}\Omega_{cc}(f)\,df\right).$$

In the argument of the second exponential we replace Ω by (76) and therein D by (43). In view of (55) the term containing Γ^b_a reduces to $\Gamma^c_c = 0$. The argument of the remaining term is

$$\begin{aligned} &-\frac{1}{2}\sum_{i,a=2}^{N}\int\xi_a^{(i)}(f)\frac{\partial\xi_a^{+(i)}(f)}{\partial f}df \\ &\quad = -\frac{1}{2}\sum_{i,a=2}^{N}\int(\xi_a^{+(i)}(f))^{-1}\frac{\partial\xi_a^{+(i)}(f)}{\partial f}df \\ &\quad = -\frac{1}{2}\mathrm{Tr}\int\frac{\partial}{\partial f}(\ln\xi^+)\,df \\ &\quad = -\tfrac{1}{2}\mathrm{Tr}\ln\xi^+ \\ &\quad = \ln(\det\xi^+)^{-1/2}. \end{aligned}$$

Hence

$$d(f) = e^{iE_1 f/\epsilon}/(\det\xi^+)^{1/2}$$

and

$$\chi_0(f,\boldsymbol{\eta}) = \frac{e^{iE_1 f/\epsilon}}{(\det\xi^+)^{1/2}}\exp\{-\tfrac{1}{2}\Omega_{ab}\eta^a\eta^b\} \tag{79}$$

or [using (37)]

$$\psi_0(q,\boldsymbol{\eta}) = \frac{e^{i\epsilon S_0(q)}}{[f'\mathbf{r}_f^2]^{1/2}}\frac{e^{i/\epsilon E_1 f(q)}}{(\det\xi^+)^{1/2}}e^{-(1/2)\Omega_{ab}(f(q))\eta^a\eta^b}. \tag{80}$$

In Sec. II we pointed out that large fluctuations η_a can lead to ambiguities which invalidate the perturbation procedure. For reasons of consistency it is therefore necessary to demand that the wave function Ψ falls off sufficiently rapidly away from the classical path. We see from (80) that this is ensured if the real part of Ω is positive definite.

The Bohr–Sommerfeld quantization condition is obtained from the requirement that the ground-state wave functional be periodic, i.e.,

$$\psi_0(f+T) = \psi_0(f), \tag{81}$$

where ψ_0 is given by (80). In this expression we have to insert

$$S_0(q) \equiv S_0(f(q)) = \pm\int^f\{2\,\mathbf{r}_f^2(f')|E_0 - V(\mathbf{r}(f'))|\}^{1/2}df'.$$

Then

$$S_0(f+T) = S_0(f) + W(E_0), \tag{82}$$

where $W(E)$ is the classical action over one complete period, i.e.,

$$W(E) = \oint\{2\mathbf{r}_f^2|E - V(\mathbf{r}(f))|\}^{1/2}df, \tag{83}$$

and, since $E_0 - V(\mathbf{r}) = \frac{1}{2}\mathbf{r}_f^2$,

$$W(E_0) = \oint\mathbf{r}_f\cdot d\mathbf{r}. \tag{84}$$

Further, if λ_j, $j = 2,\ldots,N$ are the eigenvalues of $(\xi_a^{+(i)})$ we have

$$\det(\xi^+(f)) = \prod_{j=2}^{N}\lambda_j(f)$$

and

$$\det(\xi^+(f+T)) = \det(\xi^+(f)) \exp + \left(\frac{i}{\epsilon}\sum_i \nu_i T\right). \qquad (85)$$

Finally we observe that

$$\Omega_{ab}(f+T) = \Omega_{ab}(f). \qquad (86)$$

Inserting (80) into (81) and using (82), (85), and (86) we obtain

$$\exp\left\{i\epsilon W(E_0) - \frac{i}{2\epsilon}\sum_{j=2}^{N} \nu_j T + \frac{i}{\epsilon} E_1^{(0)} T\right\} = \exp\{2mi\pi\},$$

where m is an integer. Hence with $\epsilon^2 = 1$, i.e., considering the classically allowed region,

$$W(E_0) + T\left\{E_1^{(0)} - \sum_{i=2}^{N} \frac{1}{2}\nu_i\right\} = 2m\pi. \qquad (87)$$

We can rewrite this equation in a more familiar form[1,10,11] by setting [see, e.g., (35), (70b), or (111) below]

$$E_1 = E_1^{(0)} + \sum_{a=2}^{N} n_a \nu_a, \qquad (88)$$

where $n_a = 0,1,2,...$ is the occupation number of the ath state. Now

$$W(E) \simeq W(E_0) + (E - E_0)\left(\frac{\partial W}{\partial E}\right)_{E_0}$$

and

$$\left(\frac{\partial W}{\partial E}\right)_{E_0} = \oint \frac{\mathbf{r}_f^2\, df}{\{2\mathbf{r}_f^2|E_0 - V(\mathbf{r})|\}^{1/2}}, \qquad (89)$$

where $E_0 - V(\mathbf{r}) = \frac{1}{2}\mathbf{r}_f^2$. Hence $(\partial W/\partial E)_{E_0} = T$ and

$$W(E) = W(E_0) + (E - E_0)T = W(E_0) + E_1 T. \qquad (90)$$

Using (87) and (88) we can rewrite this equation as

$$W(E) = 2m\pi + \sum_{a=2}^{N}\left(n_a + \frac{1}{2}\right)\nu_a T. \qquad (91)$$

Equation (91) may be looked at as the multidimensional generalization of the one-dimensional WKB quantization formula. The quantum number m is the main quantum number; in semiclassical language it describes the number of complete waves that can be supported by the classical path. The $N-1$ quantum numbers n_a describe the quantum excitations orthogonal to the classical path.

VIII. WAVE FUNCTIONALS IN THE ADIABATIC CASE

In Sec. II it was shown that the Hamilton operator of the system can be written as

$$H = V(\mathbf{r}(f)) + H_S + O(g^3), \qquad (92)$$

where

$$H_S = \frac{\partial V(\mathbf{r}(f))}{\partial \eta_a}\eta_a + \frac{1}{2}(p_a^2 + \eta_a M_{ab}\eta_b)$$
$$+ \frac{1}{\mathbf{r}_f^2}\left(\frac{p^2}{2f'^2} - \frac{p\Gamma_a^b \eta_a p_b}{f'}\right) \qquad (93)$$

(summations understood) the terms in H_S being, respectively, of order -1, 0, and 2 in g for p of order 0. In deriving the fundamental wave functional (i.e., the so-called unperturbed wave functional in terms of which the perturbation expansion is to be constructed) we have to take each of these terms into account (as in the Schrödinger equation), and the perturbation H_I is of order g^3.

As before, we assume that $\mathbf{r}(f)$ is periodic in Re f with period T. Then in approaching a minimum of $V(\mathbf{r}(f))$ at $\mathbf{r} = \mathbf{r}_0$ (or Im $f \to -\infty$)

$$\frac{\partial V}{\partial \eta_a} \to 0, \quad \Gamma_a^b \to 0$$

(as explained in Sec. VI), so that in a small domain around $\mathbf{r}_0$

$$V(\mathbf{r}) + H_S = -\frac{1}{2}\frac{\partial^2}{\partial \eta_a^2} + \frac{1}{2}\eta_a M_{ab}\eta_b$$
$$-\frac{1}{2\mathbf{r}_f^2}\frac{\partial^2}{\partial f^2} + V(\mathbf{r}(f)). \qquad (94)$$

Ultimately we wish to solve the Schrödinger equation

$$H\Psi = E\Psi.$$

To lowest order, H is given by (28) and Ψ by

$$\Psi^{(0)} = \phi(\mathbf{r}(f))\,\tilde{\phi}(\eta), \qquad (95)$$

with $E = E_1^{(0)} + \Delta \simeq E_1^{(0)}$ since the variables of (94) [$f(q)$ and η] are separable. The ansatz (95) also implies that we search for periodic solutions of period T. We thus obtain the two sets of equations

$$\left[-\frac{1}{2\mathbf{r}_f^2}\frac{\partial^2}{\partial f^2} + V(\mathbf{r}(f))\right]\phi(\mathbf{r}(f)) = \epsilon^{(0)}\phi(\mathbf{r}(f)), \qquad (96)$$

$$\left[-\frac{1}{2}\frac{\partial^2}{\partial \eta_a^2} + \frac{1}{2}\eta_a M_{ab}\eta_b\right]\tilde{\phi}(\eta) = \tilde{\epsilon}^{(0)}\tilde{\phi}(\eta), \qquad (97)$$

with

$$E_1^{(0)} = \epsilon^{(0)} + \tilde{\epsilon}^{(0)}. \qquad (98)$$

Now, the classical equation of motion is

$$\frac{d}{d\mathbf{r}}\left(\tfrac{1}{2}\mathbf{r}_f^2\right) = \mathbf{r}_{ff} = -\frac{\partial V(\mathbf{r})}{\partial \mathbf{r}} \qquad (99)$$

or

$$\tfrac{1}{2}\mathbf{r}_f^2 = V(\mathbf{r}_0) - V(\mathbf{r}). \qquad (100)$$

The classical energy is zero for $V(\mathbf{r}_0) = 0$, so that the turning points are shifted to the extremum (this is an important ingredient of the perturbation method of Refs. 12). We now let $\mathbf{r}$ be the vector with components x_j and we write

$$\dot{x}_j \equiv \frac{dx_j}{df}.$$

Expanding $V(\mathbf{r})$ around $\mathbf{r}_0$ we have

$$\frac{1}{2}\sum_j \dot{x}_j^2 = -\frac{1}{8}\sum_j (x_j - x_{j0})^2 h_j^4 + \cdots, \qquad (101)$$

where

$$V_{jj}^{(2)}(\mathbf{r}_0) = \tfrac{1}{4}h_j^4.$$

(For simplicity we ignore off-diagonal quadratic terms or assume that the quadratic terms have already been rotated to normal form.) From (101) we deduce on physical grounds that to lowest order

$$\dot{x}_j^2 = -\tfrac{1}{4}(x_j - x_{j0})^2 h_j^4 + \delta_j, \qquad (102)$$

for every independent coordinate x_j.

The solutions of (102) are of the form

$$x_j - x_{j0} \propto e^{\pm (i/2) h_j^2 f} \tag{103}$$

[where Re f is the variable in terms of which $\mathbf{r}(f)$ is periodic]. Of course, (103) would also follow from Newton's equation. In approaching the minimum at $\mathbf{r}_0$ the lower sign of (103) has to be chosen ($\mathrm{Im} f \to -\infty$). Close to a minimum of $V(\mathbf{r}(f))$, we can therefore write [since $\partial/\partial f$ acts on $\phi(\mathbf{r}(f))$]

$$\begin{aligned}
&-\frac{1}{2\mathbf{r}_f^2}\frac{\partial^2}{\partial f^2} \\
&= -\frac{1}{2\Sigma_j \dot{x}_j^2}\sum_{l,k}\dot{x}_l\frac{\partial}{\partial x_l}\dot{x}_k\frac{\partial}{\partial x_k} \\
&\simeq -\frac{1}{2}\frac{\Sigma_{l,k} h_l^2 h_k^2 (x_l - x_{l0})(\partial/\partial x_l)(x_k - x_{k0})(\partial/\partial x_k)}{\Sigma_j h_j^4 (x_j - x_{j0})^2} \\
&= -\frac{1}{2}\sum_j\left[\frac{\partial^2}{\partial x_j^2} + \frac{1}{(x_j - x_{j0})}\frac{\partial}{\partial x_j} + o\left(\frac{x_k - x_{k0}}{x_j - x_{j0}}\right)^2\right].
\end{aligned} \tag{104}$$

The last expression results from expanding the contribution of the jth degree of freedom for $|x_k - x_{k0}/x_j - x_{j0}| \ll 1$ in case $k \neq j$, this being the condition for singling out the independent oscillators, one for each degree of freedom. In this adiabatic (i.e., oscillatorlike) approximation, valid in a narrow domain of $o(1/h_j)$ around the minimum of the potential, we can therefore rewrite (96) as

$$\left[-\frac{1}{2}\frac{\partial^2}{\partial x_j^2} - \frac{1}{2}\frac{1}{(x_j - x_{j0})}\frac{\partial}{\partial x_j} + \frac{1}{2}(x_j - x_{j0})\, V_{jk}^{(2)}(\mathbf{r}_0)(x_k - x_{k0})\right]\phi = \epsilon^{(0)}\phi. \tag{105}$$

The η-dependent equation (97) can be brought into normal form by the substitution

$$\eta = A^T y,$$

with

$$A^T A = 1, \quad (A^T M A)_{ab} = \tfrac{1}{4} h_a^4 \delta_{ab}. \tag{106}$$

In (105) we ignore the x_j, x_k coupling terms for simplicity (or as an approximation) and set

$$V_{ij}^{(2)}(\mathbf{r}_0) = \tfrac{1}{4} h_i^4 \cdot \delta_{ij}. \tag{107}$$

We now set

$$z_j = h_j(x_j - x_{j0}), \quad z_a = h_a y_a \tag{108}$$

(no summation) so that (96) and (97) become

$$-\frac{1}{2}\sum_j h_j^2\left(\frac{\partial^2}{\partial z_j^2} + \frac{1}{z_j}\frac{\partial}{\partial z_j} - \frac{1}{4}z_j^2\right)\phi = \epsilon^{(0)}\phi, \tag{109}$$

$$-\frac{1}{2}\sum_a h_a^2\left(\frac{\partial^2}{\partial z_a^2} - \frac{1}{4}z_a^2\right)\tilde{\phi} = \tilde{\epsilon}^{(0)}\tilde{\phi}. \tag{110}$$

We now make the following ansatz for the energy E (Δ being the contribution of H_I):

$$\begin{aligned}
&E = V(\mathbf{r}_0) + E_1^{(0)} + \Delta, \quad E_1^{(0)} = \epsilon^{(0)} + \tilde{\epsilon}^{(0)}, \\
&\epsilon^{(0)} = \sum_j \frac{1}{4}(p_j - 1)h_j^2, \quad \tilde{\epsilon}^{(0)} = \sum_a \frac{1}{4}p_a h_a^2.
\end{aligned} \tag{111}$$

Here p_a is exactly or only approximately an odd integer depending on whether $V(\mathbf{R}) = V(\mathbf{r} + \boldsymbol{\eta})$ is an oscillator potential in η_a or only approximately so (i.e., allows tunneling). This type of dependence is not the subject of the present investigation. However, since we assume that $\mathbf{r}(f)$ is periodic, $V(\mathbf{r})$ is also periodic and thus possesses minima separated by humps of finite height. Thus p_j will be seen to be only approximately an odd integer and, in fact, the calculation of its deviation from an exact odd integer is the main task of these considerations. We shall see that the extra " -1" ensures that the zero-point energy is zero in the approximation of the purely classical contribution. Of course, the set of all p_a, p_j defines a set of $2N - 1$ quantum numbers. Since we have only N degrees of freedom, we can have only N independent quantum numbers. Thus the p_a can be expressed in terms of the p_j, the two sets resulting from two sets of basic eigenfunctions which are related to each other by a unitary transformation.[13] We return to this point at the end of our calculation. For our present purposes and our demonstration of the existence of a band structure of the eigenvalues this point is irrelevant. We might remark that in this respect our procedure is similar to that of Ref. 2, although, of course, the calculation itself is completely different.

Inserting (111) into (109) and (110) we obtain

$$\sum_j \frac{1}{4}h_j^2 \mathscr{D}_{p_j}\phi = 0 \tag{112}$$

and

$$\sum_a \frac{1}{4}h_a^2 \tilde{\mathscr{D}}_{p_a}\tilde{\phi} = 0, \tag{113}$$

where

$$\mathscr{D}_{p_j} = 2\left[\frac{\partial^2}{\partial z_j^2} - \frac{(-\frac{1}{2})(\frac{1}{2})}{z_j^2} + \tfrac{1}{2}(p_j - 1) + \tfrac{1}{4}z_j^2\right] \tag{114}$$

and

$$\tilde{\mathscr{D}}_{p_a} = 2\left[\frac{\partial^2}{\partial z_a^2} + \frac{1}{2}p_a - \frac{1}{4}z_a^2\right]. \tag{115}$$

The solutions of (113) are products of parabolic cylinder functions, i.e.,

$$\tilde{\phi} \equiv \prod_a \tilde{\phi} = \prod_a D_{(1/2)(p-1)}(z_a), \tag{116}$$

and describe the quantum fluctuations orthogonal to the classical path. They are the eigenfunctions that result from the application of creation operators A_a^+ to the ground state wave function ϕ_0 defined by $A_a\phi_0 = 0$, where A_a^+, A_a are the operators given by (32) for the domain around a minimum.

The solutions of (112) are of more interest since they describe the tunneling from one potential well to another in terms of the collective variable. It is therefore necessary to investigate these in some detail. Previously[3-5] the problem of matching solutions which are valid in neighboring domains has only been considered for the case of no centrifugal term. Thus we have to develop the appropriate generalization of the procedure. For this purpose we consider the equation

$$\left(\frac{d^2}{dz^2} + l + \frac{p}{2} - \frac{l(l+1)}{z^2} - \frac{z^2}{4}\right)\phi_p^l = 0. \tag{117}$$

In our case

$$\phi = \prod_j \phi(p_j, h_j^2; z_j) = \prod_j (z_j^{-1/2} \phi_p^{(-1/2)}(z_j)). \tag{118}$$

The solution of (117) which is the direct generalization of the Weber parabolic cylinder function $D_{(1/2)(p-1)}(z)$ is

$$\phi_p^l(\pm z) = \frac{\pi^{1/2} 2^{l+(1/4)(p-1)} e^{-(1/4)z^2}}{z^l \Gamma[-\frac{1}{4}(p-3)]} \times F\left(-l - \frac{1}{4}(p-1), -l + \frac{1}{2}; \frac{z^2}{2}\right) \mp \frac{\pi^{1/2} 2^{(1/4)(p+1)} z^{l+1} e^{-(1/4)z^2}}{\Gamma[-l - \frac{1}{4}(p-1)]} \times F\left(-\frac{1}{4}(p-3), l + \frac{3}{2}; \frac{z^2}{2}\right), \tag{119}$$

where F is a confluent hypergeometric function. For $l = 0$ we have

$$\phi_p^0(\pm z) = D_{(1/2)(p-1)}(\pm z).$$

We now put

$$\phi_p^{(-1/2)} \equiv \phi_p .$$

Then

$$\phi_p(\pm z) = \alpha_p\, z^{1/2} e^{-(1/4)z^2} F(-\tfrac{1}{4}(p-3), 1; \tfrac{1}{2} z^2), \tag{120}$$

where

$$\alpha_p = \frac{\pi^{1/2} 2^{(1/4)(p-1)} (2^{-1/2} \mp 2^{+1/2})}{\Gamma[-\frac{1}{4}(p-3)]}. \tag{121}$$

The asymptotic behavior of the confluent hypergeometric function F is given by

$$F(a,b;z) \simeq [\Gamma(b)/\Gamma(b-a)]\, e^{i\pi a} z^{-a} + [\Gamma(b)/\Gamma(a)]\, e^z z^{a-b} \quad (-\pi/2 < \arg z < \tfrac{3}{2}\pi).$$

Hence

$$\phi_p(z) \simeq \alpha_p z^{1/2} e^{-(1/4)z^2} \times \left\{ \frac{e^{-i(\pi/4)(p-3)} (\frac{1}{2} z^2)^{(1/4)(p-3)}}{\Gamma[\frac{1}{4}(p+1)]} + \frac{e^{(1/2)z^2} (\frac{1}{2} z^2)^{-(1/4)(p+1)}}{\Gamma[-\frac{1}{4}(p-3)]} \right\}. \tag{122}$$

We now define solutions ϕ_p^B, $\bar{\phi}_p^C$ of (117) which have the asymptotic behavior

$$\phi_p^B(z) \simeq z^{1/2} e^{-(1/4)z^2} z^{(1/2)(p-3)}, \qquad \bar{\phi}_p^C(z) \simeq z^{1/2} e^{+(1/4)z^2} (z^{1/2(p+1)})^{-1}, \tag{123}$$

and are valid in the same domain of $o(1/h)$ around a minimum of the potential [we keep the factor $z^{1/2}$ separate because this cancels the factor $1/z^{1/2}$ in (118)]. Thus (123) is the behavior of these solutions at their boundary of validity. The existence of solutions having the behavior (123) follows immediately from the equation $\mathscr{D}_p \phi_p = 0$ which remains unchanged under the combined replacements

$$z \to iz, \quad (p-1) \to -(p-1),$$

i.e.,

$$z^2 \to -z^2, \quad (p-3) \to -(p+1).$$

The solutions $\phi(p_j, h_j^2; x_j)$ with the behavior (123) for each degree of freedom can be written

$$\phi_B(p_j, h_j^2; x_j) = \prod_j \alpha_{p_j} e^{-1/4 z_j^2} F(-\tfrac{1}{4}(p_j - 3), 1; \tfrac{1}{2} z_j^2) \simeq \prod_j \alpha_{p_j} \frac{\mathrm{Re}\, e^{-i(\pi/4)(p_j-3)} (\frac{1}{2} z_j^2)^{(1/4)(p_j-3)} e^{-1/4 z_j^2}}{\Gamma[\frac{1}{4}(p_j+1)]}, \quad \text{for } |z_j| \sim 1 \tag{124}$$

(where Re means "real part of") and

$$\bar{\phi}_c(p_j, h_j^2; x_j) = \prod_j \alpha_{p_j} e^{-(1/4) z_j^2} F\left(-\frac{1}{4}(p_j - 3), 1; \frac{z_j^2}{2}\right) \simeq \prod_j \alpha_{p_j} \frac{e^{+(1/4) z_j^2}}{(\frac{1}{2} z_j^2)^{(1/4)(p_j+1)} \Gamma[-\frac{1}{4}(p_j - 3)]} \quad \text{for } |z_j| \sim 1. \tag{125}$$

It should be observed that the first expression is the same in each case and is due to the special case $b = 1$ of $F(a,b;x)$. The solutions (124) and (125) are valid for $x_j - x_{j0} < o(1/h_j)$, $|z_j| < 1$. Solutions $\bar{\phi}_B, \phi_c$ valid around $-x_{j0}$ are obtained by replacing x_{j0} by $-x_{j0}$.

IX. WAVE FUNCTIONALS IN THE NONADIABATIC CASE

We return to our considerations of Sec. V which are valid for $|x_j - x_{j0}| > o(1/h)$. We are interested in the behavior of the wave functional Ψ at the boundary of the domain of validity where the adiabatic and nonadiabatic approximations overlap. We consider first S_0, i.e.,

$$S_0 = 2^{1/2} \int^f df \{ \mathbf{r}_f^2 (V(\mathbf{r}_0) - V(\mathbf{r})) \}^{1/2} = \int^r \sum_j \left[(x_j - x_{j0})^2 \frac{h_j^4}{4} \frac{dx_j}{\dot{x}_j} \right] = \int^r \sum_j \left[(x_j - x_{j0})^2 \frac{h_j^4}{4} \frac{2\, dx_j}{i h_j^2 (x_j - x_{j0})} \right] = -\frac{i}{4} \sum_j [(x_j - x_{j0})^2 - x_{j0}^2]\, h_j^2. \tag{126}$$

Hence

$$\exp(\pm i S_0) \simeq \exp\left(\pm \sum_j \frac{h_j^2}{4} [(x_j - x_{j0})^2 - x_{j0}^2] \right). \tag{127}$$

We observe that the signs $\pm$ can be looked at as arising from either of the replacements $f \to -f$ or $h_j^2 \to -h_j^2$. Now

$$\Psi = e^{\pm i S_0} \tilde{\chi}(\dot{\mathbf{r}}(f), \eta), \tag{128}$$

and the equation for $\tilde{\chi}$ is

$$\left\{ -\frac{1}{2} \frac{\partial^2}{\partial \eta_a^2} \mp i \left(\frac{\partial}{\partial f} - \Gamma_a^b \eta^a \frac{\partial}{\partial \eta_b} \right) + \frac{1}{2} W_{ab} \eta^a \eta^b + \tilde{H}_I \mp \frac{i}{2} \frac{\partial}{\partial f} (\ln f' \mathbf{r}_f^2) \right\} \tilde{\chi}(\mathbf{r}(f), \eta) = (E_1^{(0)} + \Delta) \tilde{\chi}(\mathbf{r}(f), \eta). \tag{129}$$

In approaching a minimum of the potential

$$W_{ab} \to M_{ab}, \quad \Gamma^b_a \to 0$$

and the variables f and η_a in (129) become separable with

$$\tilde{\chi}(\mathbf{r}(f),\eta) = \tilde{\Psi}(\mathbf{r}(f))\tilde{\phi}(\eta), \tag{130}$$

where $\tilde{\phi}$ is given by (97) and (116). Using (97) and (111) we can rewrite (129)

$$\mp i \left[\frac{\partial}{\partial f} + \frac{1}{2}\frac{\partial}{\partial f}(\ln f' \mathbf{r}_f^2)\right] \tilde{\psi}(\mathbf{r}(f))$$
$$= \epsilon^{(0)}\tilde{\Psi}(\mathbf{r}(f)) = \sum_j \frac{1}{4}(p_j - 1)\, h_j^2 \tilde{\psi}(\mathbf{r}(f)). \tag{131}$$

We now reexpress this differential equation in terms of x_j. Thus, using (102) we have

$$\mp i \frac{\partial}{\partial f} = \mp i \sum_j \dot{x}_j \frac{\partial}{\partial x_j} = \pm \frac{1}{2}\sum_j (x_j - x_{j0})\, h_j^2 \frac{\partial}{\partial x_j} \tag{132}$$

and

$$\frac{\partial}{\partial f}\ln(f'\mathbf{r}_f^2)$$
$$= \frac{(\partial/\partial f)\,\Sigma_j \dot{x}_j^2}{\Sigma_j \dot{x}_j^2} \simeq \frac{\Sigma_j\,(\partial/\partial f)(x_j - x_{j0})^2 h_j^4}{\Sigma_j (x_j - x_{j0})^2\, h_j^4}$$
$$\simeq i \frac{\Sigma_j\,(x_j - x_{j0})^2\, h_j^6}{\Sigma_j (x_j - x_{j0})^2\, h_j^4} \simeq i \sum_j h_j^2 + o\left(\frac{x_k - x_{k0}}{x_j - x_{j0}}\right)^2. \tag{133}$$

Equation (131) therefore becomes

$$\pm \sum_j \frac{1}{4} h_j^2 \left[2(x_j - x_{j0})\frac{\partial}{\partial x_j} \mp (p_j - 1) + 2\right]\tilde{\Psi} = 0, \tag{134}$$

and its solution is

$$\tilde{\Psi} = \prod_j (x_j - x_{j0})^{(1/2)(\pm(p_j - 1) - 2)}. \tag{135}$$

Using (103) we observe that for $p_j = 1$ (ground state) this factor corresponds precisely to the terms $(\mathbf{r}_f^2)^{1/2}[\det \xi^+]^{1/2}$ in (80). Setting as in (95) $\tilde{\Psi} = \phi(\mathbf{r}(f))\tilde{\phi}(\eta)$, we now have $\phi = e^{iS_0}\tilde{\Psi}$, and a linearly independent solution ϕ is obtained by the combined replacements $h_j^2 \to -h_j^2$ and $(p_j - 1) \to -(p_j - 1)$. From (131) together with (132) we see that these replacements are equivalent to a change of sign of f. This observation is important for the construction of solutions which are even or odd in f, as we shall see below.

We now define solutions ϕ_A, $\bar{\phi}_A$ by their behavior in the domain of overlap with ϕ_B, $\bar{\phi}_c$ but on the side excluding $\mathbf{r}_0$:

$$\phi_A(p_j - 1, h_j^2; x_j)$$
$$\simeq \exp\left(-\frac{1}{4}\sum_j h_j^2\,[(x_j - x_{j0})^2 - x_{j0}^2]\right)$$
$$\times \prod_j (x_j - x_{j0})^{(1/2)(p_j - 3)}, \tag{136}$$

$$\bar{\phi}_A(p_j - 1, h_j^2; x_j)$$
$$= \phi_A(-(p_j - 1), -h_j^2; x_j)$$
$$\simeq \exp\left(+\frac{1}{4}\sum_j h_j^2[(x_j - x_{j0})^2 - x_{j0}^2]\right)$$
$$\times \prod_j \frac{1}{(x_j - x_{j0})^{(1/2)(p_j + 1)}}. \tag{137}$$

X. MATCHING OF WAVE FUNCTIONALS

Looking at (124) and (136), we see that in their common domain of validity

$$\phi_A = \gamma \phi_B, \tag{138}$$

with

$$\gamma = \prod_j \left\{ \mathrm{Re}\, \frac{\Gamma[\frac{1}{4}(p_j + 1)]\, e^{(1/4) h_j^2 x_{j0}^2}\, e^{+i(\pi/4)(p_j - 3)}}{\alpha_{p_j}(\frac{1}{2} h_j^2)^{(1/4)(p_j - 3)}} \right\}, \tag{139}$$

and similarly

$$\bar{\phi}_A = \bar{\gamma}\,\bar{\phi}_c, \tag{140}$$

with

$$\bar{\gamma} = \prod_j \left\{ \frac{\Gamma[-\frac{1}{4}(p_j - 3)]\, e^{-(1/4) h_j^2 x_{j0}^2}}{\alpha_{p_j}} (\tfrac{1}{2} h_j^2)^{1/4(p_j + 1)} \right\}, \tag{141}$$

these being the dominant contributions. It should be observed that we arrive at these expressions by (1) expanding the solutions ϕ_B, $\bar{\phi}_c$, which are valid in a small domain of $o(1/h_a)$ around a minimum of the potential for relatively large values of $|x_j - x_{j0}|$ so that the expansion yields the behavior in the domain of the boundary, and (2) by expanding ϕ_A, $\bar{\phi}_A$ for relatively small values of $|x_j - x_{j0}|$ (i.e., selecting the most singular contribution) so that the expansion yields their behavior in the same domain, as indicated in Fig. 2.

XI. THE LEVEL SPLITTING

We now construct solutions $\phi_\pm$ which are, respectively, even or odd in the collective variable f. We have seen that the replacement $f \to -f$ is equivalent to the combined replacements of $h_j^2 \to -h_j^2$ and $(p_j - 1) \to -(p_j - 1)$ for all j. Hence we write

$$\phi_\pm = \phi_A \pm \bar{\phi}_A. \tag{142}$$

The continuation to the domain around a minimum of the potential at $\mathbf{r}_0$ is then given by

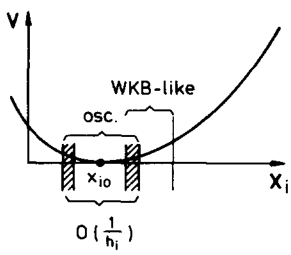


FIG. 2. Matching of oscillator and WKB solutions in overlap domains (shown hatched).

$$\phi_{\pm} = \gamma\phi_B \pm \bar{\gamma}\bar{\phi}_c. \tag{143}$$

At $\mathbf{r}_0$ these functions satisfy the following boundary conditions[14] characteristic of even and odd functions:

$$\phi_-(\mathbf{r}_0) = 0, \quad \phi_+(\mathbf{r}_0) = c_1,$$
$$\frac{d}{df}\phi_-(\mathbf{r}_0) = c_2, \quad \frac{d}{df}\phi_+(\mathbf{r}) = 0, \tag{144}$$

where c_1, c_2 are constants which determine the desired value of the Wronskian. Thus

$$\phi_{\pm}(\mathbf{r}_0) = (\gamma\phi_B \pm \bar{\gamma}\bar{\phi}_c)_{z_j=0}$$
$$= \prod_j \mathrm{Re}\, \frac{\Gamma[\frac{1}{4}(p_j+1)]\, e^{(1/4)h_j^2 x_{j0}^2} e^{i(\pi/4)(p_j-3)}}{(\frac{1}{2}h_j^2)^{(1/4)(p_j-3)}}$$
$$\pm \prod_j \Gamma[-\tfrac{1}{4}(p_j-3)]$$
$$\times e^{-(1/4)h_j^2 x_{j0}^2} (\tfrac{1}{2}h_j^2)^{(1/4)(p_j+1)}, \tag{145}$$

i.e.,

$$\prod_j \cos\frac{\pi}{4}(p_j-3)$$
$$\pm \prod_j \frac{\Gamma[-\frac{1}{4}(p_j-3)]}{\Gamma[\frac{1}{4}(p_j+1)]} (\tfrac{1}{2}h_j^2)^{(1/2)(p_j-1)} e^{-(1/2)h_j^2 x_{j0}^2}$$
$$= \phi_{\pm}(\mathbf{r}_0) \prod_j \frac{e^{-(1/4)h_j^2 x_{j0}^2}}{\Gamma[\frac{1}{4}(p_j+1)]} (\tfrac{1}{2}h_j^2)^{(1/4)(p_j-3)}.$$

For ϕ_- the right-hand side is zero; for ϕ_+ the right-hand side is of order $\Pi_j(1/h_j)$ lower than the left-hand side. Thus the right-hand side is zero or approximately zero. Using

$$\Gamma[\tfrac{1}{2}+z]\,\Gamma[\tfrac{1}{2}-z] = \pi/\cos\pi z,$$

we then have for $\phi_{\mp}$

$$\sin\frac{\pi}{2}(p_j-1) = \pm \frac{2\pi(\frac{1}{2}h_j^2)^{(1/2)(p_j-1)} e^{-(1/2)h_j^2 x_{j0}^2}}{\{\Gamma[\frac{1}{4}(p_j+1)]\}^2}. \tag{146}$$

Now $\sin(\pi/2)(p_j-1) = 0$ for $p_{j0} = 1,3,5,\ldots$. Expanding around these values we have

$$\sin(\pi/2)(p_j-1) \simeq (\pi/2)(p_j - p_{j0})(-1)^{(1/2)(p_{j0}-1)}. \tag{147}$$

Hence

$$p_j - p_{j0} \simeq \pm(-1)^{-(1/2)(p_{j0}-1)} \frac{4(\frac{1}{2}h_j^2)^{(1/2)(p_{j0}-1)} e^{-(1/2)h_j^2 x_{j0}^2}}{\{\Gamma[\frac{1}{4}(p_{j0}+1)]\}^2}, \tag{148}$$

the upper sign applying in the case of ϕ_-, the lower in the case of ϕ_+. It can be verified that the derivative conditions of (144) lead to the same relations to the same order of approximation.

We now return to (111), i.e.,

$$E(p_j) = V(\mathbf{r}_0) + \sum_a \frac{1}{4}p_a h_a^2 + \sum_j \frac{1}{4}h_j^2(p_j-1) + \Delta.$$

Expanding p_j around p_{j0} we have

$$E \simeq V(\mathbf{r}_0) + \sum_a \frac{1}{4}p_a h_a^2 + \sum_j (p_j - p_{j0})\frac{h_j^2}{4}$$
$$+ \sum_j \frac{h_j^2}{4}(p_{j_0} - 1). \tag{149}$$

Inserting (148) we obtain the splitting of the (asymptotically degenerate) energy levels associated with $\phi_{\mp}$ and thus the band structure of the eigenenergies. In particular we observe that for the ground state ($p_{j0} = 1$) the level splitting is

$$\Delta E = \frac{2}{\pi}\sum_j h_j^2 e^{-(1/2)h_j^2 x_{j0}^2}. \tag{150}$$

This type of expression is typical for the level splitting arising from degenerate minima. Of course, the existence of the splitting as such is also demonstrated by other higher-order contributions, such as the product of the exponentials[2] in (150).

It can also be observed that (149) does not involve a zero-point energy contribution in association with the quantum number related to the collective variable. This is, of course, what one would expect, because the entire procedure has been constructed in such a way that to lowest order the energy consists of the classical energy at the minimum ($V(\mathbf{r}) = 0$) plus the energy of quantum fluctuations perpendicular to the classical path. Clearly one could obtain another expression for E that would involve a zero-point energy associated with the collective variable by treating all degrees of freedom on the same quantum basis. In the above calculations it is precisely the centrifugal term in (114) which prevents this zero-point energy from occurring, and this centrifugal term, of course, results directly from the transformation to collective and fluctuation coordinates.

The band structure of the eigenenergies (149) is formally similar to that of the eigenvalues of Mathieu's equation, i.e., the Schrödinger equation for a periodic potential.[5,15] Put more precisely, (149) gives the boundaries of energy bands which are separated by forbidden regions that do not represent physically permissible eigenstates. In order to be able to obtain any desired eigenvalue in a given energy band it is necessary to introduce the equivalent of the Floquet parameter which is well known from the theory of simple differential equations with periodic differential operators.[15]

Finally we return to a point raised in connection with the ansatz (111) for the eigenenergy E. We have seen that (111) involves effectively two distinct sets of quantum numbers which are associated with different sets of basic eigenfunctions. The latter are, of course, related by a unitary transformation. The relationship between the quantum numbers $p_j = 2N_j + 1$ associated with the fixed frame of reference and the quantum numbers $p_a = 2n_a + 1$, $p_1 = 2n_1 + 1$ associated with the local reference frame is complicated but can be investigated with the help of the Bohr–Sommerfeld quantization conditions.

We do not calculate the tunneling amplitude in the present context. This problem has been discussed in Ref. 16. The application of the method of Secs. II–VI to a field theory with solitons has been investigated in Ref. 17.

APPENDIX: ALTERNATIVE DERIVATION OF EQ. (45)

Here we present an alternative derivation of the equation of small fluctuations. We have $\mathbf{R} = \mathbf{r} + \mathbf{v}$, $\mathbf{r} = \mathbf{r}\,(f(q))$, and $\dot{\mathbf{R}} = \dot{\mathbf{r}} + \dot{\mathbf{v}}$, a dot denoting differentiation with respect to $f(q)$. The total energy E is therefore

$$\begin{aligned} E &= \tfrac{1}{2}\,\dot{\mathbf{R}}^2 + V(\mathbf{R}) = \tfrac{1}{2}\,(\dot{\mathbf{r}} + \dot{\mathbf{v}})^2 + V(\mathbf{r} + \mathbf{v}) \\ &= \tfrac{1}{2}\dot{\mathbf{r}}^2 + \tfrac{1}{2}\dot{\mathbf{v}}^2 + \dot{\mathbf{r}}\cdot\dot{\mathbf{v}} + V(\mathbf{r}) \\ &\quad + \mathbf{v}\cdot(\nabla V)_{\mathbf{r}} + \frac{1}{2}\left(\frac{\partial^2 V}{\partial R_i\,\partial R_j}\right)_{\mathbf{r}} v_i v_j + \cdots . \end{aligned} \tag{A1}$$

Now $\mathbf{r}$ is a solution of

$$\ddot{\mathbf{r}} = -\nabla V(\mathbf{r}), \tag{A2}$$

so that

$$\dot{\mathbf{r}}\cdot\dot{\mathbf{v}} + \mathbf{v}(\nabla V)_{\mathbf{r}} = \dot{\mathbf{r}}\cdot\dot{\mathbf{v}} - \mathbf{v}\cdot\ddot{\mathbf{r}}. \tag{A3}$$

We choose $\mathbf{v}$ such that $\dot{\mathbf{r}}\cdot\dot{\mathbf{v}} - \mathbf{v}\cdot\ddot{\mathbf{r}} = 0$.

Alternatively we may consider $\dot{\mathbf{r}}$ as a solution of

$$\dddot{\mathbf{r}} = -\frac{d}{dt}(\nabla V(\mathbf{r})) = -\left(\frac{d^2 V}{d\mathbf{R}\,d\mathbf{R}}\right)_{\mathbf{r}}\dot{\mathbf{r}} \tag{A4}$$

obtained from (A2). Differentiating (A1) with respect to v_i we obtain

$$0 = \frac{d}{dv_i}\left(\frac{1}{2}\,\dot{\mathbf{v}}^2\right) + \left(\frac{\partial^2 V}{\partial R_i\,\partial R_j}\right)_{\mathbf{r}} v_j + \cdots , \tag{A5}$$

i.e.,

$$\ddot{v}_i = -\left(\frac{\partial^2 V}{\partial R_i\,\partial R_j}\right)_{\mathbf{r}} v_j .$$

In the following we ignore nonlinear terms. Equation (A5) is the equation of small oscillations in terms of components with respect to the fixed reference frame. Comparing (A4) and (A5) we see that $(\dot{\mathbf{r}}_i)$ is a solution of (A5).

We now set

$$\mathbf{v} = \frac{\mathbf{v}\cdot\dot{\mathbf{r}}}{\dot{\mathbf{r}}^2}\,\dot{\mathbf{r}} + \sum_{a=2}^{N}\eta_a\mathbf{n}_a . \tag{A6}$$

Then

$$\begin{aligned} (\mathbf{n}_b)_i\ddot{v}_i &= (\mathbf{n}_b)_i\,\frac{d^2}{dt^2}\left[\frac{\mathbf{v}\cdot\dot{\mathbf{r}}}{\dot{\mathbf{r}}^2}\,(\dot{\mathbf{r}})_i\right] \\ &\quad + \ddot{\eta}_b + 2\Gamma^{\,b}_{\,a}\,\dot{\eta}_a + (\mathbf{n}_b\cdot\ddot{\mathbf{n}}_a)\eta_a , \end{aligned} \tag{A7}$$

where

$$\Gamma^{\,b}_{\,a} = \dot{\mathbf{n}}_a\cdot\mathbf{n}_b .$$

Then

$$\begin{aligned} &(\mathbf{n}_b)_i\,\frac{d^2}{dt^2}\left[\frac{\mathbf{v}\cdot\dot{\mathbf{r}}}{\dot{\mathbf{r}}^2}\,(\dot{\mathbf{r}})_i\right] \\ &= \frac{\partial^2}{\partial t^2}\left(\frac{\mathbf{v}\cdot\dot{\mathbf{r}}}{\dot{\mathbf{r}}^2}\right)\dot{\mathbf{r}}\cdot\mathbf{n}_b + 2\,\frac{\partial}{\partial t}\left(\frac{\mathbf{v}\cdot\dot{\mathbf{r}}}{\dot{\mathbf{r}}^2}\right)\ddot{\mathbf{r}}\cdot\mathbf{n}_b \\ &\quad + \left(\frac{\mathbf{v}\cdot\dot{\mathbf{r}}}{\dot{\mathbf{r}}^2}\right)\dddot{\mathbf{r}}\cdot\mathbf{n}_b , \end{aligned} \tag{A8}$$

where $\dot{\mathbf{r}}\cdot\mathbf{n}_b = 0$. Also

$$\begin{aligned} \frac{\partial}{\partial t}\left(\frac{\mathbf{v}\cdot\dot{\mathbf{r}}}{\dot{\mathbf{r}}^2}\right) &= \frac{1}{\dot{\mathbf{r}}^2}\,(\mathbf{v}\cdot\ddot{\mathbf{r}} + \dot{\mathbf{v}}\cdot\dot{\mathbf{r}}) - 2\,\frac{\dot{\mathbf{r}}\cdot\ddot{\mathbf{r}}}{(\dot{\mathbf{r}}^2)^2}\,(\mathbf{v}\cdot\dot{\mathbf{r}}) \\ &= 2\,\frac{\mathbf{v}\cdot\ddot{\mathbf{r}}}{\dot{\mathbf{r}}^2} - 2\,\frac{\dot{\mathbf{r}}\cdot\ddot{\mathbf{r}}}{(\dot{\mathbf{r}}^2)^2}\,(\mathbf{v}\cdot\dot{\mathbf{r}}) \end{aligned} \tag{A9}$$

on using (A3). Next, inserting the completeness relation (12), we obtain

$$\begin{aligned} \mathbf{v}\cdot\ddot{\mathbf{r}} &= \frac{\mathbf{v}\cdot\dot{\mathbf{r}}}{\dot{\mathbf{r}}^2}\,(\dot{\mathbf{r}}\cdot\ddot{\mathbf{r}}) + (\mathbf{v}\cdot\mathbf{n}_a)(\mathbf{n}_a\cdot\ddot{\mathbf{r}}) \\ &= \frac{\mathbf{v}\cdot\dot{\mathbf{r}}}{\dot{\mathbf{r}}^2}\,(\dot{\mathbf{r}}\cdot\ddot{\mathbf{r}}) + \eta_a(\mathbf{n}_a\cdot\ddot{\mathbf{r}}). \end{aligned} \tag{A10}$$

Hence

$$\begin{aligned} &2\,\frac{\partial}{\partial t}\left(\frac{\mathbf{v}\cdot\dot{\mathbf{r}}}{\dot{\mathbf{r}}^2}\right)\ddot{\mathbf{r}}\cdot\mathbf{n}_b \\ &= \frac{4}{\dot{\mathbf{r}}^2}\,(\mathbf{v}\cdot\ddot{\mathbf{r}})(\ddot{\mathbf{r}}\cdot\mathbf{n}_b) - 4\,\frac{\ddot{\mathbf{r}}\cdot\dot{\mathbf{r}}}{(\dot{\mathbf{r}}^2)^2}\,(\mathbf{v}\cdot\dot{\mathbf{r}})(\ddot{\mathbf{r}}\cdot\mathbf{n}_b) \\ &= \frac{4}{(\dot{\mathbf{r}}^2)^2}\,(\mathbf{v}\cdot\dot{\mathbf{r}})(\dot{\mathbf{r}}\cdot\ddot{\mathbf{r}})(\ddot{\mathbf{r}}\cdot\mathbf{n}_b) + 4/\dot{\mathbf{r}}^2\,(\ddot{\mathbf{r}}\cdot\mathbf{n}_b)(\ddot{\mathbf{r}}\cdot\mathbf{n}_a)\eta_a \\ &\quad - \frac{4(\ddot{\mathbf{r}}\cdot\dot{\mathbf{r}})}{(\dot{\mathbf{r}}^2)^2}\,(\mathbf{v}\cdot\dot{\mathbf{r}})(\ddot{\mathbf{r}}\cdot\mathbf{n}_b) = \frac{4}{\dot{\mathbf{r}}^2}\,(\ddot{\mathbf{r}}\cdot\mathbf{n}_b)(\ddot{\mathbf{r}}\cdot\mathbf{n}_a)\eta_a , \end{aligned} \tag{A11}$$

where we have again used the completeness relation. Substituting (A11) into (A8) we obtain

$$(\mathbf{n}_b)_i\,\frac{d^2}{dt^2}\left[\frac{\mathbf{v}\cdot\dot{\mathbf{r}}}{\dot{\mathbf{r}}^2}\,(\dot{\mathbf{r}})_i\right] = \frac{4}{\dot{\mathbf{r}}^2}\,(\ddot{\mathbf{r}}\cdot\mathbf{n}_b)(\ddot{\mathbf{r}}\cdot\mathbf{n}_a)\eta_a + \left(\frac{\mathbf{v}\cdot\dot{\mathbf{r}}}{\dot{\mathbf{r}}^2}\right)\dddot{\mathbf{r}}\cdot\mathbf{n}_b .$$

On using (A4) this becomes

$$\frac{4}{\dot{\mathbf{r}}^2}\,(\ddot{\mathbf{r}}\cdot\mathbf{n}_b)(\ddot{\mathbf{r}}\cdot\mathbf{n}_a)\eta_a - \left(\frac{\mathbf{v}\cdot\dot{\mathbf{r}}}{\dot{\mathbf{r}}^2}\right)(\mathbf{n}_b)_i\left(\frac{\partial^2 V}{\partial R_i\,\partial R_j}\right)_{\mathbf{r}}(\dot{\mathbf{r}})_j ,$$

and on using (A6) this is

$$\frac{4}{\dot{\mathbf{r}}^2}\,(\ddot{\mathbf{r}}\cdot\mathbf{n}_b)(\ddot{\mathbf{r}}\cdot\mathbf{n}_a)\eta_a - (\mathbf{n}_b)_i\left(\frac{\partial^2 V}{\partial R_i\,\partial R_j}\right)_{\mathbf{r}}\left[v_j - \sum_{a=2}^{N}\eta_a(\mathbf{n}_a)_j\right].$$

From (A5) we obtain, by multiplying the right-hand side by $(\mathbf{n}_b)_i$,

$$\begin{aligned} &-(\mathbf{n}_b)_i\left(\frac{\partial^2 V}{\partial R_i\,\partial R_j}\right)_{\mathbf{r}} v_j \\ &= (\mathbf{n}_b)_i\ddot{v}_i = (4/\dot{\mathbf{r}}^2)(\ddot{\mathbf{r}}\cdot\mathbf{n}_b)(\ddot{\mathbf{r}}\cdot\mathbf{n}_a)\eta_a \\ &\quad - (\mathbf{n}_b)_i\left(\frac{\partial^2 V}{\partial R_i\,\partial R_j}\right)_{\mathbf{r}}\left[v_j - \sum_{a=2}^{N}\eta_a(\mathbf{n}_a)_j\right] \\ &\quad + \ddot{\eta}_b + 2\,\Gamma^{\,b}_{\,a}\dot{\eta}_a + (\mathbf{n}_b\cdot\ddot{\mathbf{n}}_a)\eta_a . \end{aligned}$$

Thus

$$\begin{aligned} &\ddot{\eta}_b + 2\Gamma^{\,b}_{\,a}\dot{\eta}_a + (\mathbf{n}_b\cdot\ddot{\mathbf{n}}_a)\eta_a + \frac{4}{\dot{\mathbf{r}}^2}\,(\ddot{\mathbf{r}}\cdot\mathbf{n}_b)(\ddot{\mathbf{r}}\cdot\mathbf{n}_a)\eta_a \\ &\quad + (\mathbf{n}_b)_i\left(\frac{\partial^2 V}{\partial R_i\,\partial R_j}\right)_{\mathbf{r}}\sum_{a=2}^{N}\eta_a(\mathbf{n}_a)_j = 0. \end{aligned} \tag{A12}$$

On the other hand, if

$$D_{ab} = \delta_{ab}\,\frac{\partial}{\partial t} - \Gamma^{\,b}_{\,a}, \tag{A13}$$

we have

$$\begin{aligned} D_{ab}D_{bc}\eta_c &= \ddot{\eta}_a - \left(\frac{\partial}{\partial t}\,\Gamma^{\,c}_{\,a}\right)\eta_c \\ &\quad - \Gamma^{\,c}_{\,a}\,\frac{\partial\eta_c}{\partial t} + \Gamma^{\,b}_{\,a}\Gamma^{\,c}_{\,b}\eta_c \\ &= \ddot{\eta}_a - 2\Gamma^{\,c}_{\,a}\dot{\eta}_c - (\dot{\mathbf{n}}_c\cdot\dot{\mathbf{n}}_a + \mathbf{n}_c\cdot\ddot{\mathbf{n}}_a)\eta_c \\ &\quad + (\mathbf{n}_b\cdot\dot{\mathbf{n}}_a)(\mathbf{n}_c\cdot\dot{\mathbf{n}}_b)\eta_c . \end{aligned} \tag{A14}$$

Using $\mathbf{n}_c \cdot \mathbf{n}_b = \delta_{cb}$ and $\mathbf{n}_c \cdot \dot{\mathbf{r}} = 0$, i.e., $\dot{\mathbf{n}}_c \cdot \mathbf{n}_b + \mathbf{n}_c \cdot \dot{\mathbf{n}}_b = 0$ and $\dot{\mathbf{n}}_c \cdot \mathbf{r} + \mathbf{n}_c \cdot \mathbf{r} = 0$, we have (again using the completeness relation)

$$\dot{\mathbf{n}}_c \cdot \dot{\mathbf{n}}_a = \frac{(\dot{\mathbf{n}}_c \cdot \dot{\mathbf{r}})(\dot{\mathbf{r}} \cdot \dot{\mathbf{n}}_a)}{\dot{\mathbf{r}}^2} + (\dot{\mathbf{n}}_c \cdot \mathbf{n}_b)(\mathbf{n}_b \cdot \dot{\mathbf{n}}_a)$$
$$= \frac{(\mathbf{n}_c \cdot \ddot{\mathbf{r}})(\mathbf{n}_a \cdot \ddot{\mathbf{r}})}{\dot{\mathbf{r}}^2} - (\mathbf{n}_c \cdot \dot{\mathbf{n}}_b) \cdot (\mathbf{n}_b \cdot \dot{\mathbf{n}}_a),$$

i.e.,

$$(\mathbf{n}_c \cdot \dot{\mathbf{n}}_b)(\mathbf{n}_b \cdot \dot{\mathbf{n}}_a) = \frac{(\mathbf{n}_a \cdot \ddot{\mathbf{r}})(\mathbf{n}_c \cdot \ddot{\mathbf{r}})}{\dot{\mathbf{r}}^2} - (\dot{\mathbf{n}}_c \cdot \dot{\mathbf{n}}_a).$$

Substituting in (A14) and using

$$0 = \frac{d^2}{dt^2}(\mathbf{n}_a \cdot \mathbf{n}_c) = \ddot{\mathbf{n}}_a \cdot \mathbf{n}_c + 2\dot{\mathbf{n}}_a \cdot \dot{\mathbf{n}}_c + \mathbf{n}_a \cdot \ddot{\mathbf{n}}_c,$$

we obtain

$$D_{ab} D_{bc} \eta_c$$
$$= \ddot{\eta}_a - 2\Gamma^c_a \dot{\eta}_c - (2\dot{\mathbf{n}}_c \cdot \dot{\mathbf{n}}_a + \mathbf{n}_c \cdot \ddot{\mathbf{n}}_a)\eta_c + \frac{(\mathbf{n}_c \cdot \ddot{\mathbf{r}})(\mathbf{n}_a \cdot \ddot{\mathbf{r}})\eta_c}{\dot{\mathbf{r}}^2}$$
$$= \ddot{\eta}_a + 2\Gamma^a_c \dot{\eta}_c + (\mathbf{n}_a \cdot \ddot{\mathbf{n}}_c)\eta_c + \frac{(\mathbf{n}_c \cdot \ddot{\mathbf{r}})(\mathbf{n}_a \cdot \ddot{\mathbf{r}})}{\dot{\mathbf{r}}^2}. \quad \text{(A15)}$$

Inserting this in (A12) we have

$$D_{ab} D_{bc} \eta_c + (3/\dot{\mathbf{r}}^2)(\ddot{\mathbf{r}} \cdot \mathbf{n}_a)(\ddot{\mathbf{r}} \cdot \mathbf{n}_b)\eta_b$$
$$+ (\mathbf{n}_a)_i \left(\frac{\partial^2 V}{\partial R_i \, \partial R_j}\right)_{\mathbf{r}} (\mathbf{n}_b)_j \eta_b = 0,$$

i.e.,

$$D_{ab} D_{bc} \eta_c + W_{ab} \eta_b = 0, \quad \text{(A16)}$$

where

$$W_{ab} = V^{(2)}_{ab} + (3/\dot{\mathbf{r}}^2)(\ddot{\mathbf{r}} \cdot \mathbf{n}_a) - (\ddot{\mathbf{r}} \cdot \mathbf{n}_b).$$

This is therefore the equation of small fluctuations with respect to the moving frame of reference.

[1] J-L. Gervais and B. Sakita, Phys. Rev. D **16**, 3507 (1977).

[2] The existence of the splitting of the ground state has been demonstrated by H. J. de Vega, J-L. Gervais, and B. Sakita, Nucl. Phys. B **139**, 20 (1978).

[3] H. M. M. Mansour and H. J. W. Müller-Kirsten, J. Math Phys. **23**, 1835 (1982).

[4] H. J. W. Müller-Kirsten, Phys. Rev. D **25**, 1552 (1982).

[5] R. B. Dingle and H. J. W. Müller, J. Reine Angew. Math. **211**, 11 (1962).

[6] See references cited in Ref. 4.

[7] H. J. W. Müller-Kirsten and A. Wiedemann, Nuovo Cimento A **78**, 61 (1983).

[8] V. N. Gribov, Lecture at the 12th Winter School of the Leningrad Nucl. Phys. Institute, 1977; SLAC-translation 176; V. N. Gribov, Nucl. Phys. B **139**, 1 (1978). Very readable discussions have also been given by H. Pagels, *New Frontiers in High-Energy Physics*, edited by A. Perlmutter and L. Scott (Plenum, New York, 1978), p. 559; and W. I. Weisberger, "Gribov Ambiguities and Quantization in the Axial Gauge," in *Asymptotic Realms of Physics, Essays in Honor of Francis E. Low*, edited by A. H. Guth, K. Huang, and R. L. Jaffe (MIT, Cambridge, MA, 1983), p. 162..

[9] H. J. W. Müller-Kirsten and A. Wiedemann, Phys. Lett. B **132**, 169 (1983).

[10] R. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D **10**, 4114, 4130, 4138 (1974); **11**, 3424 (1975).

[11] R. Rajaraman, *An Introduction to Solitons and Instantons in Quantum Field Theory* (North-Holland, Amsterdam, 1982).

[12] S. K. Bose, G. E. Hite, and H. J. W. Müller-Kirsten, J. Math. Phys. **20**, 1878 (1979); S. K. Bose and H. J. W. Müller-Kirsten, J. Math. Phys. **20**, 2471 (1979).

[13] See, e.g., A. Messiah, *Mécanique Quantique* (Dunod, Paris, 1969), Vol. I, Sec. 12.3.3.

[14] These boundary conditions have, of course, been chosen such that the Wronskian is unequal to zero and the solutions are independent. See, e.g., E. M. Harrell, Ann. Phys. (NY) **119**, 351 (1979), or any book on Mathieu or similar equations.

[15] See, e.g., J. Meixner and F. W. Schäfke, *Mathieusche Funktionen und Sphäroidfunktionen* (Springer, Berlin, 1954); F. M. Arscott, *Periodic Differential Equations* (Pergamon, Oxford, 1964).

[16] H. J. W. Müller-Kirsten, Nuovo Cimento A **63**, 1 (1981).

[17] L. Maharana and H. J. W. Müller-Kirsten, Nuovo Cimento A **83**, 229 (1984).

On the relation between gauge theories, moving frames, fiber bundles, and parastatistics

Kenneth I. Macrae
Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

(Received 27 August 1984; accepted for publication 4 January 1985)

We write down standard gauge field theory in a basis-independent manner using the ideas of moving frames (fiber bundles). Then we describe the construction of frames for gauging parafields. To do this (frames) bases for fields are replaced by Clifford matrices. These matrices are in one-to-one correspondence with the number of spinor components. We briefly examine the objects upon which they can act (through matrix multiplication on the left). These objects bear the same relations to spinors that spinors do to vectors. Finally, we show how to construct a set of inner products for the parabases that yield the same action and n-point functions as in the standard field theory.

I. INTRODUCTION

There is a quote, attributed to Einstein, which is, "A thing should be made as simple as possible and no simpler." Bearing this in mind, we will show that the physical consequences of gauging parastatistics are the same as those of conventional gauge theories.

To set the stage we will first describe a choice-of-basis independent framework for conventional gauge theories.[1] Next, we briefly review the concept of a basis for parafields.[2] Then we convert the general basis formulation of the standard theory to one natural to the parastatistical formulation and see how every quantity associated with a field theory goes through.

II. FRAMEWORK FOR GAUGE THEORIES

We will begin by considering an n-component Fermi field ψ. It has components ψ^a and basis s_a. Under an infinitesimal translation, both the field components ψ^a and the basis elements s_a can contribute to the change in $\psi = s_a \psi^a$. Thus we find[1]

$$\begin{aligned}\delta\psi &= \delta(s_a\psi^a) = s_a\,\delta\psi^a + \delta s_a\,\psi^a \\ &= s_a\,\partial_\mu\psi^a\,\delta x^\mu + s_a A^{\ a}_{\mu b}\psi^b\,\delta x^\mu \\ &= s_a(\partial_\mu\psi^a + A^{\ a}_{\mu b}\psi^b)\delta x^\mu = s_a(D_\mu\psi)^a\,\delta x^\mu.\end{aligned} \tag{1}$$

The change in the basis element s_a gives rise to the gauge field $A^{\ a}_{\mu b}$. The last object in parentheses $(D_\mu\psi)^a$ is the covariant derivative of the field.

A note on language. In view of the equality $v = v^k\hat{x}_k$, familiar from classical mechanics, showing that an object can be expressed as a tensor in an arbitrary basis (the directions and magnitudes of the basis vectors $\hat{x}_k$ are arbitrary) is the same as showing that it exists independent of the basis. We will use the analog of this vector notation for field theories. We will therefore refer to our equations as being basis independent when we show that they do not depend upon the choice of basis.

The formalism used in relation (1) was originally devised by Cartan.[3] He called s_a the moving frame (*répère mobile*). The idea of the moving frame became the foundation of the modern concept of the fiber bundle. For our purposes a fiber bundle B can be thought of as a union of copies of a given space F, the fiber, where there is a *unique* copy of F for each point x in the base manifold M. To denote this association of a fiber per point we write a given copy as $\{x\}\times F$, where $\{x\}$ denotes the set consisting of the point x. The bundle can be expressed via the following formula:

$$B = \bigcup_{x\in M}\{x\}\times F. \tag{2}$$

An aside: If one wanted to be more precise, one would replace the singleton $\{x\}$'s in Eq. (2) with a collection of open sets U_σ in M, which form an open cover of M and which have the property that the bundle is locally of product form $U_\sigma\times F$. In this case the bundle is the union of the $U_\sigma\times F$'s. The map which locally reduces the bundle of this product form is called a local trivialization. If any of the sets U_σ can be the whole manifold M then the bundle is said to be trivial.

Our notation was chosen so that one could replace the manifold M with a lattice L consisting of vertices x and links connecting them (the lattice has the structure it inherits as a substructure of R^n). The bundle B is then replaced with the obvious lattice analog

$$B = \bigcup_{x\in L}\{x\}\times F. \tag{3}$$

We will refer to these objects as lattice bundles or L bundles. This structure is especially relevant to a field theory being viewed as defined by the limit of multi-integrals on fibers over L as L goes to the limiting manifold M.

If s_a is a basis for a complex N-dimensional space C^n, there are n independent s_a's. Thus, to construct a fiber large enough to include each, one is led to $C^{n\times n}$ $(C^n \otimes \cdots \otimes C^n)$. Local coordinates for the bundle with this fiber are $(x,s_1,\ldots,s_N)$. Each $s_a \in C^n$. The real case R^n works the same way.

Thus, over a lattice L the L bundle of moving frames has the following construction:

$$B = \bigcup_{x\in L}\{x\}\times C^{n\times n}. \tag{4}$$

Its limit is the bundle of frames on M

$$B = \bigcup_{x\in M}\{x\}\times C^{n\times n}. \tag{5}$$

 0022-2488/85/071695-06$02.50

We will see that the Yang–Mills tensor is the object one finds when the frame is transported around a loop in any of these bundles.

To include scalar (or spinor) fields we will introduce a field ψ. To accomodate this new field we must give it a fiber of its own. But, because we want to couple it to the gauge field there must be a relationship between the new fiber and the old. Thus we view ψ as a multilinear mapping from the fiber $C^{n\times n}$ into an added fiber C^n

$$\psi(s_1,...,s_n) = S_1\psi^1 + ... + S_n\psi^n. \tag{6}$$

The coefficients depend upon which fiber is being mapped, but we have suppressed this x dependence. Here S_a is the image in C^n of s_a under the natural identification of the ath piece of $C^{n\times n}$ with C^n. To see that this identification is natural view the collection of n s_a's as forming a matrix which maps the canonical basis elements of C^n, that is, the vectors u_j with one in the jth place and zeros elsewhere, into the basis s_a. The components of this matrix are the inner products of s_a and u_j using the canonical metric for C^n (the identity metric on the canonical basis). Given this, components of S_a are equal to the components of s_a. In fact this identification is so natural that we will not persist in making the distinction; thus, we will write s_a instead of S_a. The coefficients ψ^a are complex numbers, if ψ is a scalar. But they become (complex) Grassmann numbers, if ψ is a spinor. In this case the image space is G_c^n, the space of complex n-component Grassmann numbers. The identification of S_a with s_a can and will still be made. The coefficients are characteristically viewed as being the field; we view the total entity ψ as being the field.

The bundle of interest in the spinor case is apparently

$$B = \bigcup_{x\in M} \{x\} \times C^{n\times n} \times G_c^n. \tag{7}$$

Here, B is equipped with ψ's. *But*, there is an important, technical point. One must wonder about the anticommunication relations of fermions at different locations. Should not all fermions be placed into a single enormous Grassmann algebra so that they will all anticommute with each other? The obvious answer to this question is yes. Fermions are not local objects. However, it is possible to achieve this end indirectly, since we are interested in functional integrals. Use the tensor product for fermions at different locations, but then project out of that the antisymmetric piece by totally antisymmetrizing the functional measure. In this way the time-ordered expectations of the fields will have the right relations and that is what matters.

This begs the question of what is the correct space. The answer is relatively straightforward in the case where the base is a lattice. In the case of a base manifold, we will view it as a limit of the base lattice case. So, if there are a total of p points in the lattice, the space of interest is an L bundle taken in product with a complex Grassmann space of dimension $n\times p$. Thus the total space is

$$S = \bigcup_{x\in L} \{x\} \times C^{n\times n} \times G_c^{n\times p}. \tag{8}$$

There is a spinor-valued map from a single fiber in the lattice bundle piece of this space ($C^{n\times n}$) into a particular (complex) n-dimensional piece of G_c^{n*p} which looks just like the map in Eq. (6).

We will now demonstrate the procedure for constructing a sample action on the manifold limit of this space. In Eq. (1) we used the equation for the change in the frame under a small displacement

$$\delta s_a = s_b A^{\,b}_{\mu a}\, \delta x^\mu. \tag{9}$$

The second time we make a displacement we pick up the change in the coefficients $A^{\,b}_{\mu a}$ as well as another change of frame. Thus in second order we find

$$\begin{aligned} d\delta s_a &= s_b\, \partial_\nu A^{\,b}_{\mu a}\, dx^\nu\, \delta x^\mu + s_b A^{\,b}_{\nu c} A^{\,c}_{\mu a}\, dx^\nu\, \delta x^\mu \\ &= [s_b(\partial_\nu A^{\,b}_{\mu a} + A^{\,b}_{\nu c} A^{\,c}_{\mu a})]dx^\nu\, \delta x^\mu. \end{aligned} \tag{10}$$

The antisymmetric part of the term in parentheses in Eq. (10) is the curvature of the frame bundle. This is also called the Yang–Mills tensor. It is what you get when you compare displacements taken in two different orders or when you take a trip around an infinitesimal loop. We can implicitly define the components of this tensor in the following way:

$$s_b F^{\,b}_{\nu\mu a} = \tfrac{1}{2}[s_b(\partial_\nu A^{\,b}_{\mu a} + A^{\,b}_{\nu c} A^{\,c}_{\mu a}) - (\nu\leftrightarrow\mu)]. \tag{11}$$

Note that the metric is a rank two tensor in the basis space. We can introduce another basis element to absorb the free index a. To achieve a formalism which works in a general basis we must do that. Thus we have a curvature tensor which lives in a subbundle of the bundle of frames. The fiber of the subbundle is just $C^n \otimes C^n$. The general basis expression for F is

$$F_{\nu\mu} = s^{*a} \otimes s_b F^{\,b}_{a\nu\mu}. \tag{12}$$

The asterisk denotes complex conjugation. Clearly, the gauge field also takes values in that fiber. In general an object with p indices will live in the subbundle whose fiber is $C^{n\times p}(= C^n \otimes \cdots \otimes C^n$, p-fold product.

It may seem inconsistent to introduce a basis-independent notation for the internal space and leave the space-time indices dangling. In fact, there is a basis-independent construction for space-time also. As we saw, the antisymmetrized tensor product of "infinitesimal" displacements was the object which arose in computing the curvature. Infinitesimal displacements are given mathematical realization as "differentials," the duals of the directional derivatives. Antisymmetric tensor products of these differentials are endemic to the geometry. The ensemble of them form the spaces of "differential forms" for the bundle, a p form being the antisymmetric product of p differentials.

The action has a natural expression in this language. It is the Hilbert square of the curvature two-form. However, it is more common in physics to keep the space-time indices, and to construct the Lagrangian density as a scalar rather than as a form. We will follow that convention since the use of forms is not required for this problem.

In order to construct the action as a scalar in our basis-independent notation we need a metric for the basis elements. We give the basis elements s_b the usual (flat) metric of C^n. We will express the metric of that space as follows:

$$(s_a | s_b) = \delta_{ab}. \tag{13}$$

This product has the usual sesquilinear structure. That is, the product of sums of entries, in either arguments, is the sum of the products. Further, scalars in the second argument can simply be factored out front, $(s_a|s_b\lambda) = \lambda(s_a|s_b)$. On the other hand, those in the first argument are complex conjugated before they can be extracted; $(s_a\lambda|s_b) = \lambda^*(s_a|s_b)$.

In order to construct a scalar such as the Lagrangian from a rank two tensor in the internal space (the Yang–Mills tensor), we must extend the metric to pairs of basis elements. Because the bundles of interest have a tensor product structure, the generalization to arbitrary products is easy:

$$\begin{aligned}(s_{a1}\otimes\cdots\otimes s_{an}|s_{b1}\otimes\cdots\otimes s_{bn}) &= (s_{a1}|s_{b1})\dots(s_{an}|s_{bn})\\ &= \delta_{a1'b1}\dots\delta_{an'bn}.\end{aligned} \tag{14}$$

The terms on the right are simply multiplied together. The gauge field contribution to the Lagrangian is

$$\begin{aligned}L &= \tfrac{1}{4}(F_{\mu\alpha}|F_{\nu\beta})g^{\mu\nu}g^{\alpha\beta}\\ &= \tfrac{1}{4}F^{*ab}_{\mu\alpha}F^{cd}_{\nu\beta}(s_a\otimes s_b|s_c\otimes s_d)g^{\mu\nu}g^{\alpha\beta}\\ &= \tfrac{1}{4}F^{a}_{\mu\alpha b}F^{b}_{\nu\beta a}g^{\mu\nu}g^{\alpha\beta}.\end{aligned} \tag{15}$$

The Hermiticity of F has been used to get the last line.

To include the contribution from a scalar, we construct the covariant derivative by recalling Eq. (1) and omitting the infinitesimal displacement

$$D_\mu\phi = s_a(D_\mu\phi)^a. \tag{16}$$

We use the inner product to make the scalar kinetic Lagrangian

$$\begin{aligned}L &= \tfrac{1}{2}(D_\mu\phi|D_\nu\phi)g^{\mu\nu}\\ &= \tfrac{1}{2}(\mathrm{D}_\mu\phi)^{*a}(\mathrm{D}_\nu\phi)^b(s_a|s_b)g^{\mu\nu}\\ &= \tfrac{1}{2}(D_\mu\phi)^{*a}(D_\nu\phi)^b\delta_{ab}g^{\mu\nu}\\ &= \tfrac{1}{2}(D_\mu\phi)^\dagger(D_\nu\phi)g^{\mu\nu}.\end{aligned} \tag{17}$$

Constructing potential terms for the action in a basis-independent way is a straightfoward continuation of the ideas we have just presented.

The fermion action is linear in the covariant derivative. To form it we introduce the following object:

$$\Gamma_\mu\psi = s_a(\gamma_\mu\psi)^a. \tag{18}$$

The fermion kinetic term is an inner product of this object with the basis-independent covariant derivative. However, the inner product appropriate to the spinors for Minkowski space equals the components of the matrix γ_0, $(s_a|s_b) = \gamma_{0ab}$. Thus,

$$\begin{aligned}L &= (\Gamma_\mu\psi|D_\nu\psi)g^{\mu\nu} \qquad (19)\\ &= (\gamma_\mu\psi)^{*a}(D_\nu\psi)^b(s_a|s_b)g^{\mu\nu}\\ &= (\gamma_\mu\psi)^{*a}g^{\mu\nu}(D_\nu\psi)^b\gamma_{0ab}.\\ &= \bar\psi\bar{\mathrm{D}}\psi.\end{aligned}$$

If this expression is not self-conjugate, then add the conjugate and divide the sum by 2.

It should now be clear that the action is independent of the choice of basis. We also must show that the relevant n-point functions do not depend upon that choice either. For example, consider a simple case in which the spinors have only one index, the spin index, which will be local (gauged). While it is true that this yields a quantum gravity formalism, we are using this as an example solely to avoid the extra complexity of giving each spinor multiple indices. A more elaborate example with pairs of indices per spinor (or basis) is dealt with later in this paper. It should, however, be easy for the reader to go from our example to the other case (especially if one remembers the Kaluza–Klein approach through which that other case can be subsumed). The two-point function for two fermions is

$$\langle(\psi(x)|\psi(y))\rangle \equiv \langle(s(x)_a\psi(x)^a|s(y)_b\psi(y)^b)\rangle. \tag{20}$$

We have made the coordinates explicit to show that one must relate bases at different points in order to compute these functions. So far we have given the inner product for frames at the same point. In a bundle with nontrivial curvature or Yang–Mills tensor this means giving the path that connects the two points. This formalism *requires* the flux link. Let $U(x,y)$ denote the path-dependent matrix obtained by integrating the gauge field along the path connecting x with y. We can define the inner product between spinor bases at two locations in terms of spinor bases at a given location by "transporting" one of the bases. Thus we have

$$(s(x)_a|s(y)_b) = (s(x)_a|s(x)_c)U(x,y)^c_b = \gamma_{0ac}U(x,y)^c_b. \tag{21}$$

If we had two indices per spinor, the rightmost expression would consist of the sum of the U matrix apropos to the spin index acting on the Dirac matrix γ_0 times (tensor product with) the identity matrix for the other index plus the other U matrix times (tensor product with) γ_0. But, in our sample case, we find the basis-independent expression for the two-point function to be

$$\begin{aligned}\langle(\psi(x)|\psi(y))\rangle &= \langle(s(x)_a\psi(x)^a|s(y)_b\psi(y)^b)\rangle\\ &= \langle\psi(x)^{*a}\psi(y)^b(s(x)_a|s(y)_b)\rangle\\ &= \langle\psi(x)^{*a}\psi(y)^bU(x,y)^c_b\rangle\gamma_{0ac}\\ &= \langle\bar\psi(x)U(x,y)\psi(y)\rangle.\end{aligned} \tag{22}$$

The last expression is the usual time-ordered functional average. The details of the choice of path are suppressed.

We will now construct some four-point functions. They are also expectations of inner products. Thus consider

$$\begin{aligned}&\langle(\psi(x)\otimes\psi(y)|\psi(z)\otimes\psi(w))\rangle\\ &\quad= \langle\psi^{*a}(x)\psi^{*b}(y)\psi^c(z)\psi^d(w)(s_a\otimes s_b|s_c\otimes s_d)\rangle\\ &\quad= \langle\psi^{*a}(x)\psi^{*b}(y)\psi^c(z)\psi^d(w)(s_a|s_c)|(s_b|s_d)\rangle \qquad (23)\\ &\quad= \langle\psi^{*a}(x)\psi^{*b}(y)\psi^c(z)\psi^d(w)U(x,y)^e_a\gamma_{0ce}U(y,w)^f_b\gamma_{0df}\rangle\\ &\quad= -\langle\bar\psi(x)U(x,z)\psi(z)\bar\psi(y)U(y,w)\psi(w)\rangle.\end{aligned}$$

Generalizing to n-point functions is straightforward. *Observe that the requirement that we consider basis-independent objects only has left us in the color-blind sector.* To see the color we must explicitly introduce basis elements into the n-point function. For example, $\langle(\psi(x)|s_a(x))(s_b(y)|\psi(y))\rangle = \langle\bar\psi_a(x)\psi_b(y)\rangle$.

III. BASIS FOR PARAFIELDS

We will now briefly review parastatistics.[2,4] The defining relations are the trilinears (think of the extra indices as momenta)

$$\begin{aligned}&[[\phi^\dagger_k,\phi_l]_\pm,\phi_m]_- = -2\,\delta_{km}\phi_l,\\&[[\phi_k,\phi_l]_\pm,\phi_m]_- = 0.\end{aligned}\tag{24}$$

We will use $[\,,\,]_+$ for $\{\,,\,\}$ and $[\,,\,]_-$ for $[\,,\,]$, interchangeably. Throughout this paper the upper subscriped sign will refer to the para-Bose case, the lower sign to the para-Fermi case. Assuming that there exists a vacuum such that $\phi_k|0\rangle = 0$, Green introduced the ansatz that each operator ϕ could be split into a sum of n operators. The number n of the components is the "order" of the parastatistics. It can take any positive integer value

$$\phi_k = \sum_{a=1}^{n}\phi_k^{(a)},\tag{25}$$

and similarly for the conjugates. These operators obey the usual commutation or anticommutation relations for equal values of a, but obey the following unusual relations for unequal values of their indices:

$$\begin{aligned}&[\phi_k^{(a)},\phi_l^{(a)\dagger}]_\mp = \delta_{kl},\\&[\phi_k^{(a)},\phi_l^{(b)\dagger}]_\pm = 0,\quad a\neq b.\end{aligned}\tag{26}$$

The inner brackets of the trilinear relationship in Eq. (24) are now given by

$$[\phi^\dagger_k,\phi_l]_\pm = \sum_a[\phi_k^{(a)\dagger},\phi_l^{(a)}]_\pm.\tag{27}$$

So, it is clear that relations (24) are satisfied by operators obeying the ansatz.

The relationship between operators and c numbers for use in functional integrals is well known. The crucial difference is that commutators and anitcommutators must vanish for the c numbers and not vanish for the operators. The analogs to relations (24) for c-number parafields are

$$[[\phi^\dagger_k,\phi_l]_\pm,\phi_m]_- = 0,\quad [[\phi_k,\phi_l]_\pm,\phi_m]_- = 0.\tag{28}$$

We can call this a para-c-number algebra. It can be satisfied by introducing an analog of Green's ansatz for q-number fields, $\phi_k = \Sigma\phi_k^{(a)}$, where now these are c numbers. The replacements for relations (26) are the following:

$$[\phi_k^{(a)},\phi_l^{(a)\dagger}]_\mp = 0,\quad [\phi_k^{(a)},\phi_l^{(b)\dagger}]_\pm = 0,\quad a\neq b.\tag{29}$$

We will use the para-c-number algebra and the functional integral formalism and not the operator construction.

Let us examine the behavior of the generalized number operator (bracket of a parafield with its conjugate) when we introduce a basis instead of using Green's ansatz. If the parafield is $\phi_k = \phi_k^a e_a$, then the generalized number operator is

$$\begin{aligned}[\phi_k^a e_a,\phi_l^{b\dagger}e_b^*]_\pm &= \phi_k^a e_a\phi_l^{b\dagger}e_b^* \pm \phi_l^{b\dagger}e_b^*\phi_k^a e_a\\&= \phi_k^a\phi_l^{b\dagger}(e_a e_b^* + e_b^* e_a)\\&= \phi_k^a\phi_l^{b\dagger}(2\,\delta_{ab})\\&= \sum_a[\phi_k^a,\phi_l^{a\dagger}]_\pm.\end{aligned}\tag{30}$$

In the second line we have used the fact that bosons commute and spinors anticommute to find the same sign in both cases. The expression in the third line is obtained by working backwards from the last line, which is what we want. We can therefore infer that the relation $e_a e_b^* + e_b^* e_a = 2\,\delta_{ab}$ is needed for our basis elements to work properly.

This is one part of the relations which define a complex Clifford algebra. In such an algebra the following two anticommutation relations (and their conjugates) hold:

$$\{e_a^*,e_b\} = 2\,\delta_{ab}I,\quad \{e_a,e_b\} = 0.\tag{31}$$

The matrix I is the unit element of the Clifford algebra (the identity matrix). These two relations will work nicely in the case of parafermions. But in the case of parabosons the second relation makes the construction of symmetric representations exceedingly awkward if we try to put the parabosons in the "vector" sector of the complex Clifford algebra.

A real basis for a Clifford algebra E_A obeys the following relation: $\{E_A,E_B\} = 2\eta_{AB}I$, where η is the (indefinite) metric. Indeed, the complex basis elements have a natural realization as linear combinations of real E_A's. Of course, there must be twice as many E_A's as there are e_a's. It is possible to avoid explicit introduction of the imaginary unit i by taking a metric η which has equal numbers of positive and negative square basis elements. Thus, let E_a be those with positive square, and let $E_{a'}$ be those with negative square. The "complex conjugate" * acts by sending $E_{a'}$ into minus itself. Define the "complex" basis $e_a = (E_a + E_{a'})/\sqrt{2}$ and $e_a^* = (E_a - E_{a'})/\sqrt{2}$. Alternately, you can use a real metric δ_{AB} and replace $E_{a'}$ with $iE_{a'}$; complex conjugation is the usual. You can verify the relations (31) in either case. Note that the subalgebra of only e's or only e^*'s is a Grassmann algebra.

We want the parabosons to have the possibility of coupling with bilinears constructed from the fermions. We have seen that gauge fields can be viewed as rank two tensors in standard basis-independent field theory [cf. Eq. (12) and remarks following it]. Of course, for the gauge field, the minimal coupling implicit in the gauge covariant derivative which arose as a piece of the derivative of the basis-independent field [Eq. (1)] made it unnecessary to explicitly utilize this tensor structure. But in the case of scalar-fermion coupling such considerations are necessary (unless the scalar is a piece of a connection which split off under dimensional reduction). In any case the obvious thing to try is to make parabosons rank two tensors in the complex Clifford algebra. The parabosons are taken to have $e_a^* e_b$ as basis elements.

Exponentials of sums of skew-Hermitian coefficients times $e_a^* e_b$ represent U(n). To obtain a U(n) rotation of e_c perform a similarity transformation on e_c with an element of that group of matrices. The infinitesimal relationship $[e_a^* e_b,e_c] = -2\,\delta_{ac}e_b$ can be derived from the anticommutation relations of the e's. The additional requirement that the metric $\{e_a^*,e_b\} = 2\,\delta_{ab}I$ be preserved restricts transformations to U(n). Parascalars can be in GL(n,C) because they are not required to preserve the metric. But the same commutator can be used to couple these scalars with the fermions [since GL(n,C) is just the complex extension of U(n), i.e., U(n) with complexified coefficients]. To get parabosons which are in the fundamental representation of U(n) we use the complex Clifford algebra with $n+1$ generators, a standard device.

So far, it all looks good. The problem is that while the anticommutator $\{e_a^* e_b, e_c^* e_d\}$ has a component which is the identity matrix multiplied by $\delta_{bc}\,\delta_{ad}$, which makes the tensor analog of Eq. (30) work, it has another component which is a rank four tensor. To get rid of this unwanted piece we take a trace (suitably normalized). When we dealt with the para-c-number fermions, we could have introduced this trace but we were lucky and wound up with a multiple of the identity. To make everything uniform we define the generalized number operator to be the (normalized) trace of either the commutator (parafermions) or the anticommutator (parabosons). *Thus, in front of the inner bracket of the defining trilinears we insert a (normalized) trace*

$$\begin{aligned}&[\mathrm{tr}[\phi_k^\dagger,\phi_l]_\pm,\phi_m]_- = -2N\,\delta_{km}\phi_1,\\&[\mathrm{tr}[\phi_k,\phi_l]_\pm,\phi_m]_- = 0.\end{aligned}\tag{32}$$

Here, N is the normalizing factor. Note, if the coefficients are quantum field theory operators, the vacuum expectation of the inner bracket of the first equation in (24) must be subtracted off, as well as taking the trace.[2] The right-hand side of the first of these relations is set to zero in the case of a para-c-number algebra.

With the basis elements obeying the relations (31) and with the coefficients being Grassmann or scalar c-numbers, one can now verify that the combined entity obeys a para-c-number algebra. The new Green trilinears (32) are satisfied.

Before going on we must consider the case where a parafermion is required to transform under two or more symmetry groups (carry two indices). Because of the fact that the e's form a Grassmann algebra when taken by themselves, we cannot use a simple tensor product basis such as $e_a \otimes h_k$, where h_k denotes some other basis (either complex Clifford or not). The reason is that if there are n e's and N h's (assume $N>n$), the obvious product of these elements (the Clifford product for the e's and either the tensor or the Clifford product for the h's) allows at most n terms, instead of $n\times N$ terms [any more terms in the product produces a zero due to the second of Eqs. (31)]. Here, $N\times n$ terms surviving is okay on account of the fact that the Fermi coefficients vanish anyway for more than $N\times n$ terms. So, the correct basis is e_{ak}, where these are the generators of an $(N\times n)$-dimensional complex Clifford algebra (as mentioned above, this is most natural from a Kaluza–Klein or totally unified perspective).

IV. EQUIVALENCE OF GAUGE THEORY AND PARAFIELD FRAMEWORKS

We first showed that the action and its gauge-invariant moments did not depend upon the usual basis elements. We also saw how to extract basis-dependent quantities. Now we have seen how to create and use a parabasis. To prove that the physics of para- and usual fields is the same we must see that the action still will not depend upon the new basis elements and that all the moments can still be extracted.

The idea now is to replace the bundle of frames piece of the space S defined in Eq. (8) with the complex Clifford algebra of parabases. Operationally, this simply means that in all the formulas given above, which use the basis $s_a(x)$, one replaces the basis elements $s_a(x)$ with the complex Clifford matrices $e_a(x)$. But, technically, there is more to it. One usually requires that all the fermionic field coefficients at all locations anticommute with each other. In order to have anticommutation relations between all spinor components at all locations, there must be an enormous Grassmann algebra. The algebra of parabases is no bigger than the Grassmann algebra of coefficients. The size of either of these algebras is found by taking into account the number of degrees of freedom per point (consider the case where there are only parafermions and coupled gauge fields) and the total number of points in the lattice. If there are n complex components at a point, the smallest (complex-irreducible) complex Clifford algebra with that many generators has rank 2^n. If there are two spaces of bases, each with that many components, the smallest complex Clifford algebra with enough generators for both is of rank 2^{n+n}. Thus, if there are p points in the lattice, the smallest matrix algebra with enough generators has rank $2^{n\times p}$. The smallest matrix algebra representation of the fermionic coefficients is twice this size, as we can infer from the fact that the subset of generators e or e^* by themselves formed a Grassmann algebra using simple matrix multiplication [see the paragraph after Eq. (31)]. So, that is the scale of these objects. Note that there is an intriguing duality in the size and nature of these objects. This can be pushed further (their dimensions can, in fact be taken equal by replacing the matrix product for the fermions with a totally antisymmetrized matrix product).

Thus, we replace the space S with the parastatistics space P which is a pair $P=(A,L)$ with $A=C(2^{n\times p})\times G_c^{n\times p}$, where $C(2^{n\times p})$ denotes the complex Clifford algebra of rank $2^{n\times p}$, $G_c^{n\times p}$ is the Grassmann algebra of spinors, and L is the lattice. The portion of $C(2^{n\times p})$ which replaces a fiber is a subalgebra $C(2^n)$. There is a section which picks out a separate copy of $C(2^n)$ for each point in L. A connection is a U(n)-valued map for each link. It can be thought of as being represented by matrices of the form $\exp(-\omega^{ab}e_a^*(x)e_b(x+\Delta)/2)$ for each link. These matrices act by conjugation on the basis $e_c(x)$ to push it forward and rotate it into a basis at $x+\Delta$, as one can see by trying the infinitesimal form and using the brackets for the e's. Scalars can be viewed as sections taking values in the two-tensor subalgebra of $C(2^{n\times p})$.

Note that there are some new entities which have arisen. Consider the space of objects that the matrices in $C(2^{n\times p})$ act upon (via left-hand multiplication). Such objects will rotate one-half as fast as spinors. That is, if $(\exp(\omega^{ij}\Sigma_{ij}))_l^k$ is a Lorentz transformation of a vector v^l, then $(\exp(\omega^{ij}\sigma_{ij})/2)_b^a$ acts on the spinor ψ^b and the new object $(\exp(\omega^{ij}\sigma_{ij}^{ab}e_a^*e_b)/4)_s^r$ acts on some new object θ^s. The reason that $\Sigma_{ij\,l}^{k}$ did not appear in the second of these matrices is that $\Sigma_{ij}^{kl}=(\delta_i^k\,\delta_j^l-\delta_j^k\,\delta_i^l)/2$. The substitution was made. Thus the last two matrices bear precisely the same relationship with their antecedents. The objects θ^s rotate one-quarter as fast as the vectors; so, they could be called spin-$\frac{1}{4}$ objects. These objects arise because in physics spinors are Grassmann elements which can be represented by matrices which in turn can act on some new space. Although, we have come to them by thinking about the complex Clifford algebra that has the same number of elements as there are spinor components (the parabasis). Furthermore, there is a real irreducible analog to this process. For example, a Majorana spinor ψ^A can be coupled to $\bar{\theta}\Gamma^A\theta$; a bilinear coupling can also be made $\psi_A G^A_B \bar{\theta}\Gamma^B\Gamma^C\theta\psi_C$. Here G is the Majorana metric. Or a Dirac mass term can be formed $\bar{\psi}^a(\bar{\theta}e^*_a e_b\theta)\psi^b$. In this case the spin-zero mass term is formed out of a spin-up and spin-down θ field. We will call θ the daughter field. Like the Russian matruska dolls, this process can continue. For those interested in the idea of constituents, the fact that spinors can be pulled apart (as vectors can be) may be of interest.

The internal metric is replaced by a trace over the complex Clifford matrices (suitably normalized). We have

$$\begin{aligned}(s_a(x)|s_b(y)) &= \gamma_{0ac}\,U(x,y)^c_b\\ &= N^{-1}\,\mathrm{tr}(e^*_a(x)e_b(y)).\end{aligned} \tag{33}$$

Here N is the normalizing factor. The Clifford matrices have the same kind of path dependence that the usual basis elements do. It is natural to define the rightmost entry as the inner product of the e's and write it as $(e_a(x)|e_b(y))$.

The next thing to do is to extend this inner product to products of bases. Here is one place where there can be some technical differences between local parastatistics and standard gauge theories. Consider first the inner product of two rank two tensors at a given point. When we defined the inner product on tensors (via tensor products of basis elements) in Eq. (14), we took that product to be simply the product of the first basis element in the left-hand slot of the product paired with the first basis element in the right-hand slot, then the second with the second, etc. That is, we defined n-fold products by induction based upon the simple inner product of two elements. However, in the Clifford matrix case the inner product can be expressed in terms of (suitably normalized) traces of matrix products. To see this for a rank two Clifford matrix observe that (at a point)

$$N^{-1}\mathrm{Tr}((e^*_a(x)e_b(x))^*(e^*_c(x)e_d(x))) = \delta_{ab}\,\delta_{cd} - \delta_{ac}\delta_{bd}. \tag{34}$$

The trace Tr refers to the trace on the daughter indices carried by the e's. We will use tr for the trace on the indices a, b, etc. If A and B are rank two tensors (in e space) we have $N^{-1}\,\mathrm{Tr}(A^*B) = \mathrm{tr}\,A^\dagger\ \mathrm{tr}\,B - \mathrm{tr}\,A^\dagger B$. This is not a definite product. Recall that the inner product of the s_a's gave the negative of the second term. To insure precise equality of the theories we must take our inner products to be the same. So, define the inner product of rank two tensors as follows:

$$\begin{aligned}(A|B) &= N^{-1}\,\mathrm{Tr}(A^*)N^{-1}\,\mathrm{Tr}(B) - N^{-1}\,\mathrm{Tr}(A^*B)\\ &= \mathrm{tr}\,A^\dagger B.\end{aligned} \tag{35}$$

The extension to the case of higher tensor products is easy and is done so that the analog of Eq. (14) holds. That is, there is only one product of δ matrices on the right-hand side. It is not essential to relate an inner product to the trace; one can simply define an inner product, but the trace is the natural inner product on the vector sector of these algebras. So, we have shown the relationship of our inner product with the trace.

V. CONCLUSION

With this space P and this set of inner products, it is clear that the lattice versions of the local parastatistical model (as we have defined it) and the standard lattice field theory have the same action and n-point functions. We assume that the continuum limit of the functional theory defined on the standard space S [Eq. (8)] exists and is well defined. Likewise, we assume that the continuum limit of the functional theory defined on the parastatistics space P exists and is well defined. Assuming this, one sees that, operationally, one set of basis elements has been replaced by another with a one-to-one correspondence; further, the inner products also correspond directly. The two theories therefore have the same predictions.

ACKNOWLEDGMENTS

I would like to thank O. W. Greenberg and H. Rubinstein for stimulating conversations.

This work was supported in part by the National Science Foundation.

[1]K. I. Macrae, Phys. Rev. D **18**, 3737, 3761, 3777 (1978), and references therein.

[2]O. W. Greenberg and K. I. Macrae, Nucl. Phys. B **219**, 358 (1983); O. W. Greenberg and K. I. Macrae, "Locally gauge invariant formulation of parastatistics," in *Workshop on Non-Perturbative Quantum Chromodynamics*, edited by K. A. Milton and M. A. Samuel (Birkhauser, Boston, 1983), pp. 181–190 and references therein.

[3]E. Cartan, *Lecons sur la Géometrie des Espaces de Riemann* (Gauthier Villars, Paris, 1963); E. Cartan, "La Méthode du Répère Mobile," in *La Théorie des Groupes Continus et les Espaces Généralisés* (Actualitiés Sci. Ind., Hermann, Paris, 1935).

[4]H. S. Green, Phys. Rev. **90**, 270 (1953).

The Dirac Coulomb Green's function and its application to relativistic Rayleigh scattering

M. K. F. Wong
Fairfield University, Fairfield, Connecticut 06430

E. H. Y. Yeh
Naval Weapons Center, China Lake, California 93555

(Received 1 August 1984; accepted for publication 22 February 1985)

The Dirac Coulomb Green's function is obtained in both coordinate and momentum space. The Green's function in coordinate space is obtained by the eigenfunction expansion method in terms of the wave functions obtained by Wong and Yeh. The result is simpler than those obtained previously by other authors, in that the radial part for each component contains one term only instead of four terms. Our Green's function reduces to the Schrödinger Green's function upon some simple conditions, chiefly by neglecting the spin and replacing λ by l. The Green's function in momentum space is obtained as the Fourier transform of the coordinate space Green's function, and is expressed in terms of basically three types of functions: (1) $F_A(\alpha;\beta_1\,\beta_2\,\beta_3;\gamma_1\,\gamma_2\,\gamma_3; z_1\,z_2\,z_3)$, (2) the hypergeometric function, and (3) spherical harmonics. The matrix element for Rayleigh scattering, or elastic Compton scattering, from relativistically bound electrons is then obtained in analytically closed form. The matrix element is written basically in terms of the coordinate space Dirac Coulomb Green's function. The technique used in the evaluation of the matrix element is based on the calculation of the momentum space Dirac Coulomb Green's function. Finally the relativistic result is compared with the nonrelativistic result.

I. INTRODUCTION

The Dirac Coulomb Green's function (DCGF) is of great importance in the quantum electrodynamics of the relativistic electron in the presence of a Coulomb field. So far, only the coordinate space DCGF has been obtained.[1] The radial part is obtained from the solutions to the homogeneous equation. Each solution ψ_i^a contains the sum of two terms. Thus the radial DCGF, being the product of two ψ's, contains four terms. In a previous paper,[2] we have obtained a simplified solution to the homogeneous Dirac Coulomb equation where each component contains one term only. Using these wave functions we calculate the DCGF by the eigenfunction expansion method, which is the most orthodox way to calculate the Green's function. We integrate the continuum wave functions by contour integration. The pole terms from the contour integration then cancel exactly the bound state sums, with the resulting Green's function in a similar form originally constructed by Brown and Schaefer[3] and Wichmann and Kroll.[1] However, our results are simpler than theirs in that each component of the radial Green's function contains one term only instead of four terms. Moreover, our results can be directly compared with the nonrelativistic case,[4] which is well known. In fact, our result reduces to the Schrödinger Green's function upon two simple conditions: (1) neglect the spin, and (2) replace λ by l.

Next we use the coordinate DCGF to obtain the momentum space DCGF by Fourier transformation. It is shown that the angular part can be easily done. It remains to evaluate the integrals over r_1 and r_2. Because of the presence of the θ function in the DCGF, the integral over, say, r_2 is finite, i.e., the limits go from zero to r_1. We find that a similar integral has been evaluated by Ogata and Asai[5] in connection with finite nuclear size effects. In the present case, by modifying the method of Ogata and Asai slightly, we obtain a result expressible in terms of ${}_2F_1$ functions. The remaining integral is a Laplace transform of $f(r)=r^{\nu-1} \times M_{\kappa_1,\mu_1-1/2}(\alpha_1 r)\cdots M_{\kappa_n,\mu_n-1/2}(\alpha_n r)$, which can be found in Ref. 6. The result is expressible in terms of the Lauricella function F_A of n variables. In the present case, $n=3$, and this function is a hypergeometric function of three variables. Thus the DCGF in momentum space is obtained in closed form.

In the second half of this paper, we apply the Green's function to a practical problem: to find the matrix element of Rayleigh scattering from relativistically bound electrons. The method used is as follows. The matrix element for Rayleigh scattering is written basically in terms of the coordinate space Green's function, which can be found in Brown and Schaeffer.[3] Then one has to evaluate an integral over the intermediate states. This is done according to the technique used in the first half of this paper for the calculation of the momentum space Green's function. Basically the angular parts are written in terms of spherical harmonics. The Fourier integral over the angular variables can then be performed. Then only the radial integrals are left. These are of two different kinds: a finite integral over, say, r_2, with limits going from zero to r_1, and an integral over r_1, with limits from zero to infinity. Both these integrals can be evaluated in closed form.

Using the ground state $1S_{1/2}$ as the initial state, we find that the integrals involved are very similar to the ones encountered in the evaluation of the momentum space Green's function. It is not hard to generalize our results to arbitrary initial states, since the technique we have developed can be generalized to more complicated structures. Moreover, it is also possible to obtain matrix elements for inelastic Compton scattering for relativistically bound electrons, since the same technique applies.

In the final section of the paper, the question of conver-

 0022-2488/85/071701-10$02.50

gence of the sums, especially over j or equivalently l, is discussed. We make a brief comparison between our result and the nonrelativistic result. If the initial state is chosen to be an arbitrary bound state, then it is shown that to within on order of 0.01 percent, the relativistic result contains the nonrelativistic result. Moreover, the relativistic result contains a correction term which corresponds to the Furry approximation, and is of the order of α, the fine structure constant $\frac{1}{137}$. However, if the initial state is chosen to be the ground state $1S_{1/2}$, then the correction term does not appear. Thus we have obtained the remarkable conclusion that the nonrelativistic result for Rayleigh scattering and Lamb shift is accurate to within 0.01% when compared with the relativistic result for the $1S_{1/2}$ state. A detailed presentation of these results will be given in a future publication.

II. COORDINATE SPACE DIRAC COULOMB GREEN'S FUNCTION

The coordinate space DCGF was obtained by Wichmann and Kroll.[1] Brown and Schaefer[3] also obtained the DCGF in general form, but did not write out the expression explicitly. Later publications basically follow the same expression given by Wichmann and Kroll; see, e.g., Mohr,[7] Gyulassy,[8] and Hylton.[9] The Green's function is obtained in terms of solutions to the homogeneous Dirac Coulomb equation. This solution, explicitly written out by Wichmann and Kroll, is basically the same form as obtained by Darwin,[10] derived in detail in Bethe and Salpeter[11] or Rose.[12] It is well known that the radial solution contains two terms for each component. Thus the DCGF so obtained, being the product of two wave functions, contains four terms for each component.

We have obtained in a previous paper[2] a simplified solution to the Dirac Coulomb equation where each component of the radial wave function contains one term only. Using this solution we are able to obtain the DCGF as one term also, instead of four terms. This is done through the eigenfunction expansion method. All notations in the following are the same as in Ref. 2, unless otherwise stated.

The DCGF $G(\mathbf{r}_2,\mathbf{r}_1,z)$ satisfies the equation

$$[H(\mathbf{r}_2)-zI\,]\,G(\mathbf{r}_2,\mathbf{r}_1,z)=\delta^3(\mathbf{r}_2-\mathbf{r}_1)I, \tag{2.1}$$

where I is the 4×4 identity matrix and $H(\mathbf{r})$ is the Dirac Hamiltonian $H^0(\mathbf{r})$ defined in Ref. 2, after the transformation S, i.e.,

$$H(\mathbf{r})=SH^0(\mathbf{r})S^{-1}, \tag{2.2}$$

$$H^0(\mathbf{r})=\boldsymbol{\alpha}\cdot\mathbf{p}-\beta m+V, \tag{2.3}$$

$$S=\exp[-\tfrac{1}{2}\rho_2\boldsymbol{\sigma}\cdot\hat{r}\tanh^{-1}(Ze^2/K)], \tag{2.4}$$

$$K=\beta(\boldsymbol{\sigma}\cdot\mathbf{L}+1). \tag{2.5}$$

The DCGF, expanded over the complete set of spin angular eigenfunctions of K [eigenvalue $\kappa=\tilde{\omega}(j+\frac{1}{2})$], and $J_3=L_3+\frac{1}{2}\sigma_3$ (eigenvalue μ), $\chi^\mu_\kappa(\hat{r})$, is

$$G(\mathbf{r}_2,\mathbf{r}_1,z)=\sum_{\kappa\mu}\begin{pmatrix}G^{11}_\kappa(r_2\,r_1z)S_2\,\chi^\mu_{-\kappa}(\hat{r}_2)\,\chi^{\mu\,+}_{-\kappa}(\hat{r}_1)S_1 & -iG^{12}_\kappa(r_2\,r_1z)S_2\,\chi^\mu_{-\kappa}(\hat{r}_2)\,\chi^{\mu\,+}_\kappa(\hat{r}_1)S_1\\ iG^{21}_\kappa(r_2\,r_2z)S_2\,\chi^\mu_\kappa(\hat{r}_2)\,\chi^{\mu\,+}_{-\kappa}(\hat{r}_1)S_1 & G^{22}_\kappa(r_2\,r_1z)S_2\,\chi^\mu_\kappa(\hat{r}_2)\,\chi^{\mu\,+}_\kappa(\hat{r}_1)S_1\end{pmatrix}, \tag{2.6}$$

where S_i, $i=1,2$, is the same transformation S as in Eq. (2.4) with $\hat{r}$ replaced by $\hat{r}_i$. The spin angular eigenfunctions $\chi^\mu_\kappa(\hat{r})$ are defined unanimously in all literature; see, e.g., Refs. 2, 7, and 9, and also Eq. (4.10). The summation over μ in Eq. (2.6) can be performed in terms of $\pi_\kappa(\hat{r}_2,\hat{r}_1)$, where

$$\begin{aligned}\pi_\kappa(\hat{r}_2,\hat{r}_1)&=\sum_\mu\chi^\mu_\kappa(\hat{r}_2)\,\chi^{\mu\,+}_\kappa(\hat{r}_1)\\&=\frac{|\kappa|}{4\pi}\left\{IP_{|\kappa+1/2|-1/2}(\xi)+\frac{1}{\kappa}\,i\boldsymbol{\sigma}\cdot(\hat{r}_2\times\hat{r}_1)\,P'_{|\kappa+1/2|-1/2}(\xi)\right\},\end{aligned} \tag{2.7}$$

where $\xi=\hat{r}_2\cdot\hat{r}_1$, P is the Legendre polynomial, and P' is the derivative of P with respect to the argument; I is the 2×2 identity matrix. Thus (2.6) can be written as follows:

$$G(\mathbf{r}_2,\mathbf{r}_1,z)=\sum_\kappa\begin{pmatrix}G^{11}_\kappa(r_2\,r_1z)S_2\,\pi_{-\kappa}(\hat{r}_2\,\hat{r}_1)S_1 & G^{12}_\kappa(r_2\,r_1z)S_2\,i\boldsymbol{\sigma}\cdot\hat{r}_2\,\pi_\kappa(\hat{r}_2\,\hat{r}_1)S_1\\ -G^{21}_\kappa(r_2\,r_1z)S_2\,i\boldsymbol{\sigma}\cdot\hat{r}_2\,\pi_{-\kappa}(\hat{r}_2\,\hat{r}_1)S_1 & G^{22}_\kappa(r_2\,r_1z)S_2\,\pi_\kappa(\hat{r}_2\,\hat{r}_1)S_1\end{pmatrix}. \tag{2.8}$$

Inserting (2.8) into (2.1), we obtain the differential equation satisfied by the radial Green's function $G^{ij}_\kappa(r_2\,r_1z)$. In Ref. 2, we have explicitly calculated the operator $H(\mathbf{r}_2)$. Thus we obtain the following equation satisfied by the radial Green's function $G^{ij}_\kappa(r_2\,r_1z)$:

$$\begin{bmatrix}\kappa z/\gamma+m & \dfrac{d}{dr_2}+\dfrac{1+\gamma}{r_2}-\dfrac{zZe^2}{\gamma}\\ \dfrac{d}{dr_2}+\dfrac{1-\gamma}{r_2}+\dfrac{zZe^2}{\gamma} & m-\kappa z/\gamma\end{bmatrix}\begin{bmatrix}G^{11}_\kappa(r_2\,r_1z) & G^{12}_\kappa(r_2\,r_1z)\\ G^{21}_\kappa(r_2\,r_1z) & G^{22}_\kappa(r_2\,r_1z)\end{bmatrix}=I\,\frac{\delta(r_2-r_1)}{r_1\,r_2}. \tag{2.9}$$

The solution for $G^{ij}_\kappa(r_2\,r_1z)$ is obtained by the eigenfunction expansion method in terms of the wave functions which are solutions to the homogeneous equation

$$\begin{bmatrix}\kappa z/\gamma+m & \dfrac{d}{dr_2}+\dfrac{1+\gamma}{r_2}-\dfrac{zZe^2}{\gamma}\\ \dfrac{d}{dr_2}+\dfrac{1-\gamma}{r_2}+\dfrac{zZe^2}{\gamma} & m-\kappa z/\gamma\end{bmatrix}\begin{bmatrix}\psi(-\tilde{\omega})\\ \psi(\tilde{\omega})\end{bmatrix}=0. \tag{2.10}$$

In Ref. 2, we have obtained the solution to (2.10). The bound state wave functions for $n_r, \tilde{\omega}, j$ are

$$\psi(\tilde{\omega}) = (E\kappa/\gamma + m)^{1/2} N'(\tilde{\omega}) \rho^{\lambda(\tilde{\omega})} e^{-\rho/2} \times {}_1F_1(-n_r, 2\lambda(\tilde{\omega}) + 2, \rho)$$

and

$$\psi(-\tilde{\omega}) = i\tilde{\omega}(E\kappa/\gamma - m)^{1/2} N'(-\tilde{\omega}) \rho^{\lambda(-\tilde{\omega})} e^{-\rho/2} \times {}_1F_1(-n_r, 2\lambda(-\tilde{\omega}) + 2, \rho). \tag{2.11}$$

Since both wave functions are completely equivalent to the Schrödinger solution, if one replaces $\lambda(\pm\tilde{\omega})$ by $l(\pm\tilde{\omega})$, we can rewrite (2.11) in terms of the normalized Schrödinger wave function $R(\pm\tilde{\omega})$. Thus

$$\psi(\tilde{\omega}) = (E\kappa/\gamma + m)^{1/2} NR(\tilde{\omega}), \tag{2.12}$$

$$\psi(-\tilde{\omega}) = i\tilde{\omega}(E\kappa/\gamma - m)^{1/2} NR(-\tilde{\omega}). \tag{2.13}$$

In (2.12) and (2.13), N is a single normalization constant for the bound state $n_r, \tilde{\omega}$, and j, and $R(\pm\tilde{\omega})$ satisfy the equation

$$\int_0^\infty R^2(\pm\tilde{\omega}) r^2\, dr = 1. \tag{2.14}$$

Then we normalize $\psi(\tilde{\omega})$ and $\psi(-\tilde{\omega})$ such that

$$\int_0^\infty r^2\, dr(|\psi(\tilde{\omega})|^2 + |\psi(-\tilde{\omega})|^2) = 1. \tag{2.15}$$

From (2.15), we obtain

$$N = (2E\kappa/\gamma)^{-1/2}. \tag{2.16}$$

Thus we shall write $\psi(\pm\tilde{\omega})$ in the form (2.12) and (2.13), with N given by (2.16). Explicitly,

$$\begin{aligned} R(\pm\tilde{\omega}) = {} & [\Gamma(2\lambda(\pm\tilde{\omega}) + 2)]^{-1}(2\mu_1)^{3/2} \\ & \times [\Gamma(n + \lambda(\pm\tilde{\omega}) + 1)\Gamma(n - \lambda(\pm\tilde{\omega}) + 1)]^{1/2} \\ & \times [2n\Gamma(n - \tilde{\lambda}(\pm\tilde{\omega}) + 1)\Gamma(n - \tilde{\lambda}(\pm\tilde{\omega}))]^{-1/2} \\ & \times \rho^{\lambda(\pm\tilde{\omega})} e^{-\rho/2}\, {}_1F_1(-n_r, 2\lambda(\pm\tilde{\omega}) + 2, \rho), \end{aligned} \tag{2.17}$$

with $\mu_1 = (m^2 - E^2)^{1/2}$ and where the terms $\Gamma(n - \tilde{\lambda}(\pm\tilde{\omega}) + 1)$ and $\Gamma(n - \tilde{\lambda}(\pm\tilde{\omega}))$ in the bracket could be written independently of $\tilde{\lambda}(\pm\tilde{\omega})$, by using the equations

$$\tilde{\lambda}(\tilde{\omega}) = n - n_r - 1, \tag{2.18}$$

$$\tilde{\lambda}(-\tilde{\omega}) = \tilde{\lambda}(\tilde{\omega}) \mp 1, \text{ for } (\tilde{\omega} \pm 1), \tag{2.19}$$

or

$$[\Gamma(n - \tilde{\lambda}(\pm\tilde{\omega}) + 1)\Gamma(n - \tilde{\lambda}(\pm\tilde{\omega}))]^{-1} = \begin{cases} [\Gamma(n_r + 2)\Gamma(n_r + 1)]^{-1} & \text{for } (\tilde{\omega}), \\ [\Gamma(n_r + 2 \pm 1)\Gamma(n_r + 1 \pm 1)]^{-1} & \text{for } (-\tilde{\omega}) \\ (\tilde{\omega} = \pm 1). \end{cases} \tag{2.20}$$

To compare our result with the Schrödinger result, we can also rewrite (2.12) as

$$\begin{aligned} \psi(\tilde{\omega}) = {} & (E\kappa/\gamma + m)^{1/2}(\gamma/2E\kappa)^{1/2}\, 2^{3/2} \mu_1^2 \\ & \times \left[\frac{\Gamma(n + \lambda + 1)}{2Ze^2E\Gamma(n - \lambda)}\right]^{1/2} \frac{1}{\Gamma(2\lambda + 2)} \\ & \times e^{-\rho/2} \rho^\lambda\, {}_1F_1(-n_r, 2\lambda + 2, \rho). \end{aligned} \tag{2.21}$$

Next we write the continuum wave function as

$$\begin{aligned} \psi^{\hat{\rho}}(\tilde{\omega}) &= (E\kappa/\gamma + m)^{1/2} \phi^{\hat{\rho}}(kr), \\ \psi^{\hat{\rho}}(-\tilde{\omega}) &= i\tilde{\omega}(E\kappa/\gamma - m)^{1/2} \phi^{\hat{\rho}}(kr), \\ \phi^{\hat{\rho}}(kr) &= \frac{\sqrt{2}mk}{E^{3/2}} \left(\frac{\gamma}{2\pi\kappa}\right)^{1/2} \Gamma(\lambda + 1 + i\eta) e^{i\sigma} \\ & \quad \times e^{\epsilon_{\hat{\rho}}\pi\eta/2}(-2ikr\,\epsilon_{\hat{\rho}})^{-1} M_{i\eta,\lambda + 1/2}(-2ikr\,\epsilon_{\hat{\rho}}), \end{aligned} \tag{2.22}$$

where $\eta = Ze^2E/k$, $\epsilon_{\hat{\rho}} = \pm 1$ for $\hat{\rho} = 1$ and 2, and $e^{i\sigma}$ is the argument of

$$[\Gamma(\lambda + 1 - i\eta)/\Gamma(\lambda + 1 + i\eta)]^{1/2}. \tag{2.23}$$

The $e^{i\sigma}$ will not contribute to the Green's function since it is always cancelled by $e^{-i\sigma}$ coming from ψ^*. Here $M_{i\eta,\lambda+1/2}(z)$ is the Whittaker function defined by

$$M_{i\eta,\lambda+1/2}(z) = z^{\lambda+1} e^{\mp z/2}/\Gamma(2\lambda + 2) \times {}_1F_1(\lambda + 1 \mp i\eta, 2\lambda + 2, \pm z). \tag{2.24}$$

We also take this opportunity to introduce W:

$$W_{i\eta,\lambda+1/2}(z) = \frac{\pi}{\sin\pi(2\lambda + 1)} \left\{\frac{-M_{i\eta,\lambda+1/2}(z)}{\Gamma(-\lambda - i\eta)} + \frac{M_{i\eta,-\lambda-1/2}(z)}{\Gamma(\lambda + 1 - i\eta)}\right\}. \tag{2.25}$$

The wave function $\phi_{\hat{\rho}}(kr)$ is normalized in the k scale.

The time-dependent wave function $\psi^{\hat{\rho}}(\mathbf{r},t)$ is written as

$$\psi^{\hat{\rho}}(\mathbf{r},t) = \psi^{\hat{\rho}}(\mathbf{r}) e^{-i\epsilon_{\hat{\rho}} Et}, \tag{2.26}$$

$$\epsilon_{\hat{\rho}} = 1, \quad \hat{\rho} = 1; \quad \epsilon_{\hat{\rho}} = -1, \quad \hat{\rho} = 2. \tag{2.27}$$

The radial Green's function $G_\kappa^{ij}(r_2 r_1 z)$ is now obtained by the eigenfunction expansion method. Before proceeding, let us clarify the meaning of the superscripts i, j. They take on four values: 1 1, 1 2, 2 1, and 2 2. In general, they refer to the "large" and "small" components, corresponding to $\psi(\tilde{\omega})$ and $\psi(-\tilde{\omega})$, respectively. In accordance with our previous papers, we have chosen to use 2 for the "large" component and 1 for the "small" component. However, this is purely a matter of convention. In what follows, we shall omit the superscripts i, j, since it is obvious that each component of the Green's function follows the corresponding component of the wave function. Thus we have

$$G_\kappa(r_2 r_1 z) = \sum_{\hat{\rho}=1}^2 \epsilon_{\hat{\rho}} \int_0^\infty dk\, \frac{\psi^{\hat{\rho}}(kr_2)\psi^{\hat{\rho}*}(kr_1)}{z - \epsilon_{\hat{\rho}} E} + \sum_{n_r} \frac{\psi_{n_r\kappa}(r_2)\psi^*_{n_r\kappa}(r_1)}{z - E_{n,\kappa}}. \tag{2.28}$$

The Green's function is obtained by doing the integration over the continuous spectrum. We follow basically the work of Hostler,[4] who did the calculation in the Klein–Gordon case. Define

$$J_\kappa = \sum_{\hat{\rho}=1}^2 \epsilon_{\hat{\rho}} \int_0^\infty dk\, \frac{\phi^{\hat{\rho}}(kr_2)\phi^{\hat{\rho}*}(kr_1)}{z - \epsilon_{\hat{\rho}} E}. \tag{2.29}$$

Use the identity

$$M_{i\eta,\lambda+1/2}(z)=\frac{e^{\pm\pi ik}W_{-i\eta,\lambda+1/2}(ze^{\pm\pi i})}{\Gamma(\lambda+1-i\eta)}+\frac{e^{\pm\pi i(i\eta-\lambda-1)}W_{i\eta,\lambda+1/2}(z)}{\Gamma(\lambda+1+i\eta)}. \tag{2.30}$$

We get

$$\begin{aligned}J_\kappa=\frac{m^2}{4\pi r_1r_2}\frac{\gamma}{\kappa}\Bigg[&\int_0^\infty\frac{dk}{E^3}\frac{\Gamma(1+\lambda-i\eta)W_{i\eta,\lambda+1/2}(-2ikr_2)e^{-\pi i(\lambda+1)}M_{-i\eta,\lambda+1/2}(2ikr_1)}{z-E}\\&+\int_0^\infty\frac{dk}{E^3}\Gamma(1+\lambda+i\eta)\frac{W_{-i\eta,\lambda+1/2}(2ikr_2)M_{-i\eta,\lambda+1/2}(2ikr_1)}{z-E}\\&-\int_0^\infty\frac{dk}{E^3}\Gamma(1+\lambda+i\eta)\frac{W_{-i\eta,\lambda+1/2}(-2ikr_2)M_{-i\eta,\lambda+1/2}(-2ikr_1)}{z+E}\\&-\int_0^\infty\frac{dk}{E^3}\Gamma(1+\lambda-i\eta)\frac{W_{i\eta,\lambda+1/2}(2ikr_2)e^{\pi i(\lambda+1)}M_{-i\eta,\lambda+1/2}(-2ikr_1)}{z+E}\Bigg],\end{aligned} \tag{2.31}$$

where $E=+(k^2+m^2)^{1/2}$.

Using the identity

$$M_{-i\eta,\lambda+1/2}(z)=e^{\mp\pi i(\lambda+1)}M_{i\eta,\lambda+1/2}(ze^{\pm\pi i}), \tag{2.32}$$

the phase factors $e^{\pm\pi i(\lambda+1)}$ occurring in the first and fourth terms of (2.31) can be absorbed in the M functions. Changing the integration variable from k to $E=+(k^2+m^2)^{1/2}$ in the first two integrals and to $E=-(k^2+m^2)^{1/2}$ in the last two, we obtain

$$\begin{aligned}J_\kappa=\frac{m^2}{4\pi r_1r_2}\frac{\gamma}{\kappa}\Bigg[&\int_\infty^m\frac{dE}{kE^2}\frac{\Gamma(1+\lambda-i\eta)W_{i\eta,\lambda+1/2}(-2ikr_2)M_{i\eta,\lambda+1/2}(-2ikr_1)}{E-z}\\&-\int_m^\infty\frac{dE}{kE^2}\frac{\Gamma(1+\lambda+i\eta)W_{-i\eta,\lambda+1/2}(2ikr_2)M_{-i\eta,\lambda+1/2}(2ikr_1)}{E-z}\\&+\int_{-\infty}^{-m}\frac{dE}{kE^2}\frac{\Gamma(1+\lambda-i\eta)W_{i\eta,\lambda+1/2}(-2ikr_2)M_{i\eta,\lambda+1/2}(-2ikr_1)}{E-z}\\&-\int_{-m}^{-\infty}\frac{dE}{kE^2}\frac{\Gamma(1+\lambda+i\eta)W_{-i\eta,\lambda+1/2}(2ikr_2)M_{-i\eta,\lambda+1/2}(2ikr_1)}{E-z}\Bigg].\end{aligned} \tag{2.33}$$

We now define k for general values of E on the complex plane less the two branch cuts $-\infty<E\leqslant-m$ and $m\leqslant E<\infty$. This definition is

$$k=(E^2-m^2)^{1/2},\quad 0<\operatorname{arc}(k)<\pi. \tag{2.34}$$

It is seen that $\operatorname{Im}(k)>0$ for all E on the cut plane. The integrands in (2.33) are then reduced to a single function

$$\frac{m^2}{4\pi r_1r_2}\frac{\gamma}{\kappa}\frac{\Gamma(1+\lambda-i\eta)W_{i\eta,\lambda+1/2}(-2ikr_2)M_{i\eta,\lambda+1/2}(-2ikr_1)}{kE^2(E-z)}, \tag{2.35}$$

when evaluated above or below the positive or negative energy branch cut. Because of the behavior of M, which is regular at the origin, and W, which is regular at infinity, the contour may be closed by semicircles "at infinity" in the upper and lower half-planes as long as $r_2>r_1$:

$$J_\kappa=\frac{m^2\gamma}{4\pi\kappa}\oint\frac{dE}{r_1r_2}\frac{\Gamma(1+\lambda-i\eta)W_{i\eta,\lambda+1/2}(-2ikr_2)M_{i\eta,\lambda+1/2}(-2ikr_1)}{kE^2(E-z)}. \tag{2.36}$$

The contour is now a closed loop enclosing the entire cut plane in the clockwise sense. (The original contour circles the disjoint branch cuts in the counterclockwise sense.) The poles of the integrand are the poles of the gamma function $\Gamma(1+\lambda-i\eta)$, the pole at $E=z$, and the double pole at $E=0$. The poles of the gamma function occur at precisely the bound state energy levels. The residues at the poles of the gamma function will be evaluated shortly. It is then found that this contribution cancels exactly the eigenfunction expansion sum coming from the bound states, i.e., the second term on the right of (2.28).

The residue of the gamma function is calculated as follows[13,14]:

$$\begin{aligned}&\operatorname*{Res}_{E=E_i}\Gamma(1+\lambda-i\eta)\\&\quad=\operatorname*{Res}_{E=E_i}\Gamma(-n_r-iZe^2(E/\mu-E_i/\mu_i))\\&\quad=(-1)^{n_r+1}\left\{iZe^2n_r!\frac{\partial}{\partial E}\left[\frac{E}{(E^2-m^2)^{1/2}}\right]_{E=E_i}\right\}^{-1}\\&\quad=(-1)^{n_r+1}\mu_i^3/Ze^2m^2n_r!\end{aligned} \tag{2.37}$$

When (2.37) is multiplied by $(-2\pi i)$, we find that it cancels exactly the bound state wave function term in (2.28). Since the bound state wave functions have been normalized exactly, we know that, because of this cancellation, the continuum wave function (2.22) must have been normalized correctly. In fact, we present this result as one of the ways of normalizing the continuum wave function. The double pole at the origin does not contribute anything. This can be seen as follows. First physically, the pole at the origin corresponds to $E=0$, and therefore there is no electron. Mathematically, one can argue as follows. Write the integrand, other than the $1/E^2$ term, as $f(k)$. Then the residue of the double pole is obtained by calculating

$$\left[\frac{df(k)}{dk}\frac{dk}{dE}\right]_{E=0} = \left[\frac{df(k)}{dk}\frac{E}{k}\right]_{E=0} = 0.$$

Finally, we obtain the coordinate space DCGF as

$$G^{ij}_{\kappa}(r_2\, r_1 z) = f(ij)\frac{m^2\gamma}{2i\kappa}\frac{1}{r_1\, r_2(z^2-m^2)^{1/2}z^2}\Gamma(1+\lambda-i\eta) \times W_{i\eta,\lambda+1/2}(-2ikr_2)M_{i\eta,\lambda+1/2}(-2ikr_1), \quad (2.38)$$

$$r_2 > r_1, \quad \eta = Ze^2z/k, \quad k=(z^2-m^2)^{1/2}, \quad \mathrm{Im}(k)>0, \quad (2.39)$$

where $f(ij)$ is the factor corresponding to $(z\kappa/\gamma+m)^{1/2}$ for $i,j=2$ and $i\tilde{\omega}(z\kappa/\gamma-m)^{1/2}$ for $i,j=1$. The appropriate λ in (2.38) is a function of $\tilde{\omega}$. For $i,j=1$, λ is $\lambda(-\tilde{\omega})$, and for $i,j=2$, $\lambda=\lambda(\tilde{\omega})$.

The connection between our Green's function and the nonrelativistic one is quite obvious. First the spin angular eigenfunction χ^{μ}_{κ} obviously reduces to the Schrödinger case if the spin is neglected, since $\mathbf{J}=\mathbf{L}+\boldsymbol{\sigma}/2$. Explicitly, one can see that the first term in (2.7) summed over $\tilde{\omega}$ is equal to $(2l+1)P_l(\xi)/4\pi$, which corresponds to the Schrödinger result. Therefore the second term in (2.7) can be considered as the relativistic correction due to the spin. In the final section of this paper, we shall show that the second term gives rise to the correction term over the nonrelativistic result in Rayleigh scattering. Furthermore, this correction is found to correspond to the Furry approximation.

We can therefore summarize our results roughly as follows.

Relativistic result: λ noninteger, spin present.

Furry approximation: replace λ by l, spin present.

Nonrelativistic or Schrödinger result: replace λ by l, put spin equal to zero.

Aside from these factors mentioned above, there are only minor factors which are peculiar to the relativistic case and cannot be found in the nonrelativistic case. The first is the fact that the Dirac equation is a first-order coupled differential equation, hence the factor $\gamma/2E\kappa$ in the normalization. The other factor is that in the relativistic case, the total energy E is considered, where $E^2=k^2+m^2$, while in the nonrelativistic case the kinetic energy E is considered, where $E=k^2/2m$. It is well known that the bound state energy spectrum between the Dirac equation and the Schrödinger equation differs by terms of the order $Z^4\alpha^4$. All these questions will be further discussed in Sec. V.

III. MOMENTUM SPACE DIRAC COULOMB GREEN'S FUNCTION

We now derive the momentum space DCGF by Fourier transforming the coordinate space DCGF obtained in the previous section. Thus

$$G(\mathbf{p}_2,\mathbf{p}_1,z) = (2\pi)^{-3}\int\int G(\mathbf{r}_2,\mathbf{r}_1,z)e^{-i\mathbf{p}_2\cdot\mathbf{r}_2+i\mathbf{p}_1\cdot\mathbf{r}_1}\,d\mathbf{r}_1\,d\mathbf{r}_2. \quad (3.1)$$

In spherical coordinates, $\mathbf{p}_2$ and $\mathbf{p}_1$ are expressed in terms of (p_2,θ_2,ϕ_2) and (p_1,θ_1,ϕ_1), respectively. The angular integrals can then be done immediately. We shall summarize the result in the form of a theorem. This result, though obviously known to many authors for a long time, has nevertheless not been stated concisely nor sufficiently appreciated. Thus we shall present this result as a theorem and give a simple proof.

Theorem: As far as the angular variables are concerned, the Fourier transform (say, from $\mathbf{r}$ to $\mathbf{p}$) of a spherical harmonic $Y^m_l(\theta_r,\phi_r)$ is equal to the same spherical harmonic $Y^m_l(\theta_p,\phi_p)$ multiplied by a definite factor (of r). Explicitly

$$\int_{-1}^{1} d\cos\theta_r \int_0^{2\pi} d\phi_r\, Y^m_l(\theta_r,\phi_r)e^{\pm i\mathbf{p}\cdot\mathbf{r}} = Y^m_l(\theta_p,\phi_p)\left[(2\pi)^{3/2}(\pm i)^l(rp)^{-1/2}J_{l+1/2}(rp)\right]. \quad (3.2)$$

This result was obtained by Podolsky and Pauling[15] by explicitly integrating out the left side of (3.2). We shall, however, give a simple proof of Eq. (3.2) based on the following three relations. The first one is an expansion due to Bauer, according to Watson,[16]

$$e^{\pm ipr\cos\omega} = \sum_{L=0}^{\infty}(\pm i)^L(2L+1)(\pi/2rp)^{1/2} \times J_{L+1/2}(rp)P_L(\cos\omega). \quad (3.3)$$

The second one is the spherical harmonic addition theorem[17]

$$P_L(\cos\omega) = \frac{4\pi}{2L+1}\sum_m Y^{m*}_L(\theta_r\,\phi_r)\,Y^m_L(\theta_p\,\phi_p). \quad (3.4)$$

The third one is the orthonormality of the spherical harmonics[18]

$$\int_0^{2\pi}\int_0^{\pi} Y^{m*}_L(\theta_r\,\phi_r)\,Y^{m'}_{L'}(\theta_r\,\phi_r)\sin\theta_r\,d\theta_r\,d\phi_r = \delta_{LL'}\,\delta_{mm'}. \quad (3.5)$$

Applying Eqs. (3.3), (3.4), and (3.5) successively to the left side of (3.2), we get the right side of (3.2).

Since the spin angular momentum eigenfunctions χ^{μ}_{κ} contain only spherical harmonics, we can immediately apply the theorem and obtain the result

$$\int_{-1}^{1} d\cos\theta_r\int_0^{2\pi} d\phi_r\, e^{\pm i\mathbf{p}_i\cdot\mathbf{r}_i}\chi^{\mu}_{\kappa}(\theta_r\,\phi_r) = \chi^{\mu}_{\kappa}(\theta_p\,\phi_p)\left[(2\pi)^{3/2}(\pm i)^l(r_i\,p_i)^{-1/2}J_{l+1/2}(r_i\,p_i)\right], \quad (3.6)$$

where $\pm i$ refer to $r_1\,p_1$ and $r_2\,p_2$, respectively. Thus the angular integration is done. In particular, the angular dependence of $G(\mathbf{p}_2,\mathbf{p}_1,z)$ is entirely contained in the matrix

$$\sum_{\kappa}\begin{bmatrix}\pi_{-\kappa}(\mathbf{p}_2,\mathbf{p}_1) & i\boldsymbol{\sigma}\cdot\mathbf{p}_2\,\pi_{\kappa}(\mathbf{p}_2\mathbf{p}_1)\\ -i\boldsymbol{\sigma}\cdot\mathbf{p}_2\,\pi_{-\kappa}(\mathbf{p}_2\mathbf{p}_1) & \pi_{\kappa}(\mathbf{p}_2\mathbf{p}_1)\end{bmatrix}. \quad (3.7)$$

The "radial" Green's function in momentum space $G^{ij}_{\kappa}(p_2\,p_1 z)$ is now

$$G_\kappa^{ij}(p_2 p_1 z) = \int\int r_2^2\, dr_2\, r_1^2\, dr_1 (r_1 r_2 p_1 p_2)^{-1/2} J_{l+1/2}(r_1 p_1) J_{l+1/2}(r_2 p_2)$$

$$\times f(ij) \frac{m^2\gamma}{2i2\kappa r_1 r_2 (z^2-m^2)^{1/2} z^2} \Gamma(1+\lambda-i\eta)[W_{i\eta,\lambda+1/2}(-2ikr_2) M_{i\eta,\lambda+1/2}(-2ikr_1)\,\theta(r_2-r_1)$$

$$+ W_{i\eta,\lambda+1/2}(-2ikr_1) M_{i\eta,\lambda+1/2}(-2ikr_2)\,\theta(r_1-r_2)]. \tag{3.8}$$

It is only necessary to evaluate one of the terms in the bracket, say, $\theta(r_1 - r_2)$. The other term can be done in a similar way. Thus we wish to evaluate the integral I,

$$I = \int_0^\infty dr_1\, r_1^{3/2} J_{l+1/2}(r_1 p_1) W_{i\eta,\lambda+1/2}(-2ikr_1) \int_0^{r_1} dr_2\, r_2^{3/2} J_{l+1/2}(r_2 p_2)(2r_2 p_2)^\lambda e^{-i\mu r_2}\, {}_1F_1(\lambda+1+i\eta,\ 2\lambda+2,\ 2i\mu r_2), \tag{3.9}$$

where

$$\mu = (z^2 - m^2)^{1/2}, \tag{3.10}$$

$$\eta = Ze^2 z/(z^2 - m^2)^{1/2}. \tag{3.11}$$

The Bessel function can be changed into a confluent hypergeometric function[19]

$$J_{l+1/2}(r_2 p_2) = \frac{2^{-l-1/2}}{\Gamma(l+\frac{3}{2})} (r_2 p_2)^{l+1/2} e^{-ir_2 p_2}\, {}_1F_1(l+1,\ 2l+2,\ 2ip_2 r_2). \tag{3.12}$$

The integral over r_2 in (3.9) is similar to the one considered by Ogata and Asai[5] in connection with finite nuclear size effects. We shall briefly outline the procedures.

First, we distinguish between two cases (1) $p_2 < \mu$ and (2) $p_2 > \mu$. In case 1, the ${}_1F_1$ functions are expressed in terms of the integral representation, Eq. (A6) of Ref. 5. We associate u with the Bessel function and v with the Dirac wave function. Then the integral over r_2 can be done, resulting in an incomplete gamma function, which in turn can be expressed in terms of ${}_1F_1$, Eq. (6) of Ref. 5. This ${}_1F_1$ function is now expanded by means of a multiplication theorem, Eq. (A12) of Ref. 5. The terms containing $[1-(2\mu/(p_2+\mu-2p_2u)v]$ are then separated out and expanded in terms of the binomial theorem. Then the integral over v can be performed, resulting in an ${}_2F_1$ function with argument $2\mu/(p_2+\mu-2p_2u)$. The argument of this ${}_2F_1$ function is converted into its inverse by means of Kummer's relations, Eq. (A11) of Ref. 5. Then the ${}_2F_1$ functions are expanded according to its definition

$${}_2F_1(a,b,c,z) = \sum_n \frac{(a)_n\,(b)_n\, z^n}{(c)_n\, n!}. \tag{3.13}$$

Finally, the integral over u can be performed, resulting in another ${}_2F_1$ function. Thus from Eq. (3.9) we wish to evaluate the integral $\mathscr{I}_2$, where

$$\mathscr{I}_2 = \int_0^{r_1} dr_2\, r_2^{\alpha-1} e^{-i(\mu+p_2)r_2}\, {}_1F_1(a,b;2ip_2 r_2)\, {}_1F_1(\bar a,\bar b;2i\mu r_2), \tag{3.14}$$

where

$$\alpha = \lambda + l + 3,\quad a = l+1,\quad b = 2l+2,\quad \bar a = \lambda + 1 + i\eta,\quad \bar b = 2\lambda + 2. \tag{3.15}$$

The final result is the following.

Case 1: $p_2 < \mu$,

$$\mathscr{I}_2 = r_1^\alpha \alpha^{-1} \sum_{q=0}^\infty \sum_{n=0}^\infty \sum_{m=0}^n \Gamma\begin{bmatrix}\alpha+n\\ \alpha\end{bmatrix} \frac{1}{n!}(-1)^m \binom{n}{m} (-i)^{\alpha+m} \Gamma\begin{bmatrix}\bar a, \bar b - \bar a\\ \bar b\end{bmatrix}$$

$$\times\, {}_1F_1(\alpha+n, 1+\alpha, -r_1) \left\{ \Gamma\begin{bmatrix} b, \alpha+m-a\\ \alpha+m, b-a\end{bmatrix} \left(\frac{p_2+\mu}{-2\mu}\right)^{\bar a} (p_2+\mu)^{-(\alpha+m)} \frac{(\bar a)_q (\bar a - \bar b + 1)_q}{(\bar a - \alpha - m + 1)_q\, q!} \left(\frac{p_2+\mu}{2\mu}\right)^q \right.$$

$$\times\, \Gamma\begin{bmatrix} b-a, a\\ b\end{bmatrix} {}_2F_1\!\left(a, \alpha+m-a-q; b, \frac{2p_2}{p_2+\mu}\right) + \Gamma\begin{bmatrix}\bar b, a - \bar a - m\\ \bar a, \bar b - \alpha - m\end{bmatrix} (-2\mu)^{-(\alpha+m)}$$

$$\left. \times \frac{(\alpha+m)_q(\alpha+m-\bar b+1)_q}{(\alpha+m-\bar a+1)_q\, q!} \left(\frac{p_2+\mu}{2\mu}\right)^q \Gamma\begin{bmatrix} b-a, a\\ b\end{bmatrix} {}_2F_1\!\left(a, -q; b, \frac{2p_2}{p_2+\mu}\right)\right\}. \tag{3.16}$$

Case 2: $p_2 > \mu$. This case is evaluated in the same way as case 1, except that the integration over u is carried out before integration over v. The result for case 2 also looks very similar to case 1:

$$\mathcal{J}_2 = r_1^a \alpha^{-1} \sum_{q=0}^{\infty} \sum_{n=0}^{\infty} \sum_{m=0}^{n} \Gamma\begin{bmatrix} \alpha+n \\ \alpha \end{bmatrix} \frac{1}{n!}(-1)^m \binom{n}{m} \quad (-i)^{\alpha+m} \Gamma\begin{bmatrix} a, b-a \\ b \end{bmatrix}$$

$$\times {}_1F_1(\alpha+n,1+\alpha,-r_1) \left\{ \Gamma\begin{bmatrix} b,\alpha+m-a \\ \alpha+m,b-a \end{bmatrix} \left(\frac{p_2+\mu}{-2p_2}\right)^a (p_2+\mu)^{-(\alpha+m)} \frac{(a)_q(a-b+1)_q}{(a-\alpha-m+1)_q\, q!} \left(\frac{p_2+\mu}{2\mu}\right)^q \right.$$

$$\times \Gamma\begin{bmatrix} \bar{b}-\bar{a},\bar{a} \\ \bar{b} \end{bmatrix} {}_2F_1\left(\bar{a},\alpha+m-\bar{a}-q;\bar{b},\frac{2\mu}{p_2+\mu}\right) + \Gamma\begin{bmatrix} b,a-\alpha-m \\ a,b-\alpha-m \end{bmatrix} (-2p_2)^{-(\alpha+m)}$$

$$\left. \times \frac{(\alpha+m)_q(\alpha+m-b+1)_q}{(\alpha+m-a+1)_q\, q!} \left(\frac{p_2+\mu}{2p_2}\right)^q \Gamma\begin{bmatrix} \bar{b}-\bar{a},\bar{a} \\ \bar{b} \end{bmatrix} {}_2F_1\left(\bar{a}-q;\bar{b},\frac{2\mu}{p_2+\mu}\right) \right\}. \tag{3.17}$$

The integral over r_1 is basically of the following form:

$$\int_0^\infty dr_1\, r_1^a\, e^{-2ir_1p_1}\, {}_1F_1(1)\, {}_1F_1(2)\, {}_1F_1(3). \tag{3.18}$$

Here, ${}_1F_1(1)$ comes from the Bessel function and is the same all the time, ${}_1F_1(2)$ comes from the incomplete gamma function and is also the same all the time, and ${}_1F_1(3)$ contains two types, ${}_1F_1(p,q,t)$ and $G(p,q,t)$, where

$$G(p,q,t) = t^{1-q} \frac{\Gamma(p+1-q)\Gamma(q)}{\Gamma(2-q)\Gamma(p)}\, {}_1F_1(p+1-q,2-q,t). \tag{3.19}$$

Thus we are required to evaluate the following integral $\mathcal{J}_1$:

$$\mathcal{J}_1 = \int_0^\infty dr_1\, r_1^{\beta_i}\, e^{-2ip_1r_1}\, {}_1F_1(l+1,2l+1,2ir_1p_1) \times {}_1F_1(\alpha+n,1+\alpha,-r_1)\, {}_1F_1(a_i,b_i,2ir_1p_1), \tag{3.20}$$

where $i=1,2$, and $i=1$ refers to the Whittaker function M, $i=2$ refers to the Whittaker function W,

$$\beta_1 = 2\lambda+2l+n+5, \quad \alpha=\lambda+l+3, \quad \beta_2 = 2l+n+4,$$
$$a_1 = \lambda+1+i\eta, \quad b_1 = 2\lambda+2, \tag{3.21}$$
$$a_2 = -\lambda+i\eta, \quad b_2 = -2\lambda.$$

Equation (3.20) is of the form given on p. 216 of Ref. 6. We shall just give the result for $i=1$ in Eq. (3.20)

$$\mathcal{J}_1 = (2ip_1)^{l+\lambda+2}(1+2ip_1)^{-2\lambda-2l-n-6} \times \Gamma(2\lambda+2l+n+6) \times F_A\left(2\lambda+2l+n+6; l+1,\lambda+l+n+3, \lambda+1-i\eta; 2l+2,\lambda+l+4,2\lambda+2: \frac{2ip_1}{1+2ip_1}, \frac{1}{1+2ip_1}, \frac{2ip_1}{1+2ip_1}\right). \tag{3.22}$$

Thus we have obtained the DCGF in momentum space.

IV. RELATIVISTIC RAYLEIGH SCATTERING

In this section we present the matrix element for relativistic Rayleigh scattering, or elastic Compton scattering, in analytically closed form.

The method used is as follows. The matrix element for Rayleigh scattering is written basically in terms of the coordinate space Dirac Coulomb Green's function, which has in fact been given by Brown and Schaefer.[3] The coordinate space Green's function has been obtained by us in Sec. II. It remains then to evaluate the integral over the intermediate states. The integral is evaluated according to the technique used in Sec. III for the calculation of the momentum space Green's function. Basically, the angular parts are written in terms of spherical harmonics. The Fourier integral over the angular variables can then be performed. Then only the radial integrals are left. These are of two different kinds: a finite integral over, say, r_2, with limits going from zero to r_1, and an integral over r_1 with limits from zero to infinity. Both these integrals can be obtained in closed form.

Using the ground state $1S_{1/2}$ as the initial state, we find that the integrals involved are very similar to the ones encountered in Sec. III in the evaluation of the momentum space Green's function. It is not hard to generalize our results to arbitrary initial states, since the technique we have developed can be generalized to more complicated structures. Moreover, it is also possible to obtain matrix elements for inelastic Compton scattering for relativistically bound electrons, since the same technique applies.

The basic diagrams for Rayleigh scattering are shown in Fig. 1, where the double solid lines represent the relativistic electron in a Coulomb field, and the dotted lines represent the photon.

For case (a) we have to evaluate the matrix element

$$M_{(a)} = \int d^4x_2 \int d^4x_1\, \bar{\psi}_i(x_2)\, \gamma_\mu\, A_\mu^{(+)}(x_2) \times S_F^{(v)}(x_2\,x_1)\, \gamma_\nu\, A_\nu^{(-)}(x_1)\, \psi_i(x_1). \tag{4.1}$$

For case (b) we have

$$M_{(b)} = \int d^4x_2 \int d^4x_1\, \bar{\psi}_i(x_2)\, \gamma_\mu\, A_\mu^{(-)}(x_2) \times S_F^{(v)}(x_2\,x_1)\, \gamma_\nu\, A_\nu^{(+)}(x_1)\, \psi_i(x_1). \tag{4.2}$$

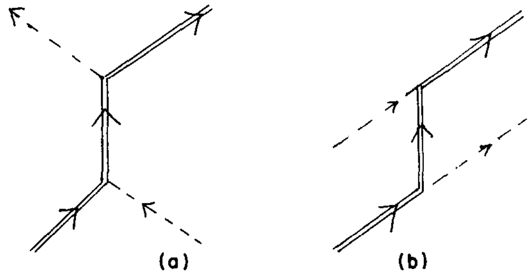


FIG. 1. Diagrams for Rayleigh scattering.

We use the fact that $\psi_i(x) = \psi_i(\mathbf{r})\, e^{-iE_it}$ and that $A_\nu^{(\pm)}(x) = A_\nu(\mathbf{r})\, e^{\pm i\omega t}$, where ω is the energy of the photon

$$S_F^{(\nu)}(x_2\, x_1) = (i/\pi)\int_{-\infty}^{\infty} G(\mathbf{r}_2\mathbf{r}_1\, E)\beta e^{-iEt}\, dE, \tag{4.3}$$

where $G(\mathbf{r}_2\mathbf{r}_1\, E)$ is the Dirac Coulomb Green's function in coordinate space. Integrating over t_2, t_1, and E we have

$$M_{(a)} = 4\pi i\delta(\omega - \omega')\int d\mathbf{r}_2 \int d\mathbf{r}_1\, \bar{\psi}_i(r_2)\, \gamma_\mu\, A_\mu(\mathbf{r}_2) \times G(\mathbf{r}_2\,\mathbf{r}_1\, E_i + \omega)\, \gamma_4\, \gamma_\nu\, A_\nu(\mathbf{r}_1)\, \psi_i(\mathbf{r}_1), \tag{4.4}$$

$$M_{(b)} = 4\pi i\delta(\omega - \omega')\int d\mathbf{r}_2 \int d\mathbf{r}_1\, \bar{\psi}_i(\mathbf{r}_2)\, \gamma_\mu\, A_\mu(\mathbf{r}_2) \times G(\mathbf{r}_2\,\mathbf{r}_1\, E_i - \omega)\, \gamma_4\, \gamma_\nu\, A_\nu(\mathbf{r}_1)\psi_i(\mathbf{r}_1). \tag{4.5}$$

Henceforth, we shall confine our attention to case (a), since case (b) can obviously be treated in a similar way.

We can now insert the Green's function obtained in Sec. II into (4.4). If the initial state is chosen to be the $1S_{1/2}$ ground state, $n_r = 0, j = \frac{1}{2}, \tilde{\omega} = -1$, then only G_κ^{22} will enter into the calculation. Thus we shall consider the integral

$$\begin{aligned} M_a = {} & \delta(\omega - \omega')(-4\pi i)\sum_{j,\tilde{\omega}} \int_0^\infty r_1^2\, dr_1 \int d\Omega_1\, \psi_i(r_1)\, \chi_{\kappa_i}^{\mu_i}(\Omega_1)\, \frac{(m + z\kappa/\gamma)m^2\gamma\Gamma(1 + \lambda - i\eta)}{r_1\, r_2 4i\kappa z^2(z^2 - m^2)^{1/2}} \\ & \times \sum_\mu S_2\, \chi_\kappa^\mu(\Omega_2)\, \chi_\kappa^{\mu +}(\Omega_1) S_1 \gamma_4\, \gamma_\mu\, e'_\mu\, e^{-i\mathbf{k}\cdot\mathbf{r}_1}\, \psi_\kappa^\infty(r_1) \int_0^{r_1} r_2^2\, dr_2 \int d\Omega_2 \\ & \times \psi_\kappa^0(r_2)\, \chi_{\kappa_i}^{\mu_i +}(\Omega_2) e^{i\mathbf{k}\cdot\mathbf{r}_2}\gamma_4\, \gamma_\nu\, e_\nu\, \psi_i(r_2), \end{aligned} \tag{4.6}$$

where $\psi_\kappa^0(r_2)$ and $\psi_\kappa^\infty(r_1)$ correspond to the Whittaker functions M and W, respectively. We shall now evaluate the angular integral.

First we write

$$\gamma_\nu = \gamma_4[x_\nu, H]. \tag{4.7}$$

Then

$$\gamma_\nu\, e_\nu = \gamma_4\, \omega x_\nu\, e_\nu, \tag{4.8}$$

$$\gamma_\nu\, e'_\mu = -\gamma_4\, \omega x_\mu\, e'_\mu. \tag{4.9}$$

Next we sum over the polarization vector of the photon and average over the position vector of the electron, obtaining for each term $x_\nu\, e_\nu$ a term $2x/3$, which can be expressed in terms of spherical harmonics. The spherical harmonic from χ_κ^μ can be coupled with the position vector by the usual angular momentum coupling technique to obtain another spherical harmonic. The spin eigenfunction is

$$\chi_\kappa^\mu = \begin{bmatrix} -\tilde{\omega}\left(\dfrac{\kappa + \frac{1}{2} - \mu}{2\kappa + 1}\right)^{1/2} & Y^{\mu - 1/2}_{|\kappa + 1/2| - 1/2} \\ \left(\dfrac{\kappa + \frac{1}{2} + \mu}{2\kappa + 1}\right)^{1/2} & Y^{\mu + 1/2}_{|\kappa + 1/2| - 1/2} \end{bmatrix}. \tag{4.10}$$

Denoting the final relevant spherical harmonic by $Y_l^m(\Omega_2)$ we can perform the angular integration over Ω_2, according to Eq. (3.6),

$$\int d\Omega_2\, e^{i\mathbf{k}\cdot\mathbf{r}_2}\, Y_l^m(\Omega_2) = Y_l^m(\hat{k})\, [(2\pi)^{3/2}\, i^l (kr_2)^{-1/2}\, J_{l+1/2}(kr_2)]. \tag{4.11}$$

The angular integration over Ω_1 can be performed in the same way. It remains for us to evaluate the radial integrals.

The integral over r_2 is of the following form:

$$\begin{aligned} \mathscr{I}_2 = {} & \int_0^{r_1} dr_2\, r_2^{2 + \lambda + l + (1 - Z^2e^4)^{1/2}}\, e^{-ir_2(\tilde{\mu} + k + s)}\, {}_1F_1(\lambda + 1 - i\eta, 2\lambda + 2, 2i\tilde{\mu}r_2) \\ & \times {}_1F_1(l + 1, 1l + 2, 1ikr_2), \end{aligned} \tag{4.12}$$

where

$$\tilde{\mu} = (E_k^2 - m^2)^{1/2}, \quad E_k = E_i + \omega, \quad s = -\tfrac{1}{2}im[2(1 - Z^2e^4)^{1/2} - 1 + Z^2e^4]^{1/2}, \quad s^2 = m^2 - E_i^2. \tag{4.13}$$

Before evaluating the integral, let us discuss the relations between the energies. We have

$$\tilde{\mu}^2 + |s|^2 = (E_i + \omega^2) - E_i^2 = 2E_i\omega + \omega^2 = 2E_ik + k^2. \tag{4.14}$$

Hence

$$\tilde{\mu}^2 + |s|^2 > k^2 \quad \text{or} \quad |\tilde{\mu} + s| > k. \tag{4.15}$$

Now the integral in (4.12) is of the same form as Eq. (3.14), except that (4.12) has an extra term e^{-ir_2s}. Equation (4.12) can be evaluated by the same technique as in Sec. III. However, because of (4.15) we only need to consider case 1, $|\tilde{\mu} + s| > k$. Let us define

$$\alpha = 3 + \lambda + l + (1 - Z^2e^4)^{1/2}, \quad a = l + 1, \quad b = 2l + 2, \quad \bar{a} = \lambda + 1 - i\eta, \quad \bar{b} = 2\lambda + 2. \tag{4.16}$$

Then

$$\begin{aligned}\mathscr{J}_2 &= \int_0^{r_1} dr_2\, r_2^{\alpha - 1}\, e^{-ir_2(\tilde{\mu} + k + s)}\, {}_1F_1(\bar{a},\bar{b},2i\tilde{\mu}r_2)\, {}_1F_1(a,b,2ikr_2) \\ &= r_1^\alpha \alpha^{-1} \sum_{q=0}^{\infty} \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \Gamma\begin{bmatrix}\alpha + n \\ \alpha\end{bmatrix} \frac{1}{n!}(-1)^m \binom{n}{m} (-i)^{\alpha + m} \Gamma\begin{bmatrix}\bar{a},\bar{b} - \bar{a} \\ \bar{b}\end{bmatrix} {}_1F_1(\alpha + n, 1 + \alpha, -r_1) \\ &\times \left\{ \Gamma\begin{bmatrix} b,\alpha + m - a \\ \alpha + m, b - a\end{bmatrix} (k + \tilde{\mu} + s)^{-\alpha - m + \bar{a}} (-2\tilde{\mu})^{-\bar{a}} \frac{(\bar{a})_q\, (\bar{a} - \bar{b} + 1)_q}{(\bar{a} - \alpha - m + 1)_q\, q!} \left(\frac{k + \tilde{\mu} + s}{2\tilde{\mu}}\right)^q \Gamma\begin{bmatrix}a,b - a \\ b\end{bmatrix} \right. \\ &\times {}_2F_1\left(a,\alpha + m - \bar{a} - q; b; \frac{2k}{k + s + \tilde{\mu}}\right) + \Gamma\begin{bmatrix}\bar{b},\bar{a} - \alpha - m \\ \bar{a},\bar{b} - \alpha - m\end{bmatrix} (-2\tilde{\mu})^{-(\alpha - m)} \frac{(\alpha + m)_q (\alpha + m - \bar{b} + 1)}{(\alpha + m - \bar{a} + 1)_q\, q!} \\ &\times \left. \left(\frac{k + s + \tilde{\mu}}{2\tilde{\mu}}\right)^q \Gamma\begin{bmatrix}a,b - a \\ b\end{bmatrix} {}_2F_1\left(a, -q; b, \frac{2k}{k + s + \tilde{\mu}}\right)\right\}.\end{aligned} \tag{4.17}$$

Case 2, where $|\tilde{\mu} + s| < k$, is applicable for the evaluation of M_b, Eq. (4.5), diagram (b) of Fig. 1.

The integral over r_1 is of the following form:

$$\mathscr{J}_1 = \int_0^\infty dr_1\, r_1^a\, e^{-ir_1(k' + \tilde{\mu} + s)}\, {}_1F_1(1)\, {}_1F_1(2)\, {}_1F_1(3), \tag{4.18}$$

where ${}_1F_1(1)$ comes from the Bessel function and is the same all the time, ${}_1F_1(2)$ comes from the incomplete gamma function and is also the same all the time, and ${}_1F_1(3)$ comes from $\psi_\kappa^\infty(r_1)$ and contains two confluent hypergeometric functions: ${}_1F_1(p,q,t)$ and $G(p,q,t)$, as defined in (3.19). All the cases are expressible in terms of the following integral $\mathscr{J}_1$:

$$\begin{aligned}\mathscr{J}_1 = \int_0^\infty dr_1\, r_1^{\beta_i}\, e^{-ir_1(k' + \tilde{\mu} + s)}\, {}_1F_1(l + 1, 2l + 2, 2ik' r_1) \\ \times {}_1F_1(\alpha + n, 1 + \alpha, -r_1)\, {}_1F_1(a_i, b_i, 2i\tilde{\mu}r_1),\end{aligned} \tag{4.19}$$

where $i = 1,2$, and $i = 1$ refers to ${}_1F_1(p,q,t)$ and $i = 2$ refers to $G(p,q,t)$

$$\begin{aligned}&\alpha = 3 + \lambda + l + (1 - Z^2e^4)^{1/2}, \\ &\beta_1 = \alpha + 2 + (1 - Z^2e^4)^{1/2} + l + \lambda, \\ &\beta_2 = \alpha + 1 + (1 - Z^2e^4)^{1/2} + l - \lambda, \\ &a_1 = \lambda + 1 - i\eta, \quad b_1 = 2\lambda + 2, \\ &a_2 = -\lambda - i\eta, \quad b_2 = -2\lambda.\end{aligned} \tag{4.20}$$

Equation (4.19) is of the form given on p. 216 of Ref. 6. We shall just give the result for $i = 1$ in Eq. (4.19)

$$\begin{aligned}\mathscr{J}_1 = (2ik')^{l+1} (2i\tilde{\mu})^{\lambda + 1}\, [1 + i(s + k' + \tilde{\mu})]^{-\nu - M} \\ \times \Gamma(\nu + M)\, F_A(\nu + M;\ l + 1, \lambda + l + n + 3, \\ \lambda + 1 - i\eta; \lambda + 1 - i\eta, 2l + 2, \\ \lambda + l + 4;\ 1/[1 + i(s + k' + \tilde{\mu})], \\ 2ik'/[1 + i(s + k' + \tilde{\mu})], \\ 2i\tilde{\mu}/[1 + i(s + k' + \tilde{\mu})]),\end{aligned} \tag{4.21}$$

where

$$\begin{aligned}&\nu - 1 = \beta_1 - (l + 1) - \tfrac{1}{2}(1 + \alpha) - \tfrac{1}{2} b_1, \\ &M = (3\lambda + 3l + 8)/2.\end{aligned} \tag{4.22}$$

V. COMPARISON OF THE RELATIVISTIC RESULT WITH THE NONRELATIVISTIC RESULT IN RAYLEIGH SCATTERING

In this section we make a comparison between the relativistic result and the nonrelativistic result in Rayleigh scattering. Our purpose is twofold. First, to understand the convergence of the sums in our formula, especially with regard to summation over j or equivalently l. Second, to see whether the nonrelativistic result is contained in the relativistic result. The conclusion we have obtained is the following. If one uses the ground state $1S_{1/2}$ as the initial state, then the relativistic sum should be convergent, and the nonrelativistic result is indeed contained in the relativistic result, to within an order of 0.01%. Moreover, in the general case, with an arbitrary initial bound state, the relativistic result contains a correction term, which is chiefly due to the spin of the electron, corresponding to the Furry approximation.[20] This correction term is of the order of the fine structure constant ($\frac{1}{137}$).

In order to compare the relativistic result with the nonrelativistic result, we shall make some approximations in our formula. The errors thus committed can be readily estimated, and are found to be very small. The approximations we make are the following.

(1) With regard to the Dirac Coulomb wave function, we shall replace λ by l. This commits an error of approximately 0.01%. This simplification reduces the radial part of the wave function and the Green's function to be almost identical with the nonrelativistic Schrödinger result, except for a scale factor $\mu_i = (m^2 - E_i^2)^{1/2}$ for the bound state wave function and $k = (z^2 - m^2)^{1/2}$ for the Green's function. In particular, if one chooses the ground state $1S_{1/2}$ to be the initial state, then the "small component" vanishes according to the Wong–Yeh solution and only G_κ^{22} will enter into the calculation.

(2) The second point to be examined is the interaction term. According to the relativistic formulation, the electromagnetic interaction term is basically $\boldsymbol{\alpha}\cdot\mathbf{A}$, while for the nonrelativistic formulation it is $\mathbf{p}\cdot\mathbf{A}$. However, from (4.8) and (4.9), we see that the first expression can be written basically as $\pm i\omega x_\nu e_\nu$, while the second expression can be written as $-im\omega x_\nu e_\nu$. Thus the interaction terms are of the same form.

Now when we compare the relativistic result with the nonrelativistic result in Rayleigh scattering, we find that the only difference arises from the summation over μ of the spin eigenfunctions in the Green's function, Eq. (2.7). Equation (2.7), for the relativistic case, contains two terms. However, the first term, when written out in terms of l, is exactly the same as the nonrelativistic result, i.e., $\Sigma_{l=0}^{\infty} [(2l+1)/4\pi] \times P_l(\cos\theta)$. Therefore, we conclude that the second term in (2.7), after the transformation S, given by Eq. (2), gives rise to the relativistic correction over the nonrelativistic result. This correction term adds

$$\frac{-iZe^2}{2l(l+1)} P_l'(\cos\theta)(1-\cos\theta)\boldsymbol{\alpha}\cdot(\mathbf{r}_1+\mathbf{r}_2) \tag{5.1}$$

to each P_l term for $l = 1,2,3,\ldots,\infty$, and is therefore of the order of $\alpha = \frac{1}{137}$. Furthermore, it has been shown by Hostler[4] that this correction term corresponds to the Furry approximation.

However, if the initial state is chosen to be the $1S_{1/2}$ state, then the correction term in (5.1) does not contribute, because $\boldsymbol{\alpha}$ has the matrix

$$\begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}$$

in ρ space and we have seen that only the G_κ^{22} term will contribute, which means that the correction term is zero in this case. Thus we conclude that for the $1S_{1/2}$ ground state the nonrelativistic result in Rayleigh scattering and in the Lamb shift is accurate to within 0.01% when compared with the relativistic result.

For a general bound state, we conclude that the relativistic result converges, and does contain the nonrelativistic result as a first approximation. Moreover, we have identified the correction term with the Furry approximation, and found that it is of the order of the fine structure constant $(\frac{1}{137})$, and is basically due to the spin of the electron. A detailed calculation of the correction term both in the case of Rayleigh scattering and in the Lamb shift will be presented in a future publication.

[1]E. H. Wichmann and N. M. Kroll, Phys. Rev. **101**, 843 (1956).
[2]M. K. F. Wong and H. Y. Yeh, Phys. Rev. D **25**, 3396 (1982).
[3]G. E. Brown and G. W. Schaefer, Proc. R. Soc. London Ser. A **233**, 527 (1956).
[4]L. Hostler, J. Math. Phys. **5**, 591 (1964).
[5]H. Ogata and J. Asai, J. Math. Phys. **23**, 1968 (1982).
[6]A. Erdelyi, W. Magnus, F. Oberhettinger, and F. Tricomi, *Higher Transcendental Functions* (McGraw–Hill, New York, 1953), Vol. 1; and *Tables of Integral Transforms* (McGraw–Hill, New York, 1954), Vol. 1.
[7]P. J. Mohr, Ann. Phys. **88**, 26 (1974).
[8]M. Gyulassy, Nucl. Phys. A **244**, 497 (1975).
[9]D. J. Hylton, J. Math. Phys. **25**, 1125 (1984).
[10]C. G. Darwin, Proc. R. Soc. London A Ser. **118**, 654 (1928).
[11]H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-electron Atoms* (Springer, Berlin, 1957).
[12]M. E. Rose, *Relativistic Electron Theory* (Wiley, New York, 1961).
[13]P. C. Martin and R. J. Glauber, Phys. Rev. **109**, 1307 (1958).
[14]R. A. Mapleton, J. Math. Phys. **2**, 478 (1961).
[15]E. Podolsky and L. Pauling, Phys. Rev. **34**, 109 (1929).
[16]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U. P., Cambridge, England, 1922), p. 128.
[17]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, NJ, 1957), p. 63.
[18]Reference 17, p. 21.
[19]Reference 3, p. 265.
[20]W. H. Furry, Phys. Rev. **46**, 391 (1934).

The dynamical structure of gravitational theories with GL(4,R) connections

Marek Antonowicz
Institute of Theoretical Physics, Warsaw University, Hoża 69, 00-681 Warsaw, Poland
Wiktor Szczyrba
Institute of Mathematics, Polish Academy of Sciences, Śniadeckich 8, 00-950 Warsaw, Poland

(Received 29 October 1984; accepted for publication 28 December 1984)

We investigate here GL(4,R)-gauge theories of gravity based on variational principles. The components of tetrad fields $e_\mu^{(\alpha)}$, the components of metrics $g_{(\alpha)(\beta)}$, and the components of connections $\Gamma_\lambda{}^{(\alpha)}{}_{(\beta)}$ are taken as the gravitational potentials. Matter potentials are the components of GL(4,R)-tensor fields ϕ^Σ. We derive the conservation laws for a general theory, that is, the Belinfante–Rosenfeld and Bianchi identities, and find minimal systems of independent variational equations. The natural GL(4,R)-covariant Hamiltonian formulation of the theory induces a GL(3,R)-covariant Hamiltonian formulation related to a chosen slicing of space-time. The Hamiltonian field equations corresponding to this formulation describe the dynamics of the system. We determine 20 symplectic constraints, 20 gauge transformations, and 20 gauge variables generic for a general gravitational Lagrangian. As an example, we consider the GL(4,R)–Einstein theory in vacuum as well as in the presence of a vector field and find the complete canonical formulation in both cases.

I. INTRODUCTION

In the 1960's astronomical observations confirmed that the Einstein theory of gravity explains several questions concerning the structure of the universe at large and can also predict the existence of such nonclassical objects as quasars and black holes. On the other hand, after 30 or 40 years of active research in quantum theory it became clear that the gravitational field should play an essential role in any future consistent theory of fundamental interactions.[1] The problem of the "initial singularity" of the universe and the difficulties in the quantization of the Einstein theory of relativity[2] attracted the interest of physicists in other theories of gravity. Physicists started to investigate the classical structure of such theories. One possible direction is to consider theories based on alternative Lagrangians for the gravitational field. We should pose the question here: What kind of conceivable geometries can be taken into account for reasonable theories of gravity? We recall that in the Einstein theory space-time is endowed with the structure of a pseudo-Riemannian manifold, that is, connections on space-time have to be metric compatible and torsionless. One can relax these conditions and investigate spaces with torsion or even admit metric noncompatible connections. For gravitational theories with the presence of tensor fields the Palatini variational principle generally gives rise to non-Riemannian space-times.

If the variational variables are (holonomic) components $g_{\mu\nu}$ of a metric $\mathbf{g}$ and holonomic components $\Gamma_\lambda{}^\mu{}_\nu$ of a connection Γ then we have two natural kinds of theories: (i) spaces with symmetric connection coefficients $\Gamma_\lambda{}^\mu{}_\nu$ (torsion equal zero),[3] and (ii) spaces with general affine connections $\Gamma_\lambda{}^\mu{}_\nu$ (with torsion and nonmetricity).[4,5] For both kinds of theories the metric $\mathbf{g}$ is generally incompatible with the connection Γ, that is, $\mathscr{D}_\lambda \mathbf{g}\neq 0$.

In recent years, however, another approach has been commonly accepted. According to Weyl[6] instead of a metric $\mathbf{g}$ on M we take four linearly independent fields of covectors $\mathbf{x}\to(\mathbf{e}^{(\alpha)}(\mathbf{x}))_{\alpha=0}^3$ on space-time. These fields are supposed to be *a priori* orthonormal, that is, $\mathbf{e}^{(\alpha)}\cdot\mathbf{e}^{(\beta)}=\eta^{(\alpha)(\beta)}$, where $\eta^{(\alpha)(\beta)}=\operatorname{diag}(-1,1,1,1)$ is the constant diagonal Minkowski metric. Such a structure is richer than the metric structure of M. It also gives us a spinor structure of space-time.[7] Fields of tetrads of covectors and their dual fields of tetrads of vectors $\mathbf{x}\to(\mathbf{e}_\alpha(\mathbf{x}))_{\alpha=0}^3$ are subject to an action of the local Lorentz group SO (3,1). If $\mathbf{x}\to[L^{(\alpha)}{}_{(\beta)}(\mathbf{x})]\in$SO (3,1) is an element of the local Lorentz group, that is, a field of special orthochronous Lorentz matrices on M, then the transformed tetrads are given by the following formulas:

$$'\mathbf{e}_{(\alpha)}=L^{(\beta)}{}_{(\alpha)}\mathbf{e}_{(\beta)},\quad '\mathbf{e}^{(\alpha)}=L^{-1(\alpha)}{}_{(\beta)}\mathbf{e}^{(\beta)}. \tag{1.1}$$

Besides its metric structure, space-time carries another structure that enables us to transport tensor fields in a parallel manner along curves in M. It is given by means of an $\mathfrak{so}(3,1)$-valued one-form $\Gamma^{(\alpha)}{}_{(\beta)}$. In local coordinates (x^λ) on M, $\Gamma^{(\alpha)}{}_{(\beta)}=\Gamma_\lambda{}^{(\alpha)}{}_{(\beta)}\,dx^\lambda$. The Lie algebra $\mathfrak{so}(3,1)$ of the Lorentz group SO(3,1) consists of Minkowski–skew-symmetric 4×4 real matrices, that is,

$$\Gamma_\lambda{}^{(\alpha)}{}_{(\epsilon)}\eta^{(\epsilon)(\beta)}=\Gamma_\lambda{}^{(\alpha)(\beta)}=-\Gamma_\lambda{}^{(\beta)(\alpha)}$$
$$=-\Gamma_\lambda{}^{(\beta)}{}_{(\epsilon)}\eta^{(\epsilon)(\alpha)}. \tag{1.2}$$

The action of the local Lorentz group in the space of connections is given by the following formula:

$$'\Gamma_\lambda{}^{(\alpha)(\beta)}=L^{-1(\alpha)}{}_{(\epsilon)}L^{-1(\beta)}{}_{(\tau)}\Gamma_\lambda{}^{(\epsilon)(\tau)}$$
$$-L^{(\epsilon)}{}_{(\tau)}\eta^{(\tau)(\beta)}\,\partial_\lambda L^{-1(\alpha)}{}_{(\epsilon)}. \tag{1.3}$$

For anholonomic components of a tensor field $\boldsymbol{\phi}$ we have

$$'\phi^{(\alpha_1)\cdots(\alpha_k)}{}_{(\beta_1)\cdots(\beta_s)}=L^{-1(\alpha_1)}{}_{(\mu_1)}\cdots L^{-1(\alpha_k)}{}_{(\mu_k)}$$
$$\times L^{(\nu_1)}{}_{(\beta_1)}\cdots L^{(\nu_s)}{}_{(\beta_s)}\phi^{(\mu_1)\cdots(\mu_k)}{}_{(\nu_1)\cdots(\nu_s)}. \tag{1.4}$$

By virtue of (1.3), (1.4), the covariant derivative $\mathscr{D}_\lambda\phi^\Sigma$ commutes with local Lorentz rotations (see Sec. II). It follows from (1.2) that the covariant derivative of the metric tensor $\boldsymbol{\eta}$ vanishes, that is,

$$\mathscr{D}_\lambda\eta_{(\alpha)(\beta)}=0. \tag{1.5}$$

 0022-2488/85/071711-17$02.50

We see that theories invariant with respect to the action (1.1), (1.3), and (1.4) of the local Lorentz group are metric compatible, that is to say, the parallel transport given by the connection $\boldsymbol{\Gamma}$ preserves the metric $\boldsymbol{\eta}$.

Theories under considerations are to be invariant with respect to the natural action of the diffeomorphism group of space-time. We say that the gauge group of these theories is the semidirect (semisimple) product of the local Lorentz group and the group of diffeomorphisms of space-time. The study of such theories was initiated by Sciama and Kibble[8] who, in the early 1960's generalized Utiyama's results[9] and rediscovered the Einstein–Cartan theory of gravity.

Gravitational theories formulated in the tetrad language are usually called "gauge theories of gravity." In the literature one can find several interpretations of that notion, either closer or further to the original Yang–Mills idea.[10] Especially important are (i) the approach based on Cartan connections in the bundle of affine frames on space-time[11–15]; (ii) the "Poincaré group approach" of Hehl and von der Heyde[16–18]; and (iii) methods of principal fiber bundles with more general structure groups [SO(3,1), Sp(4)].[19–22] Our approach is fairly close to that of (ii). As opposed to Refs. 16–18, however, we do not put the emphasis on the construction of the full gravitational gauge group, that is, a mathematical object containing local SO(3,1) [or GL(4,R)] rotations and the transformations given by diffeomorphisms of space-time. Nevertheless, as we show in Appendix F, such a construction exists but it differs from that presented in Refs. 16–18.

Investigations on the Einstein–Cartan–Sciama–Kibble (ECSK) theory were later continued by Hehl,[16] Trautman,[11] Nester,[23] and others.

In the seventies, physicists also paid attention to theories of gravity with gravitational Lagrangians different from R (the scalar curvature).[18,24–28] Sezgin and Nieuwenhuizen[29] investigated general gravitational Lagrangians quadratic in curvature and torsion, and selected such of them that could be interesting for quantum gravity. Nieh, Rauch, and other authors[26] investigated spherically symmetric solutions for gravitational Lagrangians quadratic in curvature and torsion and found the conditions when Birkhoff's theorem held. Baekler, Hehl, Mielke, McCrea, and others[28] presented several solutions for the theory based on the Hehl–von der Heyde Lagrangian and, thereby, showed its viability. The linearized version of theories with quadratic Lagrangians was investigated by Hayashi and Shirafuji.[29]

In the present paper we deal with a more general case, with theories of gravity in nonorthonormal tetrads and with metric noncompatible connections. The geometry of space-time M is given by four linearly independent fields of covectors $\mathbf{e}^{(\alpha)}$ on M, a symmetric metric tensor $g_{(\alpha)(\beta)}$ (with the signature $+2$) defining a scalar product for vectors (covectors) tangent to M, and a $\mathscr{gl}(4,R)$-valued connection one-form $\boldsymbol{\Gamma}^{(\alpha)}{}_{(\beta)} = \Gamma_\lambda{}^{(\alpha)}{}_{(\beta)}\, dx^\lambda$ [$\mathscr{gl}(4,R)$ is the Lie algebra of the general linear group GL(4,R), that is, the algebra of all 4×4 real matrices]. Matter in our theory is described by a tensor field $\boldsymbol{\phi}$. The basic gravitational gauge group for such theories is the bundle product of the local GL(4,R) group and the diffeomorphism group of space-time. The construction of the bundle product GL(4,R)$\times_b$ Diff M is presented in Appendix F.

The main goal of the present paper is to give a Hamiltonian formulation for GL(4,R) theories and to develop methods that could enable us to pose the initial value problem as well as to investigate the number of independent degrees of freedom for particular Lagrangians. For theories with metric compatible connections [SO(3,1)-gauge theories] such a problem was investigated in Ref. 30, and the general theory of gravitational Hamiltonian systems developed therein was applied to the Einstein–Cartan theory with tensor matter fields in Ref. 31. By means of the methods of Refs. 30 and 31, one of us has recently investigated the canonical structure of the Yang theory of gravity[24,28] presenting the complete set of its canonical (symplectic) variables, symplectic constraints, and gauge transformations.[32] Simultaneously, we should like to emphasize the fact that the complete canonical analysis of the theory with a given Lagrangian is not easy. The general method given in Ref. 30 enables us only to start the procedure and to separate the dynamical symplectic variables. It is relatively easy to compute primary and secondary symplectic constraints for these variables[33] but it is difficult to find the complete set of gauge variables as well as to determine the evolution of the remaining nondynamical variables.

The results of the present and previous papers show the following differences between SO(3,1)- and GL(4,R)-gauge theories of gravity.

(i) In SO(3,1) theories matter is characterized by its spin and energy-momentum distribution. For GL(4,R) theories the skew-symmetric spin tensor $\mathscr{s}^\lambda{}_{(\alpha)(\beta)}$ is replaced with the hypermomentum tensor $\mathscr{h}^\lambda{}_{(\alpha)}{}^{(\beta)}$ having no special symmetry properties.[4,34] In GL(4,R) theories, besides the canonical energy-momentum tensor ${}_c\mathscr{T}^\lambda{}_{(\alpha)}$ of matter, we have also a symmetric energy-momentum tensor ${}_s\mathscr{T}^{(\alpha)(\beta)}$. Conservation laws for these quantities and their implications are discussed in Sec. III.

(ii) For the Lorentz group we have its natural decomposition into boost transformations and three-dimensional rotations (the Cartan decomposition). It has an elegant geometric interpretation and enables us to define symplectic variables of the theory.[30] Generalizations of these constructions for the GL(4,R) case are not obvious. One reasonable construction is given in Sec. IV and further used throughout the paper. We can pose the question: is our choice of the $3+1$ decomposition for the GL(4,R) group canonical in any sense? This problem remains open.

(iii) For SO(3,1) theories we have ten basic gravitational gauge variables, that is, such variables that are completely arbitrary. Six of them are related to the action of the local Lorentz group and four to the action of Diff M in the space of solutions. Generally, for GL(4,R) theories we have 20 gravitational gauge variables: 16 corresponding to the action of the local GL(4,R) group and four corresponding to the action of Diff M. Therefore, we have at least ten gravitational gauge variables in SO(3,1)-gauge theories and at least 20 gravitational gauge variables in GL(4,R) theories. For particular Lagrangians in both cases the complete set of gauge transformations may be larger than the basic set of gauge transformations generated by the action of SO(3,1) $\times_b$ Diff M or GL(4,R) $\times_b$ Diff M. Simultaneously, we have more gravitational gauge variables. For instance, in the SO(3,1)–Yang theory we have 13 gravitational gauge variables. For the SO(3,1)–Einstein theory (the Einstein–Cartan theory) we have exactly ten gravitational gauge variables, but for the GL(4,R)–Einstein theory in vacuum we have four additional gauge variables and the total number of gravitational gauge variables is 24 (see Sec. V).

(iv) The methods developed in the previous[30,31,35] and the present paper enable us to determine the number of independent degrees of freedom for SO(3,1) and GL(4,R) theories. Of course, it depends on the choice of gravitational and matter Lagrangians. For instance, for a general canonically regular SO(3,1) theory we have 40 gravitational degrees of freedom (in the phase space) but for a highly degenerate ECSK theory we have only four of them.[31] Though for the GL(4,R)–Einstein theory in vacuum we have four additional gauge transformations, the number of independent degrees of freedom equals 4 as in the SO(3,1) case. In vacuum, the GL(4,R)–Einstein theory is equivalent to the Einstein–Cartan theory. The situation changes if we couple both theories to matter fields. We prove in Sec. VI that for the GL(4,R)–Einstein theory with a Klein–Gordon-type vector field we have only six independent matter degrees of freedom comparing to eight matter degrees of freedom for the SO(3,1) theory.

(v) Both in the present and in our previous papers,[30,31,35] matter was described by tensor fields. We know, however, that real physical particles are hadrons and leptons, that is, half-spin fermions. (Bosons should be considered as quanta of gauge fields.) We can generalize the theory presented in Ref. 30 for spinor fields considering the bundle of spinor frames instead of the bundle of orthonormal covector (vector) frames and the local SL(2,C) group instead of the local Lorentz group. Such a structure describing gravity coupled to Dirac, Fierz–Pauli, Weyl, and other spinor fields has recently been presented in Ref. 36. The general linear group has no finite-dimensional double-valued representations and therefore the problem of GL(4,R) spinors seems to be incorrectly posed. Ne'eman and Sijacki[37] have shown, however, that infinite-dimensional double-valued representations of GL(4,R) have a relation to physics and could describe hadrons. Mickelsson[38] has recently constructed a first-order differential equation covariant with respect to the double-covering $\overline{\mathrm{GL}(4,R)}$ of GL(4,R). The corresponding Dirac-type differential operator acts in the spaces of infinite-dimensional representations of $\overline{\mathrm{GL}(4,R)}$. With some natural assumptions this construction restricted to the subgroup SL(2,C) of $\overline{\mathrm{GL}(4,R)}$ is reduced to the standard Dirac equation.

II. GEOMETRY OF SPACE-TIME AND MATTER FIELDS IN GL(4,R)-GAUGE THEORIES OF GRAVITY

In the Einstein theory of general relativity space-time M is endowed with a metric tensor $\mathbf{g} = (g_{\mu\nu})$. This metric determines a scalar product for vectors tangent to M and space-time intervals between infinitesimally close events (points in M). A parallel transport of vectors tangent to M between two infinitesimally close points in M is given by means of the Riemannian (Christoffel) connection $\gamma = (\gamma_\lambda{}^\nu{}_\mu)$ determined by $\mathbf{g}$.

In gauge approaches to gravity the metric is replaced by four linearly independent fields of covectors (or vectors) on M:

$$\mathbf{x} \to (\mathbf{e}^{(\alpha)}(\mathbf{x}))^3_{\alpha=0} \quad [\text{or } \mathbf{x} \to (\mathbf{e}_{(\alpha)}(\mathbf{x}))^3_{\alpha=0}]. \tag{2.1}$$

Tensor fields on M are described by means of their components with respect to dual bases $(\mathbf{e}^{(\alpha)})$ and $(\mathbf{e}_{(\alpha)})$, that is,

$$\boldsymbol{\phi} = \phi^{(\alpha_1)\cdots(\alpha_k)}_{(\beta_1)\cdots(\beta_s)} \otimes \mathbf{e}_{(\alpha_1)} \otimes \cdots \otimes \mathbf{e}_{(\alpha_k)} \otimes \mathbf{e}^{(\beta_1)} \otimes \cdots \otimes \mathbf{e}^{(\beta_s)}. \tag{2.2}$$

Components $\phi^{\Sigma} = \phi^{(\alpha_1)\cdots(\alpha_k)}_{(\beta_1)\cdots(\beta_s)}$ are called anholonomic components of $\boldsymbol{\phi}$.

Locally in a given coordinate system (x^λ) on M we can describe a tensor field $\boldsymbol{\phi}$ by means of its holonomic components $\phi^{\mu_1\cdots\mu_k}_{\nu_1\cdots\nu_s}$, that is, its components with respect to coordinate bases $\partial/\partial x^\mu$, dx^ν. We have the following relations between holonomic and anholonomic components of a tensor field $\boldsymbol{\phi}$:

$$\begin{aligned}
\phi^{(\alpha_1)\cdots(\alpha_k)}_{(\beta_1)\cdots(\beta_s)} &= \phi^{\mu_1\cdots\mu_k}_{\nu_1\cdots\nu_s} e^{(\alpha_1)}_{\mu_1} \cdots e^{(\alpha_k)}_{\mu_k} e^{\nu_1}_{(\beta_1)} \cdots e^{\nu_s}_{(\beta_s)}, \\
\phi^{\mu_1\cdots\mu_k}_{\nu_1\cdots\nu_s} &= \phi^{(\alpha_1)\cdots(\alpha_k)}_{(\beta_1)\cdots(\beta_s)} e^{\mu_1}_{(\alpha_1)} \cdots e^{\mu_k}_{(\alpha_k)} e^{(\beta_1)}_{\nu_1} \cdots e^{(\beta_s)}_{\nu_s},
\end{aligned} \tag{2.3}$$

where $e^{(\alpha)}_\mu$, $e^\nu_{(\beta)}$ are components of tetrad fields (2.1) in coordinate bases

$$\mathbf{e}^{(\alpha)} = e^{(\alpha)}_\mu \, dx^\mu, \quad \mathbf{e}_{(\beta)} = e^\nu_{(\beta)} \frac{\partial}{\partial x^\nu}. \tag{2.4}$$

In the tetrad formalism affine connections are described by their anholonomic components

$$\Gamma_\lambda{}^{(\alpha)}{}_{(\beta)} = -e^\nu_{(\beta)} \, \partial_\lambda e^{(\alpha)}_\nu + \Gamma_\lambda{}^\sigma{}_\nu e^{(\alpha)}_\sigma e^\nu_{(\beta)}, \tag{2.5}$$

where $\Gamma_\lambda{}^\sigma{}_\nu$ are standard holonomic components of an affine connection $\boldsymbol{\Gamma}$ on M (see Ref. 39). Quantities $\Gamma_\lambda{}^{(\alpha)}{}_{(\beta)}$ are tensors with respect to coordinate transformations in M. Therefore they represent a $\mathfrak{gl}(4,R)$-valued one-form $\boldsymbol{\Gamma}^{(\alpha)}{}_{(\beta)}$.

The metric structure of space-time is given by a scalar product for fields of tetrads (2.1). Such a scalar product is defined by means of fields of symmetric 4×4 matrices $[g_{(\alpha)(\beta)}]$, $[g^{(\alpha)(\beta)}]$ (with the signature $+2$). We postulate

$$\mathbf{e}_{(\alpha)} \cdot \mathbf{e}_{(\beta)} = g_{(\alpha)(\beta)}, \quad \mathbf{e}^{(\alpha)} \cdot \mathbf{e}^{(\beta)} = g^{(\alpha)(\beta)}, \tag{2.6}$$

and $[g^{(\alpha)(\beta)}] = [g_{(\alpha)(\beta)}]^{-1}$. From (2.3) and (2.6) we have

$$\begin{aligned}
\frac{\partial}{\partial x^\mu} \cdot \frac{\partial}{\partial x^\nu} &= g_{\mu\nu} = e^{(\alpha)}_\mu e^{(\beta)}_\nu g_{(\alpha)(\beta)}, \\
dx^\mu \cdot dx^\nu &= g^{\mu\nu} = e^\mu_{(\alpha)} e^\nu_{(\beta)} g^{(\alpha)(\beta)}.
\end{aligned} \tag{2.7}$$

Formulas (2.6) and (2.7) define natural scalar products in spaces of tensors on M.

For a tensor density $\mathbf{F}$ of weight r we define the following GL(4,R)-covariant derivative:

$$\begin{aligned}
\mathscr{D}_\lambda \mathscr{F}^{\mu(\alpha)}_{\nu(\beta)} = {}& \partial_\lambda \mathscr{F}^{\mu(\alpha)}_{\nu(\beta)} - r\gamma_\lambda{}^\tau{}_\tau \mathscr{F}^{\mu(\alpha)}_{\nu(\beta)} + \gamma_\lambda{}^\mu{}_\tau \mathscr{F}^{\tau(\alpha)}_{\nu(\beta)} \\
& - \gamma_\lambda{}^\tau{}_\nu \mathscr{F}^{\mu(\alpha)}_{\tau(\beta)} + \Gamma_\lambda{}^{(\alpha)}{}_{(\tau)} \mathscr{F}^{\mu(\tau)}_{\nu(\beta)} - \Gamma_\lambda{}^{(\tau)}{}_{(\beta)} \mathscr{F}^{\mu(\alpha)}_{\nu(\tau)},
\end{aligned} \tag{2.8}$$

where the $\gamma_\lambda{}^\epsilon{}_\tau$ are holonomic components of the Riemannian connection γ determined by the metric $g_{\mu\nu}$ on M.

For GL(4,R)-tensor-valued differential forms the $\mathscr{D}$ derivative reduces to Cartan's exterior covariant derivative.[11]

If $\mathbf{r} = (r_\lambda{}^\mu{}_\nu)$ is the difference tensor,

$$r_\lambda{}^\mu{}_\nu = \Gamma_\lambda{}^\mu{}_\nu - \gamma_\lambda{}^\mu{}_\nu, \tag{2.9}$$

then we have

$$\mathscr{D}_\lambda e^{(\alpha)}_\mu = r_\lambda{}^{(\alpha)}{}_\mu, \quad \mathscr{D}_\lambda e^\nu_{(\beta)} = -r_\lambda{}^\nu{}_{(\beta)}. \tag{2.10}$$

In his Ph.D. dissertation Yasskin[27] presented an interesting formalism of "two tangent spaces" and gave a geometric interpretation for the $\mathscr{D}$ derivative. We have the standard definitions of the curvature and torsion tensors

$$R^{(\alpha)}{}_{(\beta)\mu\nu} = \partial_\mu \Gamma_\nu{}^{(\alpha)}{}_{(\beta)} - \partial_\nu \Gamma_\mu{}^{(\alpha)}{}_{(\beta)} + \Gamma_\mu{}^{(\alpha)}{}_{(\epsilon)}\Gamma_\nu{}^{(\epsilon)}{}_{(\beta)} - \Gamma_\nu{}^{(\alpha)}{}_{(\epsilon)}\Gamma_\mu{}^{(\epsilon)}{}_{(\beta)}, \tag{2.11}$$

$$Q^{(\alpha)}{}_{\mu\nu} = \mathscr{D}_\mu e^{(\alpha)}_\nu - \mathscr{D}_\nu e^{(\alpha)}_\mu = \partial_\mu e^{(\alpha)}_\nu - \partial_\nu e^{(\alpha)}_\mu + \Gamma_\mu{}^{(\alpha)}{}_{(\epsilon)} e^{(\epsilon)}_\nu - \Gamma_\nu{}^{(\alpha)}{}_{(\epsilon)} e^{(\epsilon)}_\mu. \tag{2.12}$$

Remark: It is reasonable to define $\mathscr{D}_\mu \Gamma_\nu{}^{(\alpha)}{}_{(\beta)} = R^{(\alpha)}{}_{(\beta)\mu\nu}$. Connections on space-time for GL(4,R) theories are metric noncompatible in general. That is, the nonmetricity tensor

$$M_{\lambda(\alpha)(\beta)} = \tfrac{1}{2}\mathscr{D}_\lambda g_{(\alpha)(\beta)} \tag{2.13}$$

is different from zero. There are the important relations between the difference, nonmetricity, and torsion tensors:

$$M_{\lambda\mu\nu} = -\tfrac{1}{2}(r_{\lambda\mu\nu} + r_{\lambda\nu\mu}), \quad Q^\lambda{}_{\mu\nu} = r_\mu{}^\lambda{}_\nu - r_\nu{}^\lambda{}_\mu. \tag{2.14}$$

A fixed field of tetrads $(\mathbf{e}^{(\alpha)})$, a metric $[g_{(\alpha)(\beta)}]$, and a $g\ell$(4,R)-valued connection one-form $[\Gamma^{(\alpha)}{}_{(\beta)}]$ define the geometric structure of space-time. It enables us to compute angles between vectors tangent to M and to transport them in a parallel manner. It is clear, however, that the structure defined by those three entities is too rich for these two purposes. If we rotate the given field of tetrads by means of a GL(4,R) matrix $[G^{(\alpha)}{}_{(\beta)}]$ and simultaneously transform the metric according to the appropriate tensorial representation of GL(4,R), then the scalar products at every point of M remain unchanged. Of course, at every point $\mathbf{x}$ of M we can choose another matrix $[G^{(\alpha)}{}_{(\beta)}]$, that is why we have an action of the local GL(4,G) group. Let $\mathbf{x}\rightarrow[G^{(\alpha)}{}_{(\beta)}]\in$GL(4,$R$) be a smooth field of 4×4 invertible matrices on M. We define the transformations

$$\begin{aligned} '\mathbf{e}^{(\alpha)} &= G^{-1(\alpha)}{}_{(\beta)}\mathbf{e}^{(\beta)}, \quad '\mathbf{e}_{(\alpha)} = G^{(\beta)}{}_{(\alpha)}\mathbf{e}_{(\beta)}, \\ 'g_{(\alpha)(\beta)} &= G^{(\epsilon)}{}_{(\alpha)}G^{(\tau)}{}_{(\beta)}g_{(\epsilon)(\tau)}, \\ '\phi^{(\alpha)}{}_{(\beta)} &= G^{-1(\alpha)}{}_{(\epsilon)}G^{(\tau)}{}_{(\beta)}\phi^{(\epsilon)}{}_{(\tau)}. \end{aligned} \tag{2.15}$$

For components of connections one-forms we have

$$'\Gamma_\lambda{}^{(\alpha)}{}_{(\beta)} = G^{-1(\alpha)}{}_{(\epsilon)}G^{(\tau)}{}_{(\beta)}\Gamma_\lambda{}^{(\epsilon)}{}_{(\tau)} - G^{(\epsilon)}{}_{(\beta)}\,\partial_\lambda G^{-1(\alpha)}{}_{(\epsilon)}. \tag{2.16}$$

The formulas (2.15) and (2.16) give rise to tensor transformation rules for the covariant derivative $\mathscr{D}$ with respect to the action of the local GL(4,R) group.

In the present paper we deal with Lagrangian GL(4,R)-gauge theories of gravity. Physical systems under consideration are described by fields of tetrads $(\mathbf{e}^{(\alpha)})$, metric tensors $[g_{(\alpha)(\beta)}]$, $g\ell$(4,R)-valued connection one-forms $[\Gamma^{(\alpha)}{}_{(\beta)}]$, and tensor fields $\boldsymbol{\phi}$ on M. These quantities are the variational potentials for our theories. The dynamics of a particular theory is determined by a Lagrangian being the sum of gravitational and matter fields Lagrangians. We assume that Lagrangians depend on variational potentials and their partial derivatives of the first order.

In a local coordinate system we write

$$\begin{aligned} \mathscr{L}_g &= \mathscr{L}_g(e^{(\alpha)}_\mu;\Gamma_\mu{}^{(\alpha)}{}_{(\beta)};g_{(\alpha)(\beta)};Q^{(\alpha)}{}_{\mu\nu}; R^{(\alpha)}{}_{(\beta)\mu\nu};M_{\mu(\alpha)(\beta)}), \\ \mathscr{L}_m &= \mathscr{L}_m(e^{(\alpha)}_\mu;\Gamma_\mu{}^{(\alpha)}{}_{(\beta)};g_{(\alpha)(\beta)};\phi^\Sigma;\mathscr{D}_\mu\phi^\Sigma). \end{aligned} \tag{2.17}$$

Let $\mathscr{L}$ be the total Lagrangian density

$$\mathscr{L} = \mathscr{L}_g + \mathscr{L}_m. \tag{2.18}$$

The Euler–Lagrange equations read

$$\begin{aligned} (\mathscr{E}1)^\mu{}_{(\alpha)} &= \delta\mathscr{L}/\delta e^{(\alpha)}_\mu = 0, \\ (\mathscr{E}2)^\mu{}_{(\alpha)}{}^{(\beta)} &= \delta\mathscr{L}/\delta\Gamma^{(\alpha)}_{\mu(\beta)} = 0, \\ (\mathscr{E}3)^{(\alpha)(\beta)} &= 2\delta L/\delta g_{(\alpha)(\beta)} = 0, \\ (\mathscr{E}\mathscr{M})_\Sigma &= \delta\mathscr{L}/\delta\phi^\Sigma = 0. \end{aligned} \tag{2.19}$$

We assume that Lagrangians $\mathscr{L}_g$ and $\mathscr{L}_m$ are invariant with respect to the action (2.15) and (2.16) of the local GL(4,R) group and with respect to the standard action of the diffeomorphism group of space-time. In the next section we derive important differential identities following from these assumptions.

Remark: In the present paper the matter variational potentials are anholonomic components of tensor fields. We may say that matter is described by fields of GL(4,R) tensors. With respect to (local) GL(4,R) rotations such quantities are tensors but with respect to transformations of local coordinates in M they are scalars. In physics, however, we often need more general objects to describe matter fields. For instance, in Yang–Mills theories potentials are differential forms on space-time with values in the Lie algebra of the corresponding Yang–Mills group. For the Rarita–Schwinger field we have SL(2,C)-spinor-valued one-forms on space-time. In a general case, we may assume that matter potentials are differential k forms on space-time with values in an appropriate representation of the chosen gauge group.[11] In gravity we consider k forms with values in GL(4,R) tensors. Such an approach to matter fields was investigated profoundly in Ref. 36 for SL(2,C) [SO(3,1)] gauge theories of gravity. The results of this paper show that cases $k>0$ are essentially different from the case $k=0$, which we treat in the present paper. In fact, by combining the methods and results of Ref. 36 and those of the present paper we are able to solve the problem also for GL(4,R)-gauge theories of gravity.

III. FIELD EQUATIONS, CONSERVATION LAWS, AND CONTRACTED BIANCHI IDENTITIES

This section contains a brief summary of the formulas and relations generalizing some facts well known for particular classical field theories. We derive them from the invariance properties of gravitational and matter Lagrangians with respect to the action of the gauge group. These formulas are the starting point for the explicit Hamiltonian formulation of our theory.

Invariance properties of Lagrangians with respect to the action of the gauge group lead to differential identities among partial derivatives of Lagrangians, which for classical field theories are due to Belinfante, Rosenfeld, and Pauli[40] (at least in their simplest form). The conditions of the invariance of the Lagrangian with respect to the action of the local GL(4,R) group and the diffeomorphism group of space-time can be formulated as

$$\begin{aligned}\delta_X \mathscr{L} = {} & \frac{\partial \mathscr{L}}{\partial e^{(\alpha)}_\mu}\delta_X e^{(\alpha)}_\mu + \frac{\partial \mathscr{L}}{\partial Q^{(\alpha)}{}_{\mu\nu}}\delta_X Q^{(\alpha)}{}_{\mu\nu} \\ & + \frac{\partial \mathscr{L}}{\partial \Gamma_\mu{}^{(\alpha)}{}_{(\beta)}}\delta_X \Gamma_\mu{}^{(\alpha)}{}_{(\beta)} \\ & + \frac{\partial \mathscr{L}}{\partial R^{(\alpha)}{}_{(\beta)\mu\nu}}\delta_X R^{(\alpha)}{}_{(\beta)\mu\nu} + \frac{\partial \mathscr{L}}{\partial g_{(\alpha)(\beta)}}\delta_X g_{(\alpha)(\beta)} \\ & + \frac{\partial \mathscr{L}}{\partial M_{\mu(\alpha)(\beta)}}\delta_X M_{\mu(\alpha)(\beta)} + \frac{\partial \mathscr{L}}{\partial \phi^\Sigma}\delta_X \phi^\Sigma \\ & + \frac{\partial \mathscr{L}}{\partial \mathscr{D}_\mu \phi^\Sigma}\delta_X \mathscr{D}_\mu \phi^\Sigma, \end{aligned} \tag{3.1}$$

where δ_X denotes the action of a generator X of the local GL(4,R) or the diffeomorphism group on the dynamical variables. For loc GL(4,R), $X = [L^{(\alpha)}{}_{(\beta)}]$ is a field of real 4×4 matrices on space-time, δ_L can be computed from (2.15):

$$\begin{aligned} & \delta_L e^{(\alpha)}_\mu = -L^{(\alpha)}{}_{(\beta)} e^{(\beta)}_\mu, \quad \delta_L \Gamma_\mu{}^{(\alpha)}{}_{(\beta)} = \mathscr{D}_\mu L^{(\alpha)}{}_{(\beta)}, \\ & \delta_L g_{(\alpha)(\beta)} = L^{(\epsilon)}{}_{(\alpha)} g_{(\epsilon)(\beta)} + L^{(\epsilon)}{}_{(\beta)} g_{(\alpha)(\epsilon)}, \\ & \delta_L \phi^{(\alpha)}{}_{(\beta)} = -L^{(\alpha)}{}_{(\epsilon)}\phi^{(\epsilon)}{}_{(\beta)} + L^{(\epsilon)}{}_{(\beta)}\phi^{(\alpha)}{}_{(\epsilon)}, \text{ etc.,} \end{aligned} \tag{3.2}$$

and the invariance of the Lagrangian gives rise to $\delta_L \mathscr{L} = 0$. For the Diff M, $X = (Z^\lambda)$ is a smooth vector field on space-time and δ_Z is given as a standard Lie derivative of geometrical objects on M, i.e., $\delta_Z = \mathscr{L}_Z$; for invariant Lagrangians we have $\mathscr{L}_Z \mathscr{L} = \partial_\lambda (Z^\lambda \mathscr{L})$.

In order to give an elegant and natural form of differential identities equivalent to (3.1) we have to introduce the notions of canonical momenta, hypermomentum, and energy-momentum tensors. The canonical momenta of the gravitational field are

$$\mathscr{U}^{\mu\nu}{}_{(\alpha)} = \frac{2\partial \mathscr{L}_g}{\partial Q^{(\alpha)}{}_{\mu\nu}}, \quad \mathscr{P}^{\mu\nu}{}_{(\alpha)}{}^{(\beta)} = \frac{2\partial \mathscr{L}_g}{\partial R^{(\alpha)}{}_{(\beta)\mu\nu}}, \tag{3.3}$$

$$\mathscr{V}^{\mu(\alpha)(\beta)} = \frac{\partial \mathscr{L}_g}{\partial M_{\mu(\alpha)(\beta)}}$$

(in the above formulas we treat $Q^{(\alpha)}{}_{\mu\nu}$, $R^{(\alpha)}{}_{(\beta)\mu\nu}$, $M_{\mu(\alpha)(\beta)}$ as independent quantities).

The four-momenta and hypermomentum of the matter field are defined by (cf. Refs. 4, 34, and 41)

$$p^\mu{}_\Sigma = \frac{\partial \mathscr{L}_m}{\partial \mathscr{D}_\mu \phi^\Sigma}, \quad h^\mu{}_{(\alpha)}{}^{(\beta)} = p^\mu{}_\Sigma f_{(\alpha)}{}^{(\beta)\Sigma}{}_\Lambda \phi^\Lambda, \tag{3.4}$$

where $f_{(\alpha)}{}^{(\beta)\Sigma}{}_\Lambda$ are the generators of the GL(4,R) representation corresponding to the matter field ϕ^Σ.

The hypermomentum tensor can be decomposed into its spin, dilatation, and the proper hypermomentum parts. The spin part is related to torsion of space-time. Two remaining parts correspond to nonmetricity—to its trace and traceless components, respectively. We recall that a connection with purely tracelike (diagonal) nonmetricity preserves angles of vectors under the parallel transport and a connection with traceless nonmetricity preserves the volume element.[4]

In theories with tetrads and metrics as independent variables we can construct two kinds of energy-momentum tensors: the canonical and the symmetric ones. There are some relations between them resulting from (3.1). The canonical energy-momentum tensors (tensor densities) of matter and gravitational fields are defined as

$${}_{cm}\mathscr{T}^\mu{}_{(\alpha)} = \frac{\partial \mathscr{L}_m}{\partial e^{(\alpha)}_\mu}, \quad {}_{cg}\mathscr{T}^\mu{}_{(\alpha)} = \frac{\partial \mathscr{L}_g}{\partial e^{(\alpha)}_\mu}. \tag{3.5}$$

The symmetric energy-momentum tensors (tensor densities) of matter and gravity are

$${}_{sm}\mathscr{T}^{(\alpha)(\beta)} = 2\frac{\partial \mathscr{L}_m}{\partial g_{(\alpha)(\beta)}}, \quad {}_{sg}\mathscr{T}^{(\alpha)(\beta)} = 2\frac{\partial \mathscr{L}_g}{\partial g_{(\alpha)(\beta)}}. \tag{3.6}$$

The invariance conditions (3.1) for gravitational and matter Lagrangians are equivalent to the following differential identities:

$$\begin{aligned} & \frac{\partial \mathscr{L}_g}{\partial \Gamma_\mu{}^{(\alpha)}{}_{(\beta)}} = 0, \quad \frac{\partial \mathscr{L}_m}{\partial \Gamma_\mu{}^{(\alpha)}{}_{(\beta)}} = 0, \\ & {}_{cm}\mathscr{T}^\mu{}_\nu = \delta^\mu{}_\nu \mathscr{L}_m - p^\mu{}_\Sigma \mathscr{D}_\nu \phi^\Sigma, \\ & {}_{cg}\mathscr{T}^\mu{}_\nu = \delta^\mu{}_\nu \mathscr{L}_g - \mathscr{P}^{\mu\epsilon}{}_{(\alpha)}{}^{(\beta)} R^{(\alpha)}{}_{(\beta)\nu\epsilon} \\ & \qquad - \mathscr{U}^{\mu\epsilon}{}_{(\alpha)} Q^{(\alpha)}{}_{\nu\epsilon} - \mathscr{V}^{\mu(\alpha)(\beta)} M_{\nu(\alpha)(\beta)}, \end{aligned} \tag{3.7}$$

$$\begin{aligned} & {}_{sm}\mathscr{T}^{(\alpha)}{}_{(\beta)} - {}_{cm}\mathscr{T}^{(\alpha)}{}_{(\beta)} \\ & \quad = \mathscr{D}_\lambda h^\lambda{}_{(\beta)}{}^{(\alpha)} + \text{terms linear in } (\mathscr{E}\mathscr{M})_\Sigma, \end{aligned} \tag{3.8a}$$

$$\begin{aligned} & {}_{sg}\mathscr{T}^{(\alpha)}{}_{(\beta)} - {}_{cg}\mathscr{T}^{(\alpha)}{}_{(\beta)} \\ & \quad = \tfrac{1}{2}\mathscr{U}^{\mu\nu}{}_{(\beta)} Q^{(\alpha)}{}_{\mu\nu} + \tfrac{1}{2}(\mathscr{P}^{\mu\nu}{}_{(\beta)}{}^{(\tau)} R^{(\alpha)}{}_{(\tau)\mu\nu} \\ & \qquad - \mathscr{P}^{\mu\nu}{}_{(\tau)}{}^{(\alpha)} R^{(\tau)}{}_{(\beta)\mu\nu}) - 2\mathscr{V}^{\mu(\alpha)(\tau)} M_{\mu(\beta)(\tau)}, \end{aligned} \tag{3.8b}$$

$$\begin{aligned} \mathscr{D}_\lambda {}_{cm}\mathscr{T}^\lambda{}_{(\alpha)} = {} & {}_{cm}\mathscr{T}^\lambda{}_{(\tau)} Q^{(\tau)}{}_{(\alpha)\lambda} + {}_{sm}\mathscr{T}^{(\epsilon)(\tau)} M_{(\alpha)(\epsilon)(\tau)} \\ & + h^\lambda{}_{(\tau)}{}^{(\epsilon)} R^{(\tau)}{}_{(\epsilon)(\alpha)\lambda} \\ & + \text{terms linear in } (\mathscr{E}\mathscr{M})_\Sigma. \end{aligned} \tag{3.9}$$

Formulas (3.8a) and (3.9) are called the conservation laws for matter fields.

Remark: The relations (3.8) coincide with those obtained in Szczyrba.[5(a)]

If the matter field equations are satisfied then the formula (3.9) can be transformed to

$$\mathscr{D}_\lambda {}_{cm}\mathscr{T}^\lambda{}_\mu = h^\lambda{}_{(\alpha)}{}^{(\beta)} R^{(\alpha)}{}_{(\beta)\mu\lambda} - r_\mu{}^{(\alpha)}{}_{(\beta)} \mathscr{D}_\lambda h^\lambda{}_{(\alpha)}{}^{(\beta)}.$$

Now in the case $h^\lambda{}_{(\alpha)}{}^{(\beta)} = 0$ we have from (3.8a) ${}_{sm}\mathscr{T}^{(\alpha)(\beta)} = {}_{cm}\mathscr{T}^{(\alpha)(\beta)}$ and $\mathscr{D}_\lambda {}_m\mathscr{T}^\lambda{}_\mu = 0$. It is a somewhat surprising result because it follows from the Riemann–Einstein conservation law $\nabla_\lambda {}_m\mathscr{T}^\lambda{}_\mu = 0$ that particles move along geodesics of the Riemannian metric. Therefore we see that in our theory spinless particles (or dust) move also along geodesics of the Riemannian metric.

The left-hand sides (lhs's) of the field equations (2.19), rewritten in terms of canonical momenta and energy-momentum tensors, read

$$(\mathscr{E}1)^{\lambda}{}_{(\alpha)} = {}_{cg}\mathscr{T}^{\lambda}{}_{(\alpha)} + {}_{cm}\mathscr{T}^{\lambda}{}_{(\alpha)} - \mathscr{D}_{\mu}\mathscr{U}^{\mu\lambda}{}_{(\alpha)},$$
$$(\mathscr{E}2)^{\lambda}{}_{(\alpha)}{}^{(\beta)} = \mathscr{h}^{\lambda}{}_{(\alpha)}{}^{(\beta)} + \mathscr{U}^{\lambda(\beta)}{}_{(\alpha)} - \mathscr{V}^{\lambda(\beta)}{}_{(\alpha)} - \mathscr{D}_{\mu}\mathscr{P}^{\mu\lambda}{}_{(\alpha)}{}^{(\beta)}, \tag{3.10}$$
$$(\mathscr{E}3)^{(\alpha)(\beta)} = {}_{sg}\mathscr{T}^{(\alpha)(\beta)} + {}_{sm}\mathscr{T}^{(\alpha)(\beta)} - \mathscr{D}_{\mu}V^{\mu(\alpha)(\beta)},$$
$$(\mathscr{E}\mathscr{M})_{\Sigma} = \mathscr{j}_{\Sigma} - \mathscr{D}_{\mu}\mathscr{h}^{\mu}{}_{\Sigma},$$

where $\mathscr{j}_{\Sigma} = \partial\mathscr{L}_m/\partial\phi^{\Sigma}$ is matter current.

Proposition 1: If matter field equations (2.19d) are satisfied, then the contracted Bianchi identities (CBI) hold:

$$\mathscr{D}_{\lambda}(\mathscr{E}1)^{\lambda}{}_{(\alpha)} = (\mathscr{E}1)^{\lambda}{}_{(\tau)}Q^{(\tau)}{}_{(\alpha)\lambda} + (\mathscr{E}2)^{\lambda}{}_{(\epsilon)}{}^{(\tau)}R^{(\epsilon)}{}_{(\tau)(\alpha)\lambda} + (\mathscr{E}3)^{(\epsilon)(\tau)}M_{(\alpha)(\epsilon)(\tau)}, \tag{3.11}$$
$$\mathscr{D}_{\lambda}(\mathscr{E}2)^{\lambda}{}_{(\alpha)}{}^{(\beta)} = (\mathscr{E}3)^{(\beta)}{}_{(\alpha)} - (\mathscr{E}1)^{(\beta)}{}_{(\alpha)}. \tag{3.12}$$

Invariance properties of Lagrangians with respect to the action of the gauge group give rise to the existence of two families of conserved quantities $\mathscr{I}_L$ and $\mathscr{E}_L$. The conservation laws for these quantities are equivalent to the gravitational field equations (2.19). In order to construct them we rewrite (3.1) in the Noether form

$$\partial_{\lambda}\mathscr{I}^{\lambda}_{X} + (\mathscr{E}1)^{\lambda}{}_{(\alpha)}\delta_X e^{(\alpha)}_{\lambda} + (\mathscr{E}2)^{\lambda}{}_{(\alpha)}{}^{(\beta)}\delta_X\Gamma_{\lambda}{}^{(\alpha)}{}_{(\beta)} + (\mathscr{E}3)^{(\alpha)(\beta)}\delta_X g_{(\alpha)(\beta)} + (\mathscr{E}\mathscr{M})_{\Sigma}\delta_X\phi^{\Sigma} = 0. \tag{3.13}$$

For infinitesimal rotations $X = [L^{(\alpha)}{}_{(\beta)}]$, $\mathscr{I}^{\lambda}_{L} = \mathscr{I}^{\lambda}_{L}$ is called the hypermomentum vector density, and we have

$$\mathscr{I}^{\lambda}_{L} = \mathscr{U}^{\lambda\mu}{}_{(\alpha)}\delta_L e^{(\alpha)}_{\mu} + \mathscr{P}^{\lambda\mu}{}_{(\alpha)}{}^{(\beta)}\delta_L\Gamma_{\mu}{}^{(\alpha)}{}_{(\beta)} + \tfrac{1}{2}\mathscr{V}^{\lambda(\alpha)(\beta)}\delta_L g_{(\alpha)(\beta)} + \mathscr{h}^{\lambda}{}_{\Sigma}\delta_L\phi^{\Sigma},$$

or

$$\mathscr{I}^{\lambda}_{L} = (\mathscr{E}2)^{\lambda}{}_{(\alpha)}{}^{(\beta)}L^{(\alpha)}{}_{(\beta)} + \partial_{\lambda}(\mathscr{P}^{\mu\lambda}{}_{(\alpha)}{}^{(\beta)}L^{(\alpha)}{}_{(\beta)}). \tag{3.14}$$

For infinitesimal translations $X = (Z^{\lambda})$, $\mathscr{E}^{\lambda}_{Z} = \mathscr{I}^{\lambda}_{Z}$ is called the energy-momentum vector density, and we have

$$\mathscr{E}^{\lambda}_{Z} = \mathscr{U}^{\lambda\mu}{}_{(\alpha)}\delta_Z e^{(\alpha)}_{\mu} + \mathscr{P}^{\lambda\mu}{}_{(\alpha)}{}^{(\beta)}\delta_Z\Gamma_{\mu}{}^{(\alpha)}{}_{(\beta)} + \tfrac{1}{2}\mathscr{V}^{\lambda(\alpha)(\beta)}\delta_Z g_{(\alpha)(\beta)} + \mathscr{h}^{\lambda}{}_{\Sigma}\delta_Z\phi^{\Sigma} - Z^{\lambda}\mathscr{L}. \tag{3.15a}$$

Noninvariance of the standard Lie derivative with respect to the loc GL(4,R) rotations requires its replacing with the covariant Lie derivative ${}_{\zeta}\mathscr{L}_Z$ in the formula for the generator of an infinitesimal action of Diff M in the space of dynamical variables. Substituting $\delta_Z = {}_{\zeta}\mathscr{L}_Z$ into (3.15a) and making use of (3.10) we obtain

$$\mathscr{E}^{\lambda}_{Z} = -(\mathscr{E}1)^{\lambda}{}_{\mu}Z^{\mu} - (\mathscr{E}2)^{\lambda}{}_{(\alpha)}{}^{(\beta)}Y_{\mu}{}^{(\alpha)}{}_{(\beta)}Z^{\mu} - \partial_{\tau}(\mathscr{U}^{\tau\lambda}{}_{\mu}Z^{\mu} - \partial_{\tau}(\mathscr{P}^{\tau\lambda}{}_{(\alpha)}{}^{(\beta)}Y_{\mu}{}^{(\alpha)}{}_{(\beta)}Z^{\mu}). \tag{3.15b}$$

Here, $Y_{\lambda}{}^{(\alpha)}{}_{(\beta)}\,dx^{\lambda} = (\Gamma_{\lambda}{}^{(\alpha)}{}_{(\beta)} - \zeta_{\lambda}{}^{(\alpha)}{}_{(\beta)})dx^{\lambda}$ is a tensor-valued one-form on space-time determining the difference between the physical (dynamical) connection $\Gamma_{\lambda}{}^{(\alpha)}{}_{(\beta)}$ and the auxiliary (background) connection $\zeta_{\lambda}{}^{(\alpha)}{}_{(\beta)}$ defining the covariant Lie derivative. (For details see Appendix C.) The energy-momentum vector density (3.15b) is manifestly GL(4,R)-covariant, but taking ${}_{\zeta}\mathscr{L}_Z$ as a generator of infinitesimal translations we are faced with some difficulties caused by the failure of the standard formula $[\mathscr{L}_{Z_1}, \mathscr{L}_{Z_2}] = \mathscr{L}_{[Z_1,Z_2]}$. A more detailed discussion of the problem is given in Appendix F.

From formulas (3.14) and (3.15) and the CBI, we have the following result.

Proposition 2: If matter field equations (2.19d) are satisfied, then the conservation laws $\partial_{\lambda}\mathscr{E}^{\lambda}_{Z} = 0$ and $\partial_{\lambda}\mathscr{I}^{\lambda}_{L} = 0$ are equivalent to the gravitational field equations (2.19a)–(2.19c).

It follows from the CBI that not all of field equations in the system (2.19) are independent. The following result holds.

Proposition 3: The system (2.19) is equivalent to

$$\text{(i)}\quad (\mathscr{E}1)^{\lambda}{}_{(\alpha)} = 0 \text{ and } (\mathscr{E}2)^{\lambda}{}_{(\alpha)}{}^{(\beta)} = 0 \text{ and } (\mathscr{E}\mathscr{M})_{\Sigma} = 0 \tag{3.16}$$

or

$$\text{(ii)}\quad (\mathscr{E}2)^{\lambda}{}_{(\alpha)}{}^{(\beta)} = 0 \text{ and } (\mathscr{E}3)^{(\alpha)(\beta)} = 0 \text{ and } (\mathscr{E}\mathscr{M})_{\Sigma} = 0. \tag{3.17}$$

For a special case of the Einstein gravitational Lagrangian in the GL(4,R) theory this result was proved by Trautman.[11]

IV. THE HAMILTONIAN STRUCTURE OF THE THEORY

In the papers[35,42] a general construction of the symplectic formulation of arbitrary classical field theory based on a variational principle has been presented. Following the general idea of those papers, we construct the energy-momentum function $\mathscr{E}_Z$ and symplectic two-form Ω on the space of initial data and establish the equivalence of the Hamilton equation and the Euler–Lagrange field equations.

In order to formulate the evolution problem in relativistic field theories we have to fix a slicing $\{\sigma_t\}_{t\in R}$ of space-time consisting of a one-parameter family of nonintersecting and diffeomorphic three-dimensional surfaces covering M. Let $\{f_t\}_{t\in R}$ be a one-parameter subgroup of Diff M preserving $\{\sigma_t\}$ [i.e., $f_s(\sigma_t) = \sigma_{s+t}$]. We assume that the orbits of f_t are σ transversal, i.e., the vector field Z,

$$Z(\mathbf{x}) = \frac{d}{dt}f_t(\mathbf{x})\big|_{t=0}, \tag{4.1}$$

is transversal to the surfaces σ_t.

For an arbitrary surface $\sigma\in\{\sigma_t\}$ we define the space of initial values of the field potentials and their σ-transversal derivatives $\mathscr{I}\mathscr{G}(\sigma)$ as the set of fields

$$\mathbf{x}\rightarrow F(\mathbf{x}) = (e^{(\alpha)}_{\mu}, \Gamma_{\mu}{}^{(\alpha)}{}_{(\beta)}, g_{(\alpha)(\beta)}, \phi^{\Sigma}, \partial_Z e^{(\alpha)}_{\mu}, \partial_Z\Gamma_{\mu}{}^{(\alpha)}{}_{(\beta)}, \partial_Z g_{(\alpha)(\beta)}, \partial_Z\phi^{\Sigma}). \tag{4.2}$$

The infinite-dimensional space $\mathscr{I}\mathscr{G}(\sigma)$ carries a natural geometric structure, and according to the standard rules, a vector δF tangent to $\mathscr{I}\mathscr{G}(\sigma)$ is given by the values of variations of the field potentials and the values of variations of their Z derivatives on σ:

$$\delta F = \delta e^{(\alpha)}_{\mu}\frac{\partial}{\partial e^{(\alpha)}_{\mu}} + \delta\Gamma_{\mu}{}^{(\alpha)}{}_{(\beta)}\frac{\partial}{\partial\Gamma_{\mu}{}^{(\alpha)}{}_{(\beta)}} + \delta g_{(\alpha)(\beta)}\frac{\partial}{\partial g_{(\alpha)(\beta)}} + \delta\phi^{\Sigma}\frac{\partial}{\partial\phi^{\Sigma}} + \delta\,\partial_Z e^{(\alpha)}_{\mu}\frac{\partial}{\partial\,\partial_Z e^{(\alpha)}_{\mu}} + \cdots. \tag{4.3}$$

The energy-momentum function and the symplectic two-form on the space $\mathscr{I}\mathscr{G}(\sigma)$ are defined as

$$\mathscr{E}_Z(F) = \int_\sigma \mathscr{E}^\lambda_Z(F)\eta_\lambda, \tag{4.4}$$

$$\Omega(\delta_1 F, \delta_2 F) = \int_\sigma \{\delta_1 \mathscr{U}^{\lambda\mu}{}_{(\alpha)} \wedge \delta_2 e^{(\alpha)}_\mu + \delta_1 \mathscr{P}^{\lambda\mu}{}_{(\alpha)}{}^{(\beta)} \wedge \delta_2 \Gamma_\mu{}^{(\alpha)}{}_{(\beta)} + \tfrac{1}{2}\delta_1 \mathscr{Y}^{\lambda(\alpha)(\beta)} \wedge \delta_2 g_{(\alpha)(\beta)} + \delta_1 \not{h}^\lambda{}_\Sigma \wedge \delta_2 \phi^\Sigma\}\eta_\lambda, \tag{4.5}$$

where $\eta_\lambda = \partial_\lambda \lrcorner\, dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3$, and the symbol "$\wedge$" denotes antisymmetrization with respect to the subscripts 1 and 2. The variations $\delta\mathscr{U}^{\lambda\mu}{}_{(\alpha)}$, $\delta\mathscr{P}^{\lambda\mu}{}_{(\alpha)}{}^{(\beta)}$, $\delta\mathscr{Y}^{\lambda(\alpha)(\beta)}$, $\delta\not{h}^\lambda{}_\Sigma$ are to be computed by means of (4.3) and (3.3), (3.4). The convergence of the integrals (4.4) and (4.5) can be assured by imposing appropriate boundary conditions or taking σ compact without boundary. In the present paper we deal with the latter case. Nevertheless, several of our results do not depend on this assumption.

We define the Hamiltonian vector field Y_E of the energy-momentum function $\mathscr{E}_Z$:

$$d\mathscr{E}_Z(\delta F) = -2\Omega(Y_E, \delta F) = -\Omega(Y_E \wedge \delta F). \tag{4.6}$$

Here, δF is an arbitrary (sample) vector tangent to the space $\mathscr{I}\mathscr{G}(\sigma)$. The definition of the energy-momentum vector density $\mathscr{E}^\lambda_Z$ as a Noether current of covariant translations suggests to expect that the Hamiltonian vector field Y_E of $\mathscr{E}_Z$ generates covariant translations in the space of canonical variables. That is to say, the symplectic components of the vector Y_E coincide with the covariant Lie derivatives of the corresponding canonical variables in the direction of Z:

$$\begin{aligned} Y_E = {}& {}_\zeta\mathscr{L}_Z e^{(\alpha)}_\mu \frac{\partial}{\partial e^{(\alpha)}_\mu} + {}_\zeta\mathscr{L}_Z \Gamma_\mu{}^{(\alpha)}{}_{(\beta)} \frac{\partial}{\partial \Gamma_\mu{}^{(\alpha)}{}_{(\beta)}} \\ &+ {}_\zeta\mathscr{L}_Z g_{(\alpha)(\beta)} \frac{\partial}{\partial g_{(\alpha)(\beta)}} + {}_\zeta\mathscr{L}_Z \phi^\Sigma \frac{\partial}{\partial \phi^\Sigma} \\ &+ {}_\zeta\mathscr{L}_Z \mathscr{U}^{\mu\nu}{}_{(\alpha)} \frac{\partial}{\partial \mathscr{U}^{\mu\nu}{}_{(\alpha)}} + {}_\zeta\mathscr{L}_Z \mathscr{P}^{\mu\nu}{}_{(\alpha)}{}^{(\beta)} \frac{\partial}{\partial \mathscr{P}^{\mu\nu}{}_{(\alpha)}{}^{(\beta)}} \\ &+ {}_\zeta\mathscr{L}_Z \mathscr{Y}^{\mu(\alpha)(\beta)} \frac{\partial}{\partial \mathscr{Y}^{\mu(\alpha)(\beta)}} + {}_\zeta\mathscr{L}_Z \not{h}^\mu{}_\Sigma \frac{\partial}{\partial \not{h}^\mu{}_\Sigma} + \cdots . \end{aligned} \tag{4.7}$$

The following basic result holds.

Theorem 1: The functional Hamilton equation (4.6) and the evolution postulate (4.7) are equivalent to the variational Euler–Lagrange equations (2.19).

Proof: Making use of (3.15b) we obtain

$$\begin{aligned} d\mathscr{E}^\lambda_Z(\delta F) = {}& \{ {}_\zeta\mathscr{L}_Z e^{(\alpha)}_\mu \delta\mathscr{U}^{\lambda\mu}{}_{(\alpha)} + {}_\zeta\mathscr{L}_Z \Gamma_\mu{}^{(\alpha)}{}_{(\beta)} \delta\mathscr{P}^{\lambda\mu}{}_{(\alpha)}{}^{(\beta)} \\ &+ \tfrac{1}{2}{}_\zeta\mathscr{L}_Z g_{(\alpha)(\beta)} \delta\mathscr{Y}^{\lambda(\alpha)(\beta)} + {}_\zeta\mathscr{L}_Z \phi^\Sigma \delta\not{h}^\lambda{}_\Sigma - {}_\zeta\mathscr{L}_Z \mathscr{U}^{\lambda\mu}{}_{(\alpha)} \delta e^{(\alpha)}_\mu - {}_\zeta\mathscr{L}_Z \mathscr{P}^{\lambda\mu}{}_{(\alpha)}{}^{(\beta)} \delta\Gamma_\mu{}^{(\alpha)}{}_{(\beta)} \\ &- \tfrac{1}{2}{}_\zeta\mathscr{L}_Z \mathscr{Y}^{\lambda(\alpha)(\beta)} \delta g_{(\alpha)(\beta)} - {}_\zeta\mathscr{L}_Z \not{h}^\lambda{}_\Sigma \delta\phi^\Sigma \} - \{ Z^\lambda (\mathscr{E}1)^\mu{}_{(\alpha)} \delta e^{(\alpha)}_\mu + Z^\lambda (\mathscr{E}2)^\mu{}_{(\alpha)}{}^{(\beta)} \delta\Gamma_\mu{}^{(\alpha)}{}_{(\beta)} \\ &+ \tfrac{1}{2} Z^\lambda (\mathscr{E}3)^{(\alpha)(\beta)} \delta g_{(\alpha)(\beta)} + Z^\lambda (\mathscr{E}\mathscr{M})_\Sigma \delta\phi^\Sigma \} + \partial_\mu \mathscr{B}^{\lambda\mu}, \end{aligned} \tag{4.8}$$

where

$$\begin{aligned} \mathscr{B}^{\lambda\mu} = {}& (Z^\mu \mathscr{U}^{\lambda\nu}{}_{(\alpha)} - Z^\lambda \mathscr{U}^{\mu\nu}{}_{(\alpha)} - Z^\nu \mathscr{U}^{\lambda\mu}{}_{(\alpha)}) \delta e^{(\alpha)}_\nu \\ &+ (Z^\mu \mathscr{P}^{\lambda\nu}{}_{(\alpha)}{}^{(\beta)} - Z^\lambda \mathscr{P}^{\mu\nu}{}_{(\alpha)}{}^{(\beta)} \\ &- Z^\nu \mathscr{P}^{\lambda\mu}{}_{(\alpha)}{}^{(\beta)}) \delta\Gamma_\nu{}^{(\alpha)}{}_{(\beta)} + \tfrac{1}{2}(Z^\mu \mathscr{Y}^{\lambda(\alpha)(\beta)} \\ &- Z^\lambda \mathscr{Y}^{\mu(\alpha)(\beta)}) \delta g_{(\alpha)(\beta)} + (Z^\mu \not{h}^\lambda{}_\Sigma - Z^\lambda \not{h}^\mu{}_\Sigma) \delta\phi^\Sigma \\ &- Z^\nu e^{(\alpha)}_\nu \delta\mathscr{U}^{\lambda\mu}{}_{(\alpha)} - Z^\nu Y_\nu{}^{(\alpha)}{}_{(\beta)} \delta\mathscr{P}^{\lambda\mu}{}_{(\alpha)}{}^{(\beta)}. \end{aligned}$$

The first bracket in (4.8) corresponds to the integrand in (4.5). Neglecting the boundary terms we get the formula

$$\int_\sigma Z^\lambda \{ (\mathscr{E}1)^\mu{}_{(\alpha)} \delta e^{(\alpha)}_\mu + (\mathscr{E}2)^\mu{}_{(\alpha)}{}^{(\beta)} \delta\Gamma_\mu{}^{(\alpha)}{}_{(\beta)} + \tfrac{1}{2}(\mathscr{E}3)^{(\alpha)(\beta)} \delta g_{(\alpha)(\beta)} + (\mathscr{E}\mathscr{M})_\Sigma \delta\phi^\Sigma \} \eta_\lambda = 0. \tag{4.9}$$

From transversality of Z and the free choice of $\delta e^{(\alpha)}_\mu$, $\delta\Gamma_\mu{}^{(\alpha)}{}_{(\beta)}$, $\delta g_{(\alpha)(\beta)}$, and $\delta\phi^\Sigma$, we conclude that (4.9) is equivalent to the Euler–Lagrange equations

$$(\mathscr{E}1)^\mu{}_{(\alpha)} = 0, \quad (\mathscr{E}2)^\mu{}_{(\alpha)}{}^{(\beta)} = 0,$$

$$(\mathscr{E}3)^{(\alpha)(\beta)} = 0, \quad (\mathscr{E}\mathscr{M})_\Sigma = 0.$$

It is easy to show that the analogous result is valid for the Noether current of rotations $\mathscr{I}^\lambda_L$, i.e., the vector

$$Y_I = \delta_L e^{(\alpha)}_\mu \frac{\partial}{\partial e^{(\alpha)}_\mu} + \cdots + \delta_L \mathscr{U}^{\mu\nu}{}_{(\alpha)} \frac{\partial}{\partial \mathscr{U}^{\mu\nu}{}_{(\alpha)}} + \cdots \tag{4.10}$$

is a Hamiltonian vector of the hypermomentum function $\mathscr{I}_L$ on $\mathscr{I}\mathscr{G}(\sigma)$:

$$\mathscr{I}_L(F) = \int_\sigma \mathscr{I}^\lambda_L(F)\eta_\lambda, \tag{4.11}$$

where δ_L is the operator of infinitesimal rotations given by (3.2).

Up to now we did not make any assumption about the slicing $\{\sigma_t\}$ (in general, the slices are not spacelike surfaces). The Hamiltonian approach presented in Theorem 1 is GL(4,R) covariant. For the evolution picture, however, much more important is the (3 + 1) picture for which the surfaces of the slicing are spacelike.

Let us consider only tetrads and metrics such that we have the following.

(i) $g_{(0)(0)} = \mathbf{e}_{(0)} \cdot \mathbf{e}_{(0)} < 0$.

(ii) Three-dimensional subspaces spanned by $\mathbf{e}_{(a)}$, $a = 1,2,3$, are spacelike.

(iii) All submanifolds σ_t are spacelike.

(iv) $n_{(0)} = \mathbf{e}_{(0)} \cdot \mathbf{n} < 0$, for the future directed ($\mathbf{n} \cdot \mathbf{Z} < 0$) vector field $\mathbf{n}$ orthonormal to the slicing. We look for a matrix that transforms the tetrad $(\mathbf{e}_{(\alpha)})$ into a tetrad $(\tilde{\mathbf{e}}_{(\alpha)})$ such that $\tilde{\mathbf{e}}_{(0)} = \mathbf{n}$ and $\tilde{\mathbf{e}}_{(a)}$ are tangent to σ (orthogonal to $\mathbf{n}$). Such a transformation is given up to GL(3,R) rotations. We can, however, single out the following transformations, which in some sense generalize pure Lorentz transformations (boosts).

Let us define a nonsingular matrix $[B^{(\alpha)}{}_{(\beta)}]$:

$$B^{(\alpha)}{}_{(0)} = n^{(\alpha)}, \quad B^{(0)}{}_{(b)} = n_{(b)},$$

$$B^{(a)}{}_{(b)} = \delta^{(a)}{}_{(b)}\frac{1-n_{(0)}}{1+n^{(0)}} + \frac{n^{(a)}n_{(b)}}{1+n^{(0)}},$$

$$B^{-1(0)}{}_{(\beta)} = -n_{(\beta)}, \quad B^{-1(a)}{}_{(0)} = -n^{(a)},$$

$$B^{-1(a)}{}_{(b)} = \delta^{(a)}{}_{(b)}\frac{1+n^{(0)}}{1-n_{(0)}} + \frac{n^{(a)}n_{(b)}}{1-n_{(0)}}, \tag{4.12}$$

where $\mathbf{n} = n^{(\alpha)}\mathbf{e}_{(\alpha)}$, $n_{(\beta)} = n^{(\alpha)}g_{(\alpha)(\beta)}$ [it follows from (i), (ii), (iv) that $n^{(0)} > 0$, and, of course $n_{(\alpha)}n^{(\alpha)} = -1$]. The transformation $\tilde{\mathbf{e}}_{(\alpha)} = B^{(\beta)}{}_{(\alpha)}e_{(\beta)}$ gives as we require $\tilde{\mathbf{e}}_{(0)} = n$, $\tilde{\mathbf{e}}_{(0)}\cdot\tilde{\mathbf{e}}_{(a)} = \tilde{g}_{(0)(a)} = 0$, and, in the case of the Minkowski metric $g_{(\alpha)(\beta)} = \eta_{(\alpha)(\beta)}$, it reduces to a boost transformation.

By means of the field of matrices (4.12) we transform tetrads, metrics, connections and tensor fields:

$$\tilde{e}^{\mu}_{(\alpha)} = B^{(\beta)}{}_{(\alpha)}e^{\mu}_{(\beta)}, \quad \tilde{e}^{(\alpha)}_{\mu} = B^{-1(\alpha)}{}_{(\beta)}e^{(\beta)}_{\mu},$$

$$\tilde{g}_{(\alpha)(\beta)} = B^{(\epsilon)}{}_{(\alpha)}B^{(\tau)}{}_{(\beta)}g_{(\epsilon)(\tau)},$$

$$\tilde{\phi}^{(\alpha)}{}_{(\beta)} = B^{-1(\alpha)}{}_{(\epsilon)}B^{(\tau)}{}_{(\beta)}\phi^{(\epsilon)}{}_{(\tau)},$$

$$\tilde{\Gamma}_{\mu}{}^{(\alpha)}{}_{(\beta)} = B^{-1(\alpha)}{}_{(\epsilon)}B^{(\tau)}{}_{(\beta)}\Gamma_{\mu}{}^{(\epsilon)}{}_{(\tau)} + B^{-1(\alpha)}{}_{(\tau)}\partial_{\mu}B^{(\tau)}{}_{(\beta)}, \text{ etc.} \tag{4.13}$$

From now on we use coordinate systems in M consistent with the slicing, i.e., $\sigma_t = \{x \in M: x^0 = t\}$. Such special coordinates in M are subject to the following transformations:

$$x^{0'} = x^{0'}(x^0), \quad x^{k'} = x^{k'}(x^0, x^s). \tag{4.14}$$

The quantities $\tilde{e}^{(\alpha)}_{\mu}$, $\tilde{g}_{(\alpha)(\beta)}$, $\tilde{\Gamma}_{\mu}{}^{(\alpha)}{}_{(\beta)}$, $\tilde{\phi}^{\Sigma}$, $\tilde{\mathscr{U}}^{\mu\nu}{}_{(\alpha)}$, $\tilde{\mathscr{P}}^{\mu\nu}{}_{(\alpha)}{}^{(\beta)}$, etc. are tensors with respect to coordinate transformations in M. We can decompose them into normal and tangential (with respect to σ_t) parts by means of the "bar" operation.[30,35]

The lapse function and the shift vector of the slicing $\{\sigma_t\}$ are

$$N = (-g^{00})^{-1/2}, \quad N^k = \bar{g}^{ks}g_{s0}, \tag{4.15a}$$

where $[\bar{g}^{ks}]$ is the inverse matrix of $[g_{ks}]$, $k,s = 1,2,3$. Let $A^{\bar{\mu}}{}_{\nu}$ be a 4×4 matrix given by

$$A^{\bar{0}}{}_0 = N, \quad A^{\bar{0}}{}_s = 0, \quad A^{\bar{k}}{}_0 = N^k, \quad A^{\bar{k}}{}_s = \delta^k{}_s,$$

$$(A^{-1})^0{}_{\bar{0}} = N^{-1}, \quad (A^{-1})^k{}_{\bar{0}} = -N^k/N,$$

$$(A^{-1})^0{}_{\bar{s}} = 0, \quad (A^{-1})^k{}_{\bar{s}} = \delta^k{}_s. \tag{4.15b}$$

The "bar" operation for a tensor density of weight r on M is defined by

$$\overline{\psi}^{\Sigma\mu_1\cdots\mu_k}_{\nu_1\cdots\nu_s} = N^{-r}A^{\bar{\mu}_1}{}_{\alpha_1}\cdots A^{\bar{\mu}_k}{}_{\alpha_k}(A^{-1})^{\beta_1}{}_{\bar{\nu}_1}\cdots(A^{-1})^{\beta_s}{}_{\bar{\nu}_s}\psi^{\Sigma\alpha_1\cdots\alpha_k}_{\beta_1\cdots\beta_s}. \tag{4.16}$$

The composition of the "$^{-}$" and "$^{\sim}$" operations, denoted by the caret "$^{\wedge}$," gives the $(3+1)$ variables of our theory. Especially important in further considerations are covector fields $\hat{e}^{(a)}_k$, $\hat{\Gamma}_k{}^{(\alpha)}{}_{(\beta)}$, vector densities $\hat{\mathscr{U}}^{0k}{}_{(\alpha)}$, $\hat{\mathscr{P}}^{0k}{}_{(\alpha)}{}^{(\beta)}$, scalar functions $\hat{g}_{(a)(b)}$, $\hat{\phi}^{\Sigma}$, and scalar densities $\hat{\mathscr{Y}}^{0(a)(b)}$, $\hat{\lambda}^0{}_{\Sigma}$ on σ_t. It is easy to see that the $(3+1)$ positions have clear geometrical meaning: $(\hat{e}^{(a)}_k)$ is a field of triads of covectors tangent to σ_t, $(\hat{g}_{(a)(b)})$ is the metric tensor on σ_t induced by the metric $(g_{(\alpha)(\beta)})$ on M, $(\hat{\Gamma}_k{}^{(a)}{}_{(b)})$ are anholonomic components [with respect to $(\hat{e}^{(a)}_k)$] of the connection ${}^3\Gamma$ induced on σ_t by the connection Γ on M [cf. Ref. 5(b)].

Remarks: Applying the "bar" operation to tensor densities on M we have $(\det[g_{\mu\nu}])^{-} = \det[\bar{g}_{\mu\nu}] = -\det[\bar{g}_{ks}] = -\bar{g}$. Applying the "caret" operation to tetrads and metric we obtain $\hat{e}^{(0)}_0 = 1$, $\hat{e}^{(0)}_s = 0 = \hat{e}^{(a)}_0$, $\hat{g}_{(0)(0)} = -1$, $\hat{g}_{(0)(a)} = 0$. It means that taking "caret" variables we lose some information about geometric configurations. Therefore apart of $\hat{e}^{(a)}_k$ and $\hat{g}_{(a)(b)}$ we have to take into account the lapse N, the shift N^k and seven rotational coefficients $n^{(\alpha)}$, $n_{(b)}$.

Now we rewrite the energy-momentum function and the symplectic two-form in terms of $(3+1)$ variables. For σ compact without boundary we have

$$\mathscr{E}_Z(F) = -\int_{\sigma}\{(\hat{\mathscr{E}}1)^0{}_{\mu}\bar{Z}^{\mu} + (\hat{\mathscr{E}}2)^0{}_{(\alpha)}{}^{(\beta)}\hat{Y}_{\mu}{}^{(\alpha)}{}_{(\beta)}\bar{Z}^{\mu}\}\eta_0, \tag{4.17}$$

$$\begin{aligned}\Omega(\delta_1F,\delta_2F) = \int_{\sigma}\{&\delta_1\hat{\mathscr{U}}^{0k}{}_{(a)}\wedge\delta_2\hat{e}^{(a)}_k + \delta_1\hat{\mathscr{P}}^{0k}{}_{(\alpha)}{}^{(\beta)}\wedge\delta_2\hat{\Gamma}_k{}^{(\alpha)}{}_{(\beta)}\\ &+ \delta_1\hat{\lambda}^0{}_{\Sigma}\wedge\delta_2\hat{\phi}^{\Sigma} + \tfrac{1}{2}\delta_1\hat{\mathscr{Y}}^{0(a)(b)}\wedge\delta_2\hat{g}_{(a)(b)}\\ &+ \delta_1 m_{(a)}\wedge\delta_2 n^{(a)} + \delta_1 m^{(a)}\wedge\delta_2 n_{(a)}\}\eta_0,\end{aligned} \tag{4.18}$$

where $m_{(a)}$ and $m^{(a)}$ are linear combinations of $(\hat{\mathscr{E}}2)^0{}_{(\alpha)}{}^{(\beta)}$ (see Appendix A), and $\eta_0 = dx^1_{\wedge}dx^2_{\wedge}dx^3$.

Remarks: We choose $n^{(a)}$ and $n_{(a)}$ as independent variables because $n_{(0)}$ can be determined from the condition $n^{(\alpha)}n_{(\alpha)} = -1$. The integrals of a total divergence appear in (4.17) and (4.18) if σ is not compact without boundary.

The variables appearing in (4.18) are called the $(3+1)$-symplectic variables of the theory under considerations. Similarly, as in Ref. 30, we can prove that the lhs of the field equations $(\hat{\mathscr{E}}1)^0{}_{(\alpha)} = 0$, $(\hat{\mathscr{E}}2)^0{}_{(\alpha)}{}^{(\beta)} = 0$ depend only on symplectic variables and their spatial derivatives (see Appendix B).

The Hamilton equation (4.6) and the evolution postulate (4.7) can be formulated in terms of $(3+1)$-geometric objects. To make corresponding formulas relatively simple we have to assume the following.

(i) Transformations of M generated by the vector field $\mathbf{Z}$ preserve the slicing, i.e.,

$$Z^0 = Z^0(x^0), \quad Z^k = Z^k(x^0, x^s). \tag{4.19a}$$

(ii) The auxiliary connection $\zeta_{\mu}{}^{(\alpha)}{}_{(\beta)}$ is consistent with the slicing, i.e.,

$${}_{\zeta}\mathscr{D}_{\mu}n^{(\alpha)} = 0, \quad {}_{\zeta}\mathscr{D}_{\mu}n_{(\alpha)} = 0$$

or

$$\hat{\zeta}_{\mu}{}^{(0)}{}_{(\beta)} = 0, \quad \hat{\zeta}_{\mu}{}^{(\alpha)}{}_{(0)} = 0. \tag{4.19b}$$

If the consistency conditions (4.19) hold then the σ-parallel parts of the GL(4,R)-covariant Lie derivatives of the $(3+1)$-symplectic variables consist of two terms: the GL(3,R)-covariant time derivative ${}^3_{\zeta}\mathscr{D}_0$ and the GL(3,R)-covariant σ-intrinsic Lie derivative ${}^3_{\zeta}\mathscr{L}_Z$ with respect to the σ-parallel part ${}^{\|}Z$ of the vector field Z (cf. Appendix C). The evolution postulate (4.7) reads

$$Y_E = (\bar{Z}^0\,{}^3_{\zeta}\mathscr{D}_0\hat{e}^{(a)}_k + {}^3_{\zeta}\mathscr{L}_Z\hat{e}^{(a)}_k)\frac{\partial}{\partial\hat{e}^{(a)}_k} + \cdots + (\bar{Z}^0\,{}^3_{\zeta}\mathscr{D}_0 m^{(a)} + {}^3_{\zeta}\mathscr{L}_Z m^{(a)})\frac{\partial}{\partial m^{(a)}} + \cdots . \tag{4.20}$$

In virtue of (4.19b), there are no terms proportional to $\partial/\partial n^{(\alpha)}$, $\partial/\partial n_{(a)}$ in the above formula.

We have a GL(3,R)-covariant version of the Theorem 1.

Theorem 2: The Hamilton equation (4.6) and the evolution postulate (4.20) are equivalent to the Euler–Lagrange equations (2.19).

Proof: Following the proof of Theorem 1 and taking into account that variations $\delta\hat{e}_k^{(a)}$, $\delta\hat{\Gamma}_\mu{}^{(\alpha)}{}_{(\beta)}$, $\delta\hat{g}_{(a)(b)}$, $\delta\hat{\phi}^\Sigma$, δN, δN^k, $\delta n^{(\alpha)}$, $\delta n_{(a)}$ are arbitrary, we obtain the following equations:

$$(\hat{\mathscr{E}}1)^0{}_{(\alpha)}=0,\quad (\hat{\mathscr{E}}2)^0{}_{(\alpha)}{}^{(\beta)}=0, \tag{4.21a}$$

$$(\hat{\mathscr{E}}1)^s{}_{(a)}=0,\quad (\hat{\mathscr{E}}2)^s{}_{(\alpha)}{}^{(\beta)}=0,$$

$$(\hat{\mathscr{E}}3)^{(a)(b)}=0,\quad (\hat{\mathscr{E}}\mathscr{M})_\Sigma=0, \tag{4.21b}$$

$$\tfrac{3}{\zeta}\mathscr{D}_0\, m_{(a)}=0,\quad \tfrac{3}{\zeta}\mathscr{D}_0\, m^{(a)}=0. \tag{4.21c}$$

Making use of the above equations, contracted Bianchi identities (3.11), (3.12) and explicit form of $m_{(\alpha)}$, $m^{(a)}$, we get seven missing equations $(\hat{\mathscr{E}}1)^s{}_{(0)}=0$ and $(\hat{\mathscr{E}}3)^{(0)(\alpha)}=0$.

The Hamiltonian dynamics gives us 20 constraint equations (4.21a) for the initial data F on σ, and the system (4.21b) and (4.21c) of dynamical equations for (3 + 1)-symplectic momenta. The explicit form of these equations in terms of corresponding GL(3,R)-covariant operators is given in Appendix B. It is well known that there is a deep connection between the contracted Bianchi identities and the problem of the initial value formulation of the theory. In fact, CBI ensure the consistency of the Hamiltonian formulation.

Proposition 4: If the constraints (4.21a) hold for $x^0=0$, the dynamical equations (4.21b) and $m_{(\alpha)}=0$, $m^{(a)}=0$ hold for all x^0, then (4.21a) also hold for all x^0.

Remark: In the above proposition the conditions $m_{(\alpha)}=0$, $m^{(a)}=0$ may be replaced by $(\hat{\mathscr{E}}2)^0{}_{(\alpha)}{}^{(0)}=0$ and $(\hat{\mathscr{E}}2)^0{}_{(0)}{}^{(a)}=0$ or by $(\hat{\mathscr{E}}1)^s{}_{(0)}=0$ and $(\hat{\mathscr{E}}3)^{(0)(\alpha)}=0$.

We make the following observation. In the four-covariant picture the complete set of variational field equations (2.19) is equivalent to the reduced sets (3.16) or (3.17). That is, it is sufficient to take into account either the equations $(\hat{\mathscr{E}}1)^\mu{}_{(\alpha)}=0$ or the equations $(\hat{\mathscr{E}}3)^{(\alpha)(\beta)}=0$. In the (3 + 1) picture, however, we must consider the equations $(\hat{\mathscr{E}}1)^k{}_{(a)}=0$ as well as $(\hat{\mathscr{E}}3)^{(a)(b)}=0$. The reason is that in the Hamiltonian three-covariant approach the constraints (4.21a) are *a priori* satisfied only on the initial surface and the complete set (4.21b) is necessary to ensure the time maintenance of the Hamiltonian constraints (4.21a).

In a general case, the gravitational field variables in a GL(4,R)-gauge theory of gravity may be divided into three classes: (A) dynamical symplectic variables $\hat{\mathscr{U}}^{0k}{}_{(a)}$, $\hat{e}_k^{(a)}$, $\hat{\mathscr{P}}^{0k}{}_{(\alpha)}{}^{(\beta)}$, $\hat{\Gamma}_k{}^{(\alpha)}{}_{(\beta)}$, $\hat{\mathscr{V}}^{0(a)(b)}$, $\hat{g}_{(a)(b)}$; (B) 20 gravitational gauge variables N, N^k, $n^{(\alpha)}$, $n_{(a)}$, $\hat{\Gamma}_0{}^{(a)}{}_{(b)}$; and (C) nondynamical variables $\hat{\Gamma}_0{}^{(\alpha)}{}_{(0)}$, $\hat{\Gamma}_0{}^{(0)}{}_{(\beta)}$. The matter dynamical symplectic variables are $\hat{\pi}^0{}_\Sigma$ and $\hat{\phi}^\Sigma$. For degenerate matter Lagrangians matter gauge variables may also appear, e.g., in Maxwell's electrodynamics, as well as nondynamical ones, e.g., Einstein–Cartan theory with matter fields.[31(b)] In the present paper we do not discuss the problem of matter variables more profoundly. Many interesting results concerning this subject were presented in Ref. 31(b).

Let us return to the gravitational variables. The quantities of the set (A) are called dynamical variables because we have corresponding dynamical equations (4.21b) that govern their evolution [cf. (B5)–(B7)]. The quantities of the set (B) are called gauge variables corresponding to the action of Diff M and loc GL(4,R). The group Diff M acts transitively in the space of N, N^k and the local GL(4,R) group acts transitively in the space of $n^{(\alpha)}$, $n_{(a)}$, $\hat{\Gamma}_0{}^{(a)}{}_{(b)}$. That is to say, under the actions of these groups the variables of the set (B) can be transformed to arbitrary beforehand fixed values (at least locally). Moreover, the symplectic constraints (4.21a) do not depend on gauge variables. Therefore the gauge variables may be given arbitrarily on the whole space-time.

For the nondynamical variables (C) we have no equations governing their evolution. These quantities may cause essential difficulties in the dynamical analysis of the theory in question. In Sec. V and VI we show how to solve this problem for the Einstein gravitational Lagrangian. For other Lagrangians this question should be treated individually.

Only for canonically regular gravitational Lagrangians, that is, for Lagrangians assuring a one-to-one correspondence between the symplectic momenta and the time derivatives of the symplectic positions all the variables in (A) are dynamical. For canonically degenerate Lagrangians some quantities in the set (A) cease to be dynamical and should be classified either as gauge variables or as nondynamical ones. Similarly, some of the nondynamical variables become gauge variables. In Secs. V and VI we show that for the Einstein Lagrangian (with the presence of a matter field) $\hat{\Gamma}_k{}^{(a)}{}_{(b)}$ becomes nondynamical ($\hat{\mathscr{U}}^{0k}{}_{(a)}$ and $\hat{\mathscr{V}}^{0(a)(b)}$ are also nondynamical but trivial). The nondynamical variables $\hat{\Gamma}_k{}^{(a)}{}_{(b)}$, $\hat{\Gamma}_0{}^{(a)}{}_{(0)}$, $\hat{\Gamma}_0{}^{(0)}{}_{(b)}$ can be eliminated from the field equations and $\hat{\Gamma}_\lambda{}^{(0)}{}_{(0)}$ are four new gauge variables.

V. THE EINSTEIN GRAVITATIONAL LAGRANGIAN IN THE METRIC-AFFINE THEORY

We have

$$\mathscr{L}_g=\sqrt{-g}R,$$

$$\sqrt{-g}=(-\det[g_{\mu\nu}])^{1/2}=\det[e_\mu^{(\alpha)}](-\det[g_{(\alpha)(\beta)}])^{1/2}, \tag{5.1}$$

$$\mathscr{P}^{\mu\nu}{}_{(\alpha)}{}^{(\beta)}=\sqrt{-g}(e^\mu_{(\alpha)}e^{(\beta)\nu}-e^\nu_{(\alpha)}e^{(\beta)\mu}),$$

$$\mathscr{U}^{\mu\nu}{}_{(\alpha)}=0,\quad \mathscr{V}^{\mu(\alpha)(\beta)}=0, \tag{5.2}$$

$$\hat{\mathscr{P}}^{0k}{}_{(0)}{}^{(0)}=0,\quad \hat{\mathscr{P}}^{0k}{}_{(a)}{}^{(0)}=\sqrt{\bar{g}}\hat{e}^k_{(a)}=\tfrac{1}{2}\hat{e}^k_{(a)},$$

$$\hat{\mathscr{P}}^{0k}{}_{(a)}{}^{(b)}=0,$$

$$\hat{\mathscr{P}}^{0k}{}_{(0)}{}^{(b)}=\tfrac{1}{2}\hat{e}^k_{(c)}\hat{g}^{(c)(b)}, \tag{5.3}$$

$$\hat{\mathscr{P}}^{rs}{}_{(a)}{}^{(b)}=\tfrac{1}{2}(\hat{e}^r_{(a)}\hat{e}^{(b)s}-\hat{e}^s_{(a)}\hat{e}^{(b)r}),$$

$$\hat{\mathscr{P}}^{rs}{}_{(0)}{}^{(0)}=0,\quad \hat{\mathscr{P}}^{rs}{}_{(a)}{}^{(0)}=0,\quad \hat{\mathscr{P}}^{rs}{}_{(0)}{}^{(b)}=0.$$

The symplectic two-form of the gravitational field reads

$$\begin{aligned}\Omega(\delta_1F,\delta_2F)&=\int_\sigma\{\delta_1(\tfrac{1}{2}\hat{e}^k_{(a)})\wedge\delta_2\hat{\Gamma}_k{}^{(a)}{}_{(0)}\\&\quad+\delta_1(\tfrac{1}{2}\hat{e}^k_{(c)}\hat{g}^{(c)(b)})\wedge\delta_2\hat{\Gamma}_k{}^{(0)}{}_{(b)}\\&\quad+\delta_1 m_{(\alpha)}\wedge\delta_2 n^{(\alpha)}+\delta_1 m^{(a)}\wedge\delta_2 n_{(a)}\}\eta_0.\end{aligned} \tag{5.4}$$

Unfortunately, in virtue of the symmetry properties of $\hat{g}^{(a)(b)}$, the quantities $\hat{e}^k_{(a)}$ and $\hat{e}^k_{(c)}\hat{g}^{(c)(b)}$ are not independent. We rewrite (5.4) in the following form:

$$\Omega(\delta_1 F,\delta_2 F) = \int_\sigma \{\delta_1 \check{e}^k_{(a)} \wedge \delta_2 \hat{K}^{(a)}_k + \delta_1 \hat{g}^{(a)(b)} \wedge \delta_2 \hat{\mathscr{H}}_{(a)(b)} + m \text{ terms}\}\eta_0. \quad (5.5)$$

Here,

$$\check{e}^k_{(a)} = 2\sqrt{\bar{g}}\hat{e}^k_{(a)}, \quad \hat{K}^{(a)}_k = \tfrac{1}{2}(\hat{\Gamma}_k{}^{(a)}{}_{(0)} + \hat{\Gamma}_k{}^{(0)}{}_{(c)}\hat{g}^{(c)(a)}),$$
$$\hat{\mathscr{H}}_{(a)(b)} = \tfrac{1}{2}\sqrt{\bar{g}}(\hat{\Gamma}_k{}^{(0)}{}_{(a)}\hat{e}^k_{(b)} + \hat{\Gamma}_k{}^{(0)}{}_{(b)}\hat{e}^k_{(a)}). \quad (5.6)$$

We have no primary constraints between the gravitational symplectic variables

$$\check{e}^k_{(a)}, \hat{K}^{(a)}_k, \hat{g}^{(a)(b)}, \hat{\mathscr{H}}_{(a)(b)}. \quad (5.7)$$

Remark: We have the following relations:

$$\bar{g} = \det[\bar{g}_{rs}] = (\det[\hat{e}^{(a)}_k])^2 \det[\hat{g}_{(a)(b)}],$$
$$\det[\check{e}^k_{(a)}] = 8\bar{g}\det[\hat{g}_{(a)(b)}]^{1/2}. \quad (5.8)$$

Therefore, having $\check{e}^k_{(a)}$ and $\hat{g}^{(a)(b)}$ we are able to determine $\hat{e}^k_{(a)}$.

The field equations: We have

$$(\mathscr{E}2)^\lambda{}_{(\alpha)}{}^{(\beta)} = \hslash^\lambda{}_{(\alpha)}{}^{(\beta)} - \mathscr{D}_\tau \mathscr{P}^{\tau\lambda}{}_{(\alpha)}{}^{(\beta)} = 0. \quad (5.9)$$

It follows from (5.2) that $\mathscr{P}^{\tau\lambda}{}_{(\alpha)}{}^{(\alpha)} = 0$ and therefore

$$h^\lambda{}_{(\alpha)}{}^{(\alpha)} = 0. \quad (5.10)$$

In the $(3+1)$ picture we have the following system:

$$\hat{\hslash}^0{}_{(0)}{}^{(0)} - \check{e}^s_{(c)}\hat{K}^{(c)}_s + 2\hat{\mathscr{H}}_{(c)(d)}\hat{g}^{(c)(d)} = 0, \quad (5.11a)$$
$$\hat{\hslash}^0{}_{(a)}{}^{(0)} + \tfrac{1}{2}{}^3\mathscr{D}_s\check{e}^s_{(a)} + \tfrac{1}{2}\hat{\Gamma}_s{}^{(0)}{}_{(0)}\check{e}^s_{(a)} = 0, \quad (5.11b)$$
$$\hat{\hslash}^0{}_{(0)}{}^{(b)} + \tfrac{1}{2}{}^3\mathscr{D}_s\check{e}^{(b)s} - \tfrac{1}{2}\hat{\Gamma}_s{}^{(0)}{}_{(0)}\check{e}^{(b)s} = 0, \quad (5.11c)$$
$$\hat{\hslash}^0{}_{(a)}{}^{(b)} + \check{e}^s_{(a)}\hat{K}^{(b)}_s - 2\hat{\mathscr{H}}_{(c)(a)}\hat{g}^{(c)(b)} = 0, \quad (5.11d)$$
$$\hat{\hslash}^k{}_{(0)}{}^{(0)} + \tfrac{1}{2}\check{e}^k_{(c)}\hat{\Gamma}_0{}^{(c)}{}_{(0)} - \tfrac{1}{2}\check{e}^{(c)k}\hat{\Gamma}_0{}^{(0)}{}_{(c)} = 0, \quad (5.11e)$$
$$\hat{\hslash}^k{}_{(a)}{}^{(0)} - \tfrac{1}{2}{}^3\mathscr{D}_0\check{e}^k_{(a)} - \tfrac{1}{2}\check{e}^k_{(a)}\hat{\Gamma}_0{}^{(0)}{}_{(0)} - \tfrac{1}{2}\hat{\Gamma}_s{}^{(0)}{}_{(c)}(\check{e}^s_{(a)}\check{e}^{(c)k} - \check{e}^k_{(a)}\check{e}^{(c)s}) = 0, \quad (5.11f)$$
$$\hat{\hslash}^k{}_{(0)}{}^{(b)} - \tfrac{1}{2}{}^3\mathscr{D}_0(\check{e}^k_{(c)}\hat{g}^{(c)(b)}) + \tfrac{1}{2}\check{e}^{(b)k}\hat{\Gamma}_0{}^{(0)}{}_{(0)} + \tfrac{1}{2}\hat{\Gamma}_s{}^{(c)}{}_{(0)}(\check{e}^s_{(c)}\check{e}^{(b)k} - \check{e}^k_{(c)}\check{e}^{(b)s}) = 0, \quad (5.11g)$$
$$\hat{\hslash}^k{}_{(a)}{}^{(b)} - \tfrac{1}{2}{}^3\mathscr{D}_s(\check{e}^s_{(a)}\check{e}^{(b)k} - \check{e}^k_{(a)}\check{e}^{(b)s}) - \tfrac{1}{2}\partial_s \ln N(\check{e}^s_{(a)}\check{e}^{(b)k} - \check{e}^k_{(a)}\check{e}^{(b)s}) + \tfrac{1}{2}\check{e}^{(b)k}\hat{\Gamma}_0{}^{(0)}{}_{(a)} - \tfrac{1}{2}\check{e}^k_{(a)}\hat{\Gamma}_0{}^{(b)}{}_{(0)} = 0. \quad (5.11h)$$

Equations (5.11d) are symplectic constraints, Eq. (5.11a) follow from (5.11d) and (5.10). Equations (5.11b), (5.11c), and (5.11h) enable us to compute $\hat{\Gamma}_0{}^{(0)}{}_{(a)}$, $\hat{\Gamma}_0{}^{(b)}{}_{(0)}$, and $\hat{\Gamma}_k{}^{(a)}{}_{(b)}$. We get

$$\sqrt{\bar{g}}\hat{\Gamma}_0{}^{(0)}{}_{(a)} = \tfrac{1}{2}\hat{e}^s_{(a)}\,\partial_s \ln N + \tfrac{1}{2}\hat{\hslash}_{(a)(0)}{}^{(0)} + \tfrac{1}{4}\hat{\hslash}^{(s)}{}_{(s)(a)} - \tfrac{1}{4}\hat{\hslash}^{(s)}{}_{(a)(s)} - \tfrac{1}{4}\hat{\hslash}^0{}_{(a)}{}^{(0)} - \tfrac{1}{4}\hat{\hslash}^0{}_{(0)(a)}, \quad (5.12a)$$
$$\sqrt{\bar{g}}\hat{\Gamma}_0{}^{(b)}{}_{(0)} = \sqrt{\bar{g}}\hat{\Gamma}_0{}^{(0)}{}_{(a)}\hat{g}^{(a)(b)} - \hat{\hslash}^{(b)}{}_{(0)}{}^{(0)}, \quad (5.12b)$$
$$\sqrt{\bar{g}}\hat{\Gamma}_k{}^{(a)}{}_{(b)} = \sqrt{\bar{g}}\hat{\gamma}_k{}^{(a)}{}_{(b)} + (\sqrt{\bar{g}}\hat{\Gamma}_k{}^{(0)}{}_{(0)} - \tfrac{1}{4}\hat{\hslash}^0{}_{(0)k} + \tfrac{1}{4}\hat{\hslash}^0{}_k{}^{(0)} + \tfrac{1}{4}\hat{\hslash}_{k(0)}{}^{(0)}) \times\delta^{(a)}{}_{(b)} + \tfrac{1}{2}\hat{\hslash}^{(a)}{}_{(b)k} + \tfrac{1}{2}\hat{\hslash}_{(b)k}{}^{(a)} - \tfrac{1}{2}\hat{\hslash}_k{}^{(a)}{}_{(b)} + \tfrac{1}{4}\hat{\hslash}^0{}_{(b)}{}^{(0)}\hat{e}^{(a)}_k + \tfrac{1}{4}\hat{\hslash}^0{}_{(0)(b)}\hat{e}^{(a)}_k + \tfrac{1}{4}\hat{e}^{(a)}_k(\hat{\hslash}^s{}_{(s)(b)} - \hat{\hslash}^{(s)}{}_{(b)(s)}) - \tfrac{1}{4}\hat{e}_{(b)k}(\hat{\hslash}^{0(a)(0)} + \hat{\hslash}^0{}_{(0)}{}^{(a)} + \hat{\hslash}^{(s)}{}_{(s)}{}^{(a)} - \hat{\hslash}^{(s)(a)}{}_{(s)}). \quad (5.12c)$$

Remark: In the process of calculations we use the following formulas:

$${}^3\mathscr{D}_s\hat{e}^{(a)}_k = {}^3r_s{}^{(a)}{}_{(c)}\hat{e}^{(c)}_k = {}^4\hat{r}_s{}^{(a)}{}_{(c)}\hat{e}^{(c)}_k,$$
$${}^3\mathscr{D}_s\hat{e}^k_{(b)} = -{}^3r_s{}^{(c)}{}_{(b)}\hat{e}^k_{(c)}, \quad {}^3\mathscr{D}_s\hat{e}^{(b)s} = {}^3r_s{}^{(b)}{}_{(c)}\hat{e}^{(c)s},$$
$${}^3\mathscr{D}_s\hat{g}_{(a)(b)} = -({}^3r_{s(a)(b)} + {}^3r_{s(b)(a)}).$$

We are not able to compute $\hat{\Gamma}_\lambda{}^{(0)}{}_{(0)}$ from Eqs. (5.11) (!). Equations (5.11f) and (5.11g) give us the dynamics of the symplectic variables $\check{e}^k_{(a)}$ and $\hat{g}^{(a)(b)}$. We get

$$\tfrac{1}{2}{}^3\mathscr{D}_0\check{e}^k_{(a)} + \hat{\mathscr{H}}_{(a)(c)}\check{e}^{(c)k} - \hat{\mathscr{H}}_{(c)(d)}\hat{g}^{(c)(d)}\check{e}^k_{(a)} + \tfrac{1}{2}\hat{\Gamma}_{[(a)}{}^{(0)}{}_{(c)]}\check{e}^{(c)k} + \tfrac{1}{2}\check{e}^k_{(a)}\hat{\Gamma}_0{}^{(0)}{}_{(0)} - \hat{\hslash}^k{}_{(a)}{}^{(0)} = 0, \quad (5.13a)$$
$$\sqrt{\bar{g}}\,{}^3\mathscr{D}_0\hat{g}^{(a)(b)} - \check{e}^s_{(c)}\hat{K}^{(c)}_s\hat{g}^{(a)(b)} + \sqrt{\bar{g}}(\hat{K}^{(a)(b)} + \hat{K}^{(b)(a)}) + 2\hat{\mathscr{H}}_{(c)(d)}\hat{g}^{(c)(d)}\hat{g}^{(a)(b)} - 2\hat{\mathscr{H}}^{(a)(b)} - 2\sqrt{\bar{g}}\hat{g}^{(a)(b)}\hat{\Gamma}_0{}^{(0)}{}_{(0)} + \hat{\hslash}^{((a)(b))(0)} - \hat{\hslash}^{((a)}{}_{(0)}{}^{(b))} = 0. \quad (5.13b)$$

Here, $\hat{K}^{(a)(b)} = \hat{K}^{(a)}_s\hat{e}^{(b)s}$.

Let us observe that from 18 equations (5.11f) and (5.11g) we have obtained only 15 equations (5.13). Therefore, we should have additional information from (5.11f) and (5.11g). In fact, we are able to compute the skew-symmetric part of $\hat{\Gamma}_{(a)}{}^{(0)}{}_{(b)}$. We get

$$\sqrt{\bar{g}}\hat{\Gamma}_{[(a)}{}^{(0)}{}_{(b)]} = -\tfrac{1}{2}(\hat{\hslash}^0{}_{[(a)(b)]} + \hat{\hslash}_{[(a)}{}^{(0)}{}_{(b)]} + \hat{\hslash}_{[(a)(b)]}{}^{(0)}). \quad (5.14)$$

Finally, we observe that Eqs. (5.11e) follow from (5.11h) and (5.10), and the analysis of the system (5.11) is complete.

Remark: Formulas (5.12) are special cases of the results presented in Ref. 5(b).

The Einstein equations are

$$(\mathscr{E}1)^\lambda{}_{(\alpha)} = {}_{cg}\mathscr{T}^\lambda{}_{(\alpha)} + {}_{cm}\mathscr{T}^\lambda{}_{(\alpha)} = 0, \quad (5.15)$$

where

$${}_{cg}\mathscr{T}^\lambda{}_{(\alpha)} = \sqrt{-g}Re^\lambda_{(\alpha)} - \sqrt{-g}(R^\lambda{}_{(\alpha)} + R^\lambda{}_{(\tau)(\alpha)}{}^{(\tau)}).$$

In the $(3+1)$ decomposition, Eqs. (5.15) can be written in the following form:

$$(\hat{\mathscr{E}}1)^0{}_{(0)} = \sqrt{\bar{g}}\hat{R} - \sqrt{\bar{g}}(\hat{R}^0{}_{(0)} + \hat{R}^0{}_{(\tau)(0)}{}^{(\tau)}) + {}_{cm}\hat{\mathscr{T}}^0{}_{(0)} = 0, \quad (5.16a)$$
$$(\hat{\mathscr{E}}1)^0{}_{(a)} = -\sqrt{\bar{g}}(\hat{R}^0{}_{(a)} + \hat{R}^0{}_{(\tau)(a)}{}^{(\tau)}) + {}_{cm}\hat{\mathscr{T}}^0{}_{(a)} = 0, \quad (5.16b)$$
$$(\hat{\mathscr{E}}1)^k{}_{(a)} - \tfrac{1}{2}\hat{e}^k_{(a)}\operatorname{tr}(\hat{\mathscr{E}}1) = -\sqrt{\bar{g}}(\hat{R}^k{}_{(a)} + \hat{R}^k{}_{(\tau)(a)}{}^{(\tau)}) + {}_{cm}\hat{\mathscr{T}}^k{}_{(a)} - \tfrac{1}{2}\hat{e}^k_{(a)}\operatorname{tr}{}_{cm}\hat{\mathscr{T}} = 0, \quad (5.16c)$$
$$(\hat{\mathscr{E}}1)^k{}_{(0)} = -\sqrt{\bar{g}}(\hat{R}^k{}_{(0)} + \hat{R}^k{}_{(\tau)(0)}{}^{(\tau)}) + {}_{cm}\hat{\mathscr{T}}^k{}_{(0)} = 0. \quad (5.16d)$$

For (5.16a) and (5.16b) we get explicitly

$$
\begin{aligned}
&\sqrt{\bar g}\,{}^3R + 2\hat K_s{}^{(a)}\hat e^s_{(a)}\hat{\mathcal H}_{(c)(d)}\hat g^{(c)(d)} \\
&\quad - (\bar g)^{-1/2}(\hat{\mathcal H}_{(c)(d)}\hat g^{(c)(d)})^2 \\
&\quad - 2\hat K^{(a)(b)}(\hat{\mathcal H}_{(a)(b)} + \sqrt{\bar g}\hat\Gamma_{[(a)}{}^{(0)}{}_{(b)]}) \\
&\quad + (\bar g)^{-1/2}\hat{\mathcal H}_{(c)(d)}\hat{\mathcal H}^{(c)(d)} \\
&\quad + \sqrt{\bar g}\hat\Gamma_{[(c)}{}^{(0)}{}_{(d)]}\hat\Gamma^{[(c)}{}_{(0)}{}^{(d)]} + {}_{cm}\hat{\mathcal T}^0{}_{(0)} = 0,
\end{aligned} \tag{5.16a'}
$$

$$
\begin{aligned}
&\sqrt{\bar g}\hat e^r_{(c)}\hat e^s_{(a)}\{2({}^3\mathcal D_r\hat K_s{}^{(c)} - {}^3\mathcal D_s\hat K_r{}^{(c)}) \\
&\quad + \hat\Gamma_r{}^{(0)}{}_{(d)}{}^3\mathcal D_s\hat g^{(c)(d)} - \hat\Gamma_s{}^{(0)}{}_{(d)} \\
&\quad \times {}^3\mathcal D_r\hat g^{(c)(d)} + \hat\Gamma_s{}^{(0)}{}_{(0)}(\hat\Gamma_r{}^{(c)}{}_{(0)} - \hat\Gamma_r{}^{(0)(c)}) \\
&\quad - \hat\Gamma_r{}^{(0)}{}_{(0)}(\hat\Gamma_s{}^{(c)}{}_{(0)} - \hat\Gamma_s{}^{(0)(c)})\} + {}_{cm}\hat{\mathcal T}^0{}_{(a)} = 0.
\end{aligned} \tag{5.16b'}
$$

Remarks: (i) The formula for $\hat\Gamma_k{}^{(a)}{}_{(b)}$ (5.12c) contains undetermined quantities $\hat\Gamma_k{}^{(0)}{}_{(0)}$; in the expression for ${}^3R$, however, these quantities do not appear. Taking into account formula (5.14) we see that the lhs of (5.16a′) is a function of symplectic variables, their spatial derivatives, and the components of hypermomentum and the canonical energy-momentum tensor.

(ii) If we unfold the ${}^3\mathcal D_s$ operators in (5.16b′) then, by virtue of (5.12c) all terms containing $\hat\Gamma_s{}^{(0)}{}_{(0)}$ cancel. Hence, the lhs of (5.16b′) depends only on the symplectic variables, their spatial derivatives, and the components of hypermomentum and the canonical energy-momentum tensor.

(iii) By virtue of (5.6) and (5.14), $\hat\Gamma_s{}^{(a)}{}_{(0)}$ and $\hat\Gamma_s{}^{(0)}{}_{(b)}$ are functions of the symplectic variables and the coefficients of hypermomentum. The quantities ${}^3R^k{}_{(a)} + {}^3R^k{}_{(c)(a)}{}^{(c)}$, ${}^3\mathcal D_s\hat\Gamma_0{}^{(a)}{}_{(0)} - \hat\Gamma_s{}^{(0)}{}_{(0)}\hat\Gamma_0{}^{(a)}{}_{(0)}$, and ${}^3\mathcal D_s\hat\Gamma_0{}^{(0)}{}_{(b)} + \hat\Gamma_s{}^{(0)}{}_{(0)}\hat\Gamma_0{}^{(0)}{}_{(b)}$ are independent of $\hat\Gamma_s{}^{(0)}{}_{(0)}$, cf. (5.12c).

If we use (5.12) and (5.13b) then the field equations (5.16c) read

$$
\begin{aligned}
&2\sqrt{\bar g}\,{}^3\mathcal D_0\hat K_s{}^{(a)} - 2\sqrt{\bar g}\hat K_s{}^{(a)}\hat\Gamma_0{}^{(0)}{}_{(0)} \\
&\quad + \sqrt{\bar g}({}^3R^{(a)}{}_s + {}^3R^{(a)}{}_{(c)s}{}^{(c)}) - {}_{cm}\hat{\mathcal T}^{(a)}{}_s + \tfrac12\hat e_s^{(a)}\,\mathrm{tr}\,{}_{cm}\hat{\mathcal T} \\
&\quad + (\text{terms depending on the gravitational symplectic} \\
&\qquad \text{variables, the coefficients of hypermomentum and} \\
&\qquad \partial_s \ln N) = 0.
\end{aligned} \tag{5.16c'}
$$

By virtue of the general theory (Sec. IV), Eqs. (5.16d) follow from the time-maintenance of equations (5.11b), (5.11c). We assume that Eqs. (5.11b) and (5.11c) are satisfied on the whole M. Hence, Eqs. (5.16d) follow from other equations and do not give us new information.

Equations:

$$
(\mathscr E 3)^{(\alpha)(\beta)} = {}_{sg}\mathcal T^{(\alpha)(\beta)} + {}_{sm}\mathcal T^{(\alpha)(\beta)} = 0. \tag{5.17}
$$

Making use of (5.12), (5.13), and (5.8) we get from $(\hat{\mathscr E}3)^{(a)(b)} = 0$,

$$
\begin{aligned}
&{}^3\mathcal D_0\hat{\mathcal H}_{(a)(b)} + 2\hat{\mathcal H}_{(a)(b)}\hat\Gamma_0{}^{(0)}{}_{(0)} \\
&\quad + \sqrt{\bar g}({}^3R_{(a)(b)} + {}^3R_{(b)(a)}) - {}_{sm}\hat{\mathcal T}_{(a)(b)} \\
&\quad + (\text{terms depending on the gravitational symplectic} \\
&\qquad \text{variables, coefficients of hypermomentum and} \\
&\qquad \partial_s \ln N) = 0.
\end{aligned} \tag{5.18}
$$

The general theory tells us that the equations $(\hat{\mathscr E}3)^{(0)(\alpha)} = 0$ are consequences of the time-maintenance conditions for the equations $(\hat{\mathscr E}2)^0{}_{(\alpha)}{}^{(0)} = 0$. Therefore these equations do not give us new information.

The $(3+1)$ analysis of the metric-affine theory of gravity with the Einstein gravitational Lagrangian and with the phenomenological description of matter can be summarized as follows.

The gravitational dynamical variables (5.7), the zero components of hypermomentum $\hat h^0{}_{(\alpha)}{}^{(\beta)}$, and the canonical energy-momentum tensor ${}_{cm}\hat{\mathcal T}^0{}_{(\alpha)}$ are subject to $9 + 4 = 13$ constraints (5.11d), (5.16a), and (5.16b).

The time evolution of the gravitational dynamical variables is determined by Eqs. (5.13a), (5.13b), (5.16c′), and (5.18).

Equations (5.14) enable us to express $\hat\Gamma_k{}^{(a)}{}_{(0)}$ and $\hat\Gamma_k{}^{(0)}{}_{(b)}$ by means of symplectic variables.

The variables $\hat\Gamma_k{}^{(a)}{}_{(b)}$, $\hat\Gamma_0{}^{(0)}{}_{(b)}$, $\hat\Gamma_0{}^{(a)}{}_{(0)}$ are to be computed by means of (5.12).

The variables $\hat\Gamma_k{}^{(0)}{}_{(0)}$ do not appear at all in the field equations.

The variable $\hat\Gamma_0{}^{(0)}{}_{(0)}$ appears always together with the ${}^3\mathcal D_0$ terms.

The quantities $\hat\Gamma_0{}^{(a)}{}_{(b)}$ are arbitrary GL(3,R)-gauge variables.

In the phenomenological picture the hypermomentum current as well as the canonical and symmetric energy-momentum tensors have to satisfy the conservation laws (3.8a) and (3.9). Moreover, in the Einstein case

$$
h^\lambda{}_{(\alpha)}{}^{(\alpha)} = 0 \quad \text{and} \quad {}_{cm}\mathcal T^{(\alpha)}{}_{(\alpha)} - {}_{sm}\mathcal T^{(\alpha)}{}_{(\alpha)} = 0. \tag{5.19}
$$

We assume that the components of the symmetric energy-momentum tensor ${}_{sm}\hat{\mathcal T}^{(\alpha)(b)}$, as well as the spatial components of hypermomentum $\hat h^k{}_{(\alpha)}{}^{(\beta)}$, and the spatial components of the canonical energy-momentum tensor ${}_{cm}\hat{\mathcal T}^k{}_{(\alpha)}$ are known functions on space-time. If we make use of Eqs. (5.12) and (5.14) then the conservation laws (3.8a) and (3.9) determine the dynamics of $\hat h^0{}_{(\alpha)}{}^{(\beta)}$ and ${}_{cm}\hat{\mathcal T}^0{}_{(\alpha)}$. We write

$$
\left.\begin{array}{l} {}^3\mathcal D_0\hat h^0{}_{(\alpha)}{}^{(\beta)} \\ \\ {}^3\mathcal D_0{}_{cm}\hat{\mathcal T}^0{}_{(\alpha)} \end{array}\right\}
= \left\{\begin{array}{l} \text{functions of the gravitational symplectic variables,} \\ \hat h^0{}_{(\alpha)}{}^{(\beta)},\ {}_{cm}\hat{\mathcal T}^0{}_{(\alpha)},\ \text{the gauge variables } N,\ N^k,\ \hat\Gamma_0{}^{(a)}{}_{(b)}, \\ \text{their } x^k \text{ derivatives as well as of the variable } \hat\Gamma_0{}^{(0)}{}_{(0)} \end{array}\right\}. \tag{5.20}
$$

It is very important that the quantities $\hat\Gamma_k{}^{(0)}{}_{(0)}$ do not appear in Eqs. (5.20). This fact can be checked directly. We only have to make use of relations (5.19). The manifest dependence of the right-hand sides of (5.20) on the quantity $\hat\Gamma_0{}^{(0)}{}_{(0)}$ can be supressed if we pass from the GL(3,R)-gauge variables $\hat\Gamma_0{}^{(a)}{}_{(b)}$ to the new ones ${}'\hat\Gamma_0{}^{(a)}{}_{(b)} = \hat\Gamma_0{}^{(a)}{}_{(b)} - \delta^{(a)}{}_{(b)}\hat\Gamma_0{}^{(0)}{}_{(0)}$. If such a transformation is performed then the quantity $\hat\Gamma_0{}^{(0)}{}_{(0)}$ disappears also from the dynamical field equations (5.13), (5.16c), and (5.18).

Let us sum up. In the phenomenological picture the Cauchy–Kowalewska initial value problem may be formulated as follows.

(i) We assign 30 gravitational symplectic variables and 20 components $\hat{h}^{0}{}_{(\alpha)}{}^{(\beta)}$, ${}_{cm}\hat{\mathscr{T}}^{0}{}_{(\alpha)}$ on the initial surface σ such that 13 constraints (5.11d), (5.16a), (5.16b), and Eqs. (5.19) are satisfied.

(ii) Values of 20 gauge variables N, N^{k}, $\hat{\Gamma}_{0}{}^{(a)}{}_{(b)}$, $n^{(a)}$, $n_{(b)}$, and values of the quantities $\hat{h}^{k}{}_{(\alpha)}{}^{(\beta)}$, ${}_{cm}\hat{\mathscr{T}}^{k}{}_{(\alpha)}$, ${}_{sm}\hat{\mathscr{T}}^{(a)(b)}$ are fixed on space-time.

Then, the dynamical gravitational equations and the conservation laws (5.20) give us the evolution of the system. Four quantities $\hat{\Gamma}_{\lambda}{}^{(0)}{}_{(0)}$ are not determined by this procedure. The quantity ${}_{sm}\hat{\mathscr{T}}^{(0)(0)}$ is to be determined from (5.19).

The appearance of four further arbitrary quantities in the dynamical formulation of the theory becomes clear if we recall that in the metric affine theory the Einstein gravitational Lagrangian is invariant with respect to the following transformations (cf. Refs. 11 and 43):

$$e^{(\alpha)}_{\lambda}\to{}'e^{(\alpha)}_{\lambda}=e^{(\alpha)}_{\lambda},\quad g_{(\alpha)(\beta)}\to{}'g_{(\alpha)(\beta)}=g_{(\alpha)(\beta)},$$

$$\Gamma_{\lambda}{}^{(\alpha)}{}_{(\beta)}\to{}'\Gamma_{\lambda}{}^{(\alpha)}{}_{(\beta)}=\Gamma_{\lambda}{}^{(\alpha)}{}_{(\beta)}+\chi_{\lambda}\delta^{(\alpha)}{}_{(\beta)}. \tag{5.21}$$

We have

$$'R^{(\alpha)}{}_{(\beta)\mu\nu}=R^{(\alpha)}{}_{(\beta)\mu\nu}+(\partial_{\mu}\chi_{\nu}-\partial_{\nu}\chi_{\mu})\delta^{(\alpha)}{}_{(\beta)},$$

$$'Q^{(\alpha)}{}_{\mu\nu}=Q^{(\alpha)}{}_{\mu\nu}+\chi_{\mu}e^{(\alpha)}_{\nu}-\chi_{\nu}e^{(\alpha)}_{\mu}, \tag{5.22}$$

$$'M_{\lambda(\alpha)(\beta)}=M_{\lambda(\alpha)(\beta)}-\chi_{\lambda}g_{(\alpha)(\beta)},$$

and the tensors $R^{\lambda}{}_{(\alpha)}+R^{\lambda}{}_{(\epsilon)(\alpha)}{}^{(\epsilon)}$, $R^{(\alpha)(\beta)}+R^{(\beta)(\alpha)}$ are invariant with respect to the transformations (5.21).

If

$$'h^{\lambda}{}_{(\alpha)}{}^{(\beta)}=h^{\lambda}{}_{(\alpha)}{}^{(\beta)},\quad {}'_{cm}\mathscr{T}^{\lambda}{}_{(\alpha)}={}'_{cm}\mathscr{T}^{\lambda}{}_{(\alpha)},$$

$${}'_{sm}\mathscr{T}^{(\alpha)(\beta)}={}'_{sm}\mathscr{T}^{(\alpha)(\beta)}, \tag{5.23}$$

then the field equations (5.11), (5.16), (5.17) are invariant with respect to (5.21)–(5.23). Moreover, in virtue of relations (5.19) the conservation laws (3.8a) and (3.9) are also invariant with respect to the transformations (5.21)–(5.23).

In the caret variables we have

$$'\hat{\Gamma}_{\lambda}{}^{(0)}{}_{(0)}=\hat{\Gamma}_{\lambda}{}^{(0)}{}_{(0)}+\bar{\chi}_{\lambda}. \tag{5.21$'$}$$

Therefore, gauge transformations (5.21) can change the variables $\hat{\Gamma}_{\lambda}{}^{(0)}{}_{(0)}$ arbitrarily. Now it is clear why the dynamics of these quantities cannot be obtained from the field equations. It is interesting to note that gauge transformations (5.21) do not affect the dynamical quantities of the theory.

Remark: In the phenomenological picture the quantities $\hat{\Gamma}_{\lambda}{}^{(0)}{}_{(0)}$ are arbitrary and they cannot be determined from the field equations. If, however, matter is described by a Lagrangian, then the theory is not invariant with respect to transformations (5.21), and the $\hat{\Gamma}_{\lambda}{}^{(0)}{}_{(0)}$ are not arbitrary quantities. In the next section we show how to determine them by means of relations (5.19).

The above-presented dynamical analysis of the metric affine Einstein theory of gravity shows that the results published in Ref. 5 were incomplete. In particular, those papers do not contain dynamical equations for the symplectic variables $\hat{K}_{s}^{(a)}$. Moreover, as we show in the next section, for Lagrangian matter field relations (5.19) give rise to additional symplectic constraints and the analysis of these constraints is essential for the dynamics. Also, this problem was not fully clarified in the previous papers.

VI. THE METRIC-AFFINE EINSTEIN GRAVITY COUPLED TO A VECTOR FIELD

In the previous section we discussed the Einstein gravitational equations in the framework of metric-affine geometry with matter characterized by its hypermomentum and two energy-momentum tensors. In the present section we consider a Lagrangian matter field described by a vector potential $f^{(\alpha)}$ and the Lagrangian

$$\mathscr{L}_{m}=(A/2)\sqrt{-g}\,\mathscr{D}_{\mu}f^{(\alpha)}\mathscr{D}_{\nu}f^{(\beta)}g^{\mu\nu}g_{(\alpha)(\beta)}$$
$$+\sqrt{-g}\,V(f^{(\alpha)},g_{(\alpha)(\beta)}), \tag{6.1}$$

where A is an arbitrary constant and $-V$ is a potential function. We have

$$\mathscr{P}^{\lambda}{}_{(\alpha)}=\sqrt{-g}\,\mathscr{D}_{\tau}f^{(\beta)}g^{\lambda\tau}g_{(\alpha)(\beta)},$$
$$h^{\lambda}{}_{(\alpha)}{}^{(\beta)}=\mathscr{P}^{\lambda}{}_{(\alpha)}f^{(\beta)}, \tag{6.2}$$
$$j_{(\alpha)}=\sqrt{-g}\frac{\partial V}{\partial f^{(\alpha)}}.$$

The matter field equation and the energy-momentum tensors read

$$(\mathscr{E}\mathscr{M})_{(\alpha)}=j_{(\alpha)}-\mathscr{D}_{\lambda}\mathscr{P}^{\lambda}{}_{(\alpha)}=0, \tag{6.3}$$

$${}_{cm}\mathscr{T}^{\lambda}{}_{(\alpha)}=e^{\lambda}_{(\alpha)}\mathscr{L}_{m}$$
$$-\sqrt{-g}\,\mathscr{D}_{\nu}f^{(\epsilon)}\mathscr{D}_{\tau}f^{(\beta)}g^{\lambda\nu}g_{(\epsilon)(\beta)}e^{\tau}_{(\alpha)}, \tag{6.4a}$$

$${}_{sm}\mathscr{T}^{(\alpha)(\beta)}=g^{(\alpha)(\beta)}\mathscr{L}_{m}+\sqrt{-g}\,\mathscr{D}_{\mu}f^{(\alpha)}\mathscr{D}_{\nu}f^{(\beta)}g^{\mu\nu}$$
$$-\sqrt{-g}\,\mathscr{D}_{\mu}f^{(\epsilon)}\mathscr{D}_{\nu}f^{(\tau)}g_{(\epsilon)(\tau)}e^{(\alpha)\mu}e^{(\beta)\nu}$$
$$+2\sqrt{-g}\frac{\partial V}{\partial g_{(\alpha)(\beta)}}. \tag{6.4b}$$

In the $(3+1)$ variables we obtain

$$\hat{\mathscr{P}}^{0}{}_{(0)}=A\sqrt{\bar{g}}(\bar{\partial}_{0}\hat{f}^{(0)}+\hat{\Gamma}_{0}{}^{(0)}{}_{(0)}\hat{f}^{(0)}+\hat{\Gamma}_{0}{}^{(0)}{}_{(c)}\hat{f}^{(c)}),$$
$$\hat{\mathscr{P}}^{0}{}_{(a)}=-A\sqrt{\bar{g}}({}^{3}\mathscr{D}_{0}\hat{f}^{(c)}+\hat{\Gamma}_{0}{}^{(c)}{}_{(0)}\hat{f}^{(0)})\hat{g}_{(c)(a)}, \tag{6.5a}$$
$$\hat{\mathscr{P}}^{k}{}_{(0)}=-A\sqrt{\bar{g}}(\partial_{s}\hat{f}^{(0)}+\hat{\Gamma}_{s}{}^{(0)}{}_{(0)}\hat{f}^{(0)}+\hat{\Gamma}_{s}{}^{(0)}{}_{(c)}\hat{f}^{(c)})\bar{g}^{sk},$$
$$\hat{\mathscr{P}}^{k}{}_{(a)}=A\sqrt{\bar{g}}({}^{3}\nabla_{s}\hat{f}^{(c)}+\hat{r}_{s}{}^{(c)}{}_{(d)}\hat{f}^{(d)}$$
$$+\hat{\Gamma}_{s}{}^{(c)}{}_{(0)}\hat{f}^{(0)})\bar{g}^{sk}\hat{g}_{(c)(a)}. \tag{6.5b}$$

Here, ${}^{3}\nabla_{s}$ is the Riemannian covariant derivative on slices defined by means of (B1a), where the $\hat{\Gamma}_{k}{}^{(a)}{}_{(b)}$ are replaced with the anholonomic Riemannian coefficients ${}^{3}\gamma_{k}{}^{(a)}{}_{(b)}$.

Relations (6.5a) determine the dynamics of the variables $\hat{f}^{(0)}$, $\hat{f}^{(a)}$; the dynamics of the conjugate momenta is given by matter field equations

$$(\hat{\mathscr{E}}\mathscr{M})_{(0)}=-{}^{3}\mathscr{D}_{0}\hat{\mathscr{P}}^{0}{}_{(0)}-({}^{3}\mathscr{D}_{s}+\partial_{s}\ln N)\hat{\mathscr{P}}^{s}{}_{(0)}$$
$$+\hat{\Gamma}_{0}{}^{(0)}{}_{(0)}\hat{\mathscr{P}}^{0}{}_{(0)}+\hat{\Gamma}_{s}{}^{(0)}{}_{(0)}\hat{\mathscr{P}}^{s}{}_{(0)}$$
$$+\hat{\Gamma}_{0}{}^{(c)}{}_{(0)}\hat{\mathscr{P}}^{0}{}_{(c)}+\hat{\Gamma}_{s}{}^{(c)}{}_{(0)}\hat{\mathscr{P}}^{s}{}_{(c)}+\hat{j}_{(0)}=0, \tag{6.6a}$$

$$(\hat{\mathscr{E}}\mathscr{M})_{(a)}=-{}^{3}\mathscr{D}_{0}\hat{\mathscr{P}}^{0}{}_{(a)}-({}^{3}\mathscr{D}_{s}+\partial_{s}\ln N)\hat{\mathscr{P}}^{s}{}_{(a)}$$
$$+\hat{\Gamma}_{0}{}^{(0)}{}_{(a)}\hat{\mathscr{P}}^{0}{}_{(0)}+\hat{\Gamma}_{s}{}^{(0)}{}_{(a)}\hat{\mathscr{P}}^{s}{}_{(0)}+\hat{j}_{(a)}=0. \tag{6.6b}$$

We remember that the Einstein gravitational Lagrangian gives rise to the primary constraints $h^{\lambda}{}_{(\alpha)}{}^{(\alpha)} = 0$ (5.10). In the (3 + 1) picture we have

$$\hat{h}^{0}{}_{(\alpha)}{}^{(\alpha)} = \hat{\mathscr{P}}^{0}{}_{(0)}\hat{f}^{(0)} + \hat{\mathscr{P}}^{0}{}_{(a)}\hat{f}^{(a)} = 0, \tag{6.7a}$$

$$\hat{h}^{k}{}_{(\alpha)}{}^{(\alpha)} = \hat{\mathscr{P}}^{k}{}_{(0)}\hat{f}^{(0)} + \hat{\mathscr{P}}^{k}{}_{(a)}\hat{f}^{(a)} = 0. \tag{6.7b}$$

Equation (6.7a) is a symplectic constraint, whereas Eq. (6.7b) together with (6.5b) lead to relations for the unknown quantities $\hat{\Gamma}_{k}{}^{(0)}{}_{(0)}$. Making use of (6.5b) we observe that relations (5.12c), (5.14), and (6.7b) give rise to a linear system of equations for $\hat{\Gamma}_{k}{}^{(a)}{}_{(b)}$, $\hat{\Gamma}_{[(a)}{}^{(0)}{}_{(b)]}$, and $\hat{\Gamma}_{k}{}^{(0)}{}_{(0)}$. This system can be solved by means of simple algebraic methods presented in Ref. 31. If the conditions (here $\hat{f}^2 = \hat{f}_{(a)}\hat{f}^{(a)}$)

$$\begin{aligned}
&(\hat{f}^{(0)})^2 - \hat{f}^2 \neq 0, \quad (1 + \tfrac{1}{4}A\hat{f}^2) \neq 0,\\
&(1 + \tfrac{1}{2}A\hat{f}^2) \neq 0, \quad (1 - \tfrac{1}{4}A(\hat{f}^{(0)})^2) \neq 0,\\
&(1 + (\tfrac{1}{4}A + \tfrac{1}{16}A^2\hat{f}^2)((\hat{f}^{(0)})^2 - \hat{f}^2)) \neq 0,\\
&(1 + \tfrac{1}{2}A(\hat{f}^2 - (\hat{f}^{(0)})^2)) \neq 0
\end{aligned} \tag{6.8}$$

hold, then $\hat{\Gamma}_{k}{}^{(a)}{}_{(b)}$, $\hat{\Gamma}_{[(a)}{}^{(0)}{}_{(b)]}$, and $\hat{\Gamma}_{k}{}^{(0)}{}_{(0)}$ are rational functions of the gravitational symplectic variables, matter symplectic variables, as well as of x^k derivatives of $\hat{f}^{(0)}, \hat{f}^{(a)}$, $\hat{e}^{k}_{(a)}$, and $\hat{g}^{(a)(b)}$. If we apply relations (5.12a) and (5.12b) for the quantities $\hat{\Gamma}_{0}{}^{(a)}{}_{(0)}, \hat{\Gamma}_{0}{}^{(0)}{}_{(b)}$, then the structure of the field equations is almost clear. The only undetermined quantity is $\hat{\Gamma}_{0}{}^{(0)}{}_{(0)}$. In order to find an equation for $\hat{\Gamma}_{0}{}^{(0)}{}_{(0)}$ we investigate the time-maintenance property of the symplectic constraint (6.7a).

This constraint is preserved in time if

$$\mathscr{D}_{\lambda}h^{\lambda}{}_{(\alpha)}{}^{(\alpha)} = 0, \tag{6.9}$$

or, in virtue of the field equations,

$$\begin{aligned}
A\sqrt{\bar{g}}\,\hat{j}_{(a)}\hat{f}^{(a)} + (\hat{\mathscr{P}}^{0}{}_{(0)})^2 - \hat{\mathscr{P}}^{0}{}_{(a)}\hat{\mathscr{P}}^{0}{}_{(b)}\hat{g}^{(a)(b)}&\\
- \hat{\mathscr{P}}^{k}{}_{(0)}\hat{\mathscr{P}}^{s}{}_{(0)}\bar{g}_{ks} + \hat{\mathscr{P}}^{k}{}_{(a)}\hat{\mathscr{P}}^{s}{}_{(b)}\bar{g}_{ks}\hat{g}^{(a)(b)} &= 0.
\end{aligned} \tag{6.10}$$

If we substitute relations for $\hat{\Gamma}_{k}{}^{(a)}{}_{(b)}$, $\hat{\Gamma}_{[(a)}{}^{(0)}{}_{(b)]}$, $\hat{\Gamma}_{k}{}^{(0)}{}_{(0)}$ and (6.5b) into (6.10), then we obtain a constraint for the gravitational and matter symplectic variables. The maintenance of the secondary constraint (6.10) in time leads to a new relation. The analytic form of this condition can be obtained either by means of the direct ${}^3\mathscr{D}_0$ differentiation of the relation (6.10) or, more elegantly, from the relation

$$\mathscr{D}_{\epsilon}(\mathscr{D}_{\lambda}h^{\lambda}{}_{(\alpha)}{}^{(\alpha)}) = 0. \tag{6.11}$$

Taking into account the matter field equations and commuting the second-order derivatives, we obtain

$$\begin{aligned}
A(-g)\mathscr{D}_{\epsilon}((-g)^{-1/2}j_{(\alpha)}f^{(\alpha)})&\\
+ 2\mathscr{P}^{\nu}{}_{(\alpha)}g_{\epsilon\mu}\mathscr{D}_{\nu}(\mathscr{P}^{\mu}{}_{(\beta)}g^{(\alpha)(\beta)})&\\
+ 2A\sqrt{-g}R^{(\alpha)}{}_{(\beta)\epsilon\nu}f^{(\beta)}\mathscr{P}^{\nu}{}_{(\alpha)}&\\
- \mathscr{P}^{\mu}{}_{(\alpha)}\mathscr{P}^{\nu}{}_{(\beta)}g_{\mu\nu}\mathscr{D}_{\epsilon}g^{(\alpha)(\beta)} &= 0.
\end{aligned} \tag{6.12}$$

Taking $\epsilon = 0$ in the "$\hat{\ }$" form of (6.12) we get an equation for $\hat{\Gamma}_{0}{}^{(0)}{}_{(0)}$. In Appendix G it is proved that this is a linear algebraic equation for $\hat{\Gamma}_{0}{}^{(0)}{}_{(0)}$ (it does not depend on $\partial_k\hat{\Gamma}_{0}{}^{(0)}{}_{(0)}$!) and, eventually, $\hat{\Gamma}_{0}{}^{(0)}{}_{(0)}$ is a rational function of symplectic variables, their first and second spatial derivatives as well as of $\partial_k \ln N$. Of course, the solvability of (6.12) requires some additional regularity conditions for initial values of symplectic variables [cf. (6.8)]. In the weak field approximation, that is, if we neglect terms of higher than quadratic order in $\hat{f}^{(0)}$, $\hat{f}^{(a)}$, $\hat{\mathscr{P}}^{0}{}_{(0)}$, $\hat{\mathscr{P}}^{0}{}_{(a)}$, the regularity condition has the following simple form:

$$(\hat{\mathscr{P}}^{0}{}_{(0)})^2 - \hat{\mathscr{P}}^{0}{}_{(a)}\hat{\mathscr{P}}^{0}{}_{(b)}\hat{g}^{(a)(b)} \neq 0.$$

Summing up: In the metric-affine theory of gravity with the Einstein gravitational Lagrangian and with the matter Lagrangian (6.1), we have 30 gravitational symplectic variables $\hat{e}^{k}_{(a)}$, $\hat{K}^{(a)}_{k}$, $\hat{\mathscr{H}}_{(a)(b)}$, $\hat{g}^{(a)(b)}$ and eight matter symplectic variables $\hat{\mathscr{P}}^{0}{}_{(0)}, \hat{f}^{(0)}, \hat{\mathscr{P}}^{0}{}_{(a)}, \hat{f}^{(a)}$. Thirteen gravitational symplectic constraints (5.11d), (5.16a), and (5.16b) are related to the actions of the diffeomorphism group of space-time and the local GL(3,R) group in the space of symplectic variables. Therefore they reduce $2\cdot 13 = 26$ degrees of freedom. The matter constraints (6.7a) and (6.10) eliminate two degrees of freedom. Therefore, we have four gravitational and six matter degrees of freedom (in the phase space).

The quantities $\hat{\Gamma}_{k}{}^{(a)}{}_{(b)}, \hat{\Gamma}_{[(a)}{}^{(0)}{}_{(b)]}, \hat{\Gamma}_{0}{}^{(a)}{}_{(0)}, \hat{\Gamma}_{0}{}^{(0)}{}_{(b)}, \hat{\Gamma}_{k}{}^{(0)}{}_{(0)}$, $\hat{\Gamma}_{0}{}^{(0)}{}_{(0)}$ are uniquely determined by means of algebraic formulas. The $\hat{\Gamma}_{0}{}^{(a)}{}_{(b)}$ are arbitrary GL(3,R)-gauge variables and N, N^k are arbitrary Diff M-gauge variables.

In our example, it is very important that the time-maintenance condition for the secondary constraint (6.12) leads to a linear, algebraic equation for $\hat{\Gamma}_{0}{}^{(0)}{}_{(0)}$. This feature of the theory does not depend on the particular choice of a vector matter Lagrangian and it is related to the structure of the gravitational Einstein Lagrangian.

The matter Lagrangian (6.1) is canonically regular, that is, the relations (6.5a) give us a one-to-one correspondence between the time derivatives of matter potentials ${}^3\mathscr{D}_0\hat{f}^{(\alpha)}$ and the matter momenta $\hat{\mathscr{P}}^{0}{}_{(\alpha)}$. For canonically nonregular matter Lagrangians some time derivatives of matter potentials cannot be determined by the momenta. In such cases we have additional primary matter constraints and the corresponding secondary constraints. The time-maintenance conditions for these secondary constraints give rise to linear differential equations for the values of undetermined time derivatives of matter potentials on the initial surface σ. Now, the classical Cauchy–Kowalewska procedure is based not only on differential and algebraic operations but also requires finding solutions of those linear differential equations. Several examples of such a situation were considered in Ref. 31(b) in the framework of the ECSK theory.

Let us briefly describe the main differences between the gravitational theories with the Einstein gravitational Lagrangian in metric-affine and in Cartan space-time (metic compatible connections). In metric-affine space-time the Einstein gravitational Lagrangian is invariant with respect to projective transformations[11,43] (5.21) and this fact leads to additional matter constraints and, thus, it reduces some matter degrees of freedom. In the ECSK theory, the gravitational Lagrangian loses its projective invariance, we have less matter symplectic constraints and more matter degrees of freedom. For instance, for the regular vector Lagrangian (6.1) we have six independent matter degrees of freedom in the metric-affine Einstein theory but eight of them in the ECSK theory.[31] However, the number of independent gravitational degrees of freedom is equal to 4 in both theories.

ACKNOWLEDGMENTS

One of us (WS) would like to thank Professor F. W. Hehl (University of Cologne) for fruitful discussions concerning gauge theories of gravity, his valuable remarks and comments.

Special thanks are due to Mrs. J. Perzanowska for her help in preparation of the manuscript.

The work of one of us (MA) was supported in part by the Polish Ministry of Science, Higher Education, and Technology under Grant No. MR.1.7. to Warsaw University.

APPENDIX A: THE "CARET" OPERATION FOR GL(4,R)-COVARIANT DERIVATIVES

The bar and tilde operations applied to a tensor density $\mathscr{D}_\lambda \mathscr{F}^{\mu(\alpha)}_{\nu(\beta)}$ give a family of tensorial objects $(\mathscr{D}_\lambda \mathscr{F}^{\mu(\alpha)}_{\nu(\beta)})\hat{}$ on particular slices. In Refs. 31 and 35 it was proved that for a tensor density of weight r,

$$
\begin{aligned}
(\mathscr{D}_\lambda \mathscr{F}^{\mu(\alpha)}_{\nu(\beta)})\hat{} = {} & \bar\partial_\lambda \hat{\mathscr{F}}^{\mu(\alpha)}_{\nu(\beta)} - r\hat\gamma_\lambda{}^\tau{}_\tau \hat{\mathscr{F}}^{\mu(\alpha)}_{\nu(\beta)} \\
& + \hat\gamma_\lambda{}^\mu{}_\tau \hat{\mathscr{F}}^{\tau(\alpha)}_{\nu(\beta)} - \hat\gamma_\lambda{}^\tau{}_\nu \hat{\mathscr{F}}^{\mu(\alpha)}_{\tau(\beta)} \\
& + \hat\Gamma_\lambda{}^{(\alpha)}{}_{(\tau)} \hat{\mathscr{F}}^{\mu(\tau)}_{\nu(\beta)} - \hat\Gamma_\lambda{}^{(\tau)}{}_{(\beta)} \hat{\mathscr{F}}^{\mu(\alpha)}_{\nu(\tau)}, \qquad \text{(A1)}
\end{aligned}
$$

where

$$
\begin{aligned}
\bar\partial_\lambda &= (A^{-1})^\tau{}_{\bar\lambda}\, \partial_\tau, \\
\hat\gamma_\lambda{}^\mu{}_\nu &= \bar\gamma_\lambda{}^\mu{}_\nu = A^{\bar\mu}{}_\tau (A^{-1})^\epsilon{}_{\bar\lambda} (A^{-1})^\delta{}_{\bar\nu} \gamma_\epsilon{}^\tau{}_\delta \\
&\quad + A^{\bar\mu}{}_\tau \bar\partial_\lambda (A^{-1})^\tau{}_{\bar\nu}. \qquad \text{(A2)}
\end{aligned}
$$

The symplectic momenta $m_{(\alpha)}$, $m^{(a)}$ are

$$
\begin{aligned}
m^{(a)} = {} & \frac{1-n_{(0)}}{1+n^{(0)}} (\hat{\mathscr{E}}2)^0{}_{(0)}{}^{(a)} - \frac{n^{(a)} n_{(c)}}{n^{(0)}(1+n^{(0)})} (\hat{\mathscr{E}}2)^0{}_{(0)}{}^{(c)} \\
& + \frac{n^{(a)}}{n^{(0)}(1-n_{(0)})} (\hat{\mathscr{E}}2)^0{}_{(c)}{}^{(c)} - \frac{n^{(c)}}{1+n^{(0)}} (\hat{\mathscr{E}}2)^0{}_{(c)}{}^{(a)} \\
& + \frac{n^{(a)} n^{(b)} n_{(c)}}{n^{(0)}(1+n^{(0)})(1-n_{(0)})} (\hat{\mathscr{E}}2)^0{}_{(b)}{}^{(c)}, \qquad \text{(A3a)}
\end{aligned}
$$

$$
\begin{aligned}
m_{(a)} = {} & -n_{(a)} (\hat{\mathscr{E}}2)^0{}_{(0)}{}^{(0)} + \frac{1+n^{(0)}}{1-n_{(0)}} (\hat{\mathscr{E}}2)^0{}_{(a)}{}^{(0)} \\
& + \frac{n_{(a)} n^{(c)}}{1-n_{(0)}} (\hat{\mathscr{E}}2)^0{}_{(c)}{}^{(0)} - \frac{n_{(a)} n_{(c)}}{n^{(0)}} (\hat{\mathscr{E}}2)^0{}_{(0)}{}^{(c)} \\
& + \frac{n_{(c)}}{1-n_{(0)}} (\hat{\mathscr{E}}2)^0{}_{(a)}{}^{(c)} + \frac{n_{(a)} n_{(c)} n^{(b)}}{n^{(0)}(1-n_{(0)})} (\hat{\mathscr{E}}2)^0{}_{(b)}{}^{(c)} \\
& + \frac{n_{(a)}}{n^{(0)}(1-n_{(0)})} (\hat{\mathscr{E}}2)^0{}_{(c)}{}^{(c)}, \qquad \text{(A3b)}
\end{aligned}
$$

$$
\begin{aligned}
m_{(0)} = {} & -n_{(0)} (\hat{\mathscr{E}}2)^0{}_{(0)}{}^{(0)} - n^{(c)} (\hat{\mathscr{E}}2)^0{}_{(c)}{}^{(0)} \\
& - \frac{n_{(0)} n_{(c)}}{n^{(0)}} (\hat{\mathscr{E}}2)^0{}_{(c)}{}^{(0)} \\
& + \frac{n_{(0)} - n^{(0)} + 2n^{(0)} n_{(0)}}{n^{(0)}(1+n^{(0)})(1-n_{(0)})} (\hat{\mathscr{E}}2)^0{}_{(c)}{}^{(c)} \\
& - \frac{n^{(b)} n_{(c)} (n^{(0)} - n^{(0)} n_{(0)} - n_{(0)})}{n^{(0)}(1+n^{(0)})(1-n_{(0)})} (\hat{\mathscr{E}}2)^0{}_{(b)}{}^{(c)}. \qquad \text{(A3c)}
\end{aligned}
$$

These formulas enable us to determine $(\hat{\mathscr{E}}2)^0{}_{(a)}{}^{(0)}$ and $(\hat{\mathscr{E}}2)^0{}_{(0)}{}^{(b)}$ as functions of $m_{(a)}$, $m^{(b)}$, $n^{(\alpha)}$, $n_{(b)}$ as well as of other symplectic variables and their spatial derivatives [cf. (B4)].

APPENDIX B: GL(4,R)-COVARIANT DIFFERENTIAL OPERATORS

In the (3 + 1)-picture we use caret quantities, which are GL(3,R)-covariant tensorial objects on particular slices. We define the following GL(3,R)-covariant differential operators for a tensor density (of weight r) $\hat{\mathscr{F}}^{\mu(\alpha)}_{\nu(\beta)}$ on σ:

$$
\begin{aligned}
{}^3\mathscr{D}_k \hat{\mathscr{F}}^{\mu(\alpha)}_{\nu(\beta)} = {} & \bar\partial_k \hat{\mathscr{F}}^{\mu(\alpha)}_{\nu(\beta)} - r\,{}^3\gamma_k{}^\tau{}_\tau \hat{\mathscr{F}}^{\mu(\alpha)}_{\nu(\beta)} + {}^3\gamma_k{}^\mu{}_\tau \hat{\mathscr{F}}^{\tau(\alpha)}_{\nu(\beta)} \\
& - {}^3\gamma_k{}^\tau{}_\nu \hat{\mathscr{F}}^{\mu(\alpha)}_{\tau(\beta)} + {}^3\Gamma_k{}^{(\alpha)}{}_{(\tau)} \hat{\mathscr{F}}^{\mu(\tau)}_{\nu(\beta)} \\
& - {}^3\Gamma_k{}^{(\tau)}{}_{(\beta)} \hat{\mathscr{F}}^{\mu(\alpha)}_{\nu(\tau)}, \qquad \text{(B1a)}
\end{aligned}
$$

$$
\begin{aligned}
{}^3\mathscr{D}_0 \hat{\mathscr{F}}^{\mu(\alpha)}_{\nu(\beta)} = {} & \bar\partial_0 \hat{\mathscr{F}}^{\mu(\alpha)}_{\nu(\beta)} - r\,{}^3\sigma_0{}^\tau{}_\tau \hat{\mathscr{F}}^{\mu(\alpha)}_{\nu(\beta)} \\
& + {}^3\sigma_0{}^\mu{}_\tau \hat{\mathscr{F}}^{\tau(\alpha)}_{\mu(\beta)} - {}^3\sigma_0{}^\tau{}_\nu \hat{\mathscr{F}}^{\mu(\alpha)}_{\tau(\beta)} \\
& + {}^3\Gamma_0{}^{(\alpha)}{}_{(\tau)} \hat{\mathscr{F}}^{\mu(\tau)}_{\nu(\beta)} - {}^3\Gamma_0{}^{(\tau)}{}_{(\beta)} \hat{\mathscr{F}}^{\mu(\alpha)}_{\nu(\tau)}, \qquad \text{(B1b)}
\end{aligned}
$$

where

$$
\begin{aligned}
{}^3\Gamma_\mu{}^{(\alpha)}{}_{(\beta)} &= \delta^{(\alpha)}{}_{(a)} \delta^{(b)}{}_{(\beta)} \hat\Gamma_\mu{}^{(a)}{}_{(b)}, \\
{}^3\gamma_k{}^\mu{}_\nu &= \delta^\mu{}_r \delta^s{}_\nu \hat\gamma_k{}^r{}_s, \\
{}^3\sigma_0{}^\mu{}_\nu &= \delta^\mu{}_r \delta^s{}_\nu (1/N) \partial_s N^r. \qquad \text{(B2)}
\end{aligned}
$$

Remarks: Formulas (B1) and (B2) are covariant with respect to transformations (4.14) of local coordinates and with respect to local GL(3,R) rotations (even x^0 dependent).

For ${}^3\gamma_k{}^r{}_s = \hat\gamma_k{}^r{}_s$ are Christoffel symbols of Levi–Civita connection on σ and ${}^3\Gamma_k{}^{(\alpha)}{}_{(\beta)}$ determine a GL(3,R) connection on σ, the formula (B1a) corresponds directly to (2.8) (valency of a tensor on σ is determined by the number of nonzero indices).

The lhs of the Hamiltonian constraints (4.21a) written in terms of GL(3,R)-covariant objects read

$$
(\hat{\mathscr{E}}1)^0{}_{(0)} = {}_{cg}\hat{\mathscr{T}}^0{}_{(0)} + {}_{cm}\hat{\mathscr{T}}^0{}_{(0)} + {}^3\mathscr{D}_s \hat{\mathscr{U}}^{0s}{}_{(0)} - \hat\Gamma_s{}^{(\tau)}{}_{(0)} \hat{\mathscr{U}}^{0s}{}_{(\tau)}, \qquad \text{(B3a)}
$$

$$
(\hat{\mathscr{E}}1)^0{}_{(a)} = {}_{cg}\hat{\mathscr{T}}^0{}_{(a)} + {}_{cm}\hat{\mathscr{T}}^0{}_{(a)} + {}^3\mathscr{D}_s \hat{\mathscr{U}}^{0s}{}_{(a)} - \hat\Gamma_s{}^{(0)}{}_{(a)} \hat{\mathscr{U}}^{0s}{}_{(0)}, \qquad \text{(B3b)}
$$

$$
\begin{aligned}
(\hat{\mathscr{E}}2)^0{}_{(0)}{}^{(0)} = {} & \hat{h}^0{}_{(0)}{}^{(0)} - \hat{\mathscr{V}}^{0(0)}{}_{(0)} + {}^3\mathscr{D}_s \hat{\mathscr{P}}^{0s}{}_{(0)}{}^{(0)} \\
& - \hat\Gamma_s{}^{(\tau)}{}_{(0)} \hat{\mathscr{P}}^{0s}{}_{(\tau)}{}^{(0)} + \hat\Gamma_s{}^{(0)}{}_{(\tau)} \hat{\mathscr{P}}^{0s}{}_{(0)}{}^{(\tau)}, \qquad \text{(B4a)}
\end{aligned}
$$

$$
\begin{aligned}
(\hat{\mathscr{E}}2)^0{}_{(a)}{}^{(0)} = {} & \hat{h}^0{}_{(a)}{}^{(0)} - \hat{\mathscr{V}}^{0(0)}{}_{(a)} + {}^3\mathscr{D}_s \hat{\mathscr{P}}^{0s}{}_{(a)}{}^{(0)} \\
& - \hat\Gamma_s{}^{(0)}{}_{(a)} \hat{\mathscr{P}}^{0s}{}_{(0)}{}^{(0)} + \hat\Gamma_s{}^{(0)}{}_{(\tau)} \hat{\mathscr{P}}^{0s}{}_{(a)}{}^{(\tau)}, \qquad \text{(B4b)}
\end{aligned}
$$

$$
\begin{aligned}
(\hat{\mathscr{E}}2)^0{}_{(0)}{}^{(b)} = {} & \hat{h}^0{}_{(0)}{}^{(b)} + \hat{\mathscr{U}}^{0(b)}{}_{(0)} \\
& - \hat{\mathscr{V}}^{0(b)}{}_{(0)} + {}^3\mathscr{D}_s \hat{\mathscr{P}}^{0s}{}_{(0)}{}^{(b)} \\
& - \hat\Gamma_s{}^{(\tau)}{}_{(0)} \hat{\mathscr{P}}^{0s}{}_{(\tau)}{}^{(b)} + \hat\Gamma_s{}^{(b)}{}_{(0)} \hat{\mathscr{P}}^{0s}{}_{(0)}{}^{(0)}, \qquad \text{(B4c)}
\end{aligned}
$$

$$
\begin{aligned}
(\hat{\mathscr{E}}2)^0{}_{(a)}{}^{(b)} = {} & \hat{h}^0{}_{(a)}{}^{(b)} + \hat{\mathscr{U}}^{0(b)}{}_{(a)} \\
& - \hat{\mathscr{V}}^{0(b)}{}_{(a)} + {}^3\mathscr{D}_s \hat{\mathscr{P}}^{0s}{}_{(a)}{}^{(b)} \\
& - \hat\Gamma_s{}^{(0)}{}_{(a)} \hat{\mathscr{P}}^{0s}{}_{(0)}{}^{(b)} + \hat\Gamma_s{}^{(b)}{}_{(0)} \hat{\mathscr{P}}^{0s}{}_{(a)}{}^{(0)}. \qquad \text{(B4d)}
\end{aligned}
$$

For the dynamical equations (4.21b) we have

$$
\begin{aligned}
(\hat{\mathscr{E}}1)^k{}_{(a)} = {} & {}_{cg}\hat{\mathscr{T}}^k{}_{(a)} + {}_{cm}\hat{\mathscr{T}}^k{}_{(a)} \\
& + {}^3\mathscr{D}_0 \hat{\mathscr{U}}^{0k}{}_{(a)} + ({}^3\mathscr{D}_s + \partial_s \ln N) \hat{\mathscr{U}}^{sk}{}_{(a)} \\
& - \hat\Gamma_0{}^{(0)}{}_{(a)} \hat{\mathscr{U}}^{0k}{}_{(0)} - \hat\Gamma_s{}^{(0)}{}_{(a)} \hat{\mathscr{U}}^{sk}{}_{(0)}, \qquad \text{(B5)}
\end{aligned}
$$

$$(\hat{\mathscr{E}}2)^k{}_{(0)}{}^{(0)} = \hat{h}^k{}_{(0)}{}^{(0)} + \hat{\mathscr{U}}^{k(0)}{}_{(0)} - \hat{\mathscr{V}}^{k(0)}{}_{(0)} + {}^3\mathscr{D}_0\hat{\mathscr{P}}^{0k}{}_{(0)}{}^{(0)} + ({}^3\mathscr{D}_s + \partial_s \ln N)\hat{\mathscr{P}}^{sk}{}_{(0)}{}^{(0)} - \hat{\Gamma}_\lambda{}^{(\tau)}{}_{(0)}\hat{\mathscr{P}}^{\lambda k}{}_{(\tau)}{}^{(0)} + \hat{\Gamma}_\lambda{}^{(0)}{}_{(\tau)}\hat{\mathscr{P}}^{\lambda k}{}_{(0)}{}^{(\tau)}, \tag{B6a}$$

$$(\hat{\mathscr{E}}2)^k{}_{(a)}{}^{(0)} = \hat{h}^k{}_{(a)}{}^{(0)} + \hat{\mathscr{U}}^{k(0)}{}_{(a)} - \hat{\mathscr{V}}^{k(0)}{}_{(a)} + {}^3\mathscr{D}_0\hat{\mathscr{P}}^{0k}{}_{(a)}{}^{(0)} + ({}^3\mathscr{D}_s + \partial_s \ln N)\hat{\mathscr{P}}^{sk}{}_{(a)}{}^{(0)} - \hat{\Gamma}_\lambda{}^{(0)}{}_{(a)}\hat{\mathscr{P}}^{\lambda k}{}_{(0)}{}^{(0)} + \hat{\Gamma}_\lambda{}^{(0)}{}_{(\tau)}\hat{\mathscr{P}}^{\lambda k}{}_{(a)}{}^{(\tau)}, \tag{B6b}$$

$$(\hat{\mathscr{E}}2)^k{}_{(0)}{}^{(b)} = \hat{h}^k{}_{(0)}{}^{(b)} + \hat{\mathscr{U}}^{k(b)}{}_{(0)} - \hat{\mathscr{V}}^{k(b)}{}_{(0)} + {}^3\mathscr{D}_0\hat{\mathscr{P}}^{0k}{}_{(0)}{}^{(b)} + ({}^3\mathscr{D}_s + \partial_s \ln N)\hat{\mathscr{P}}^{sk}{}_{(0)}{}^{(b)} - \hat{\Gamma}_\lambda{}^{(\tau)}{}_{(0)}\hat{\mathscr{P}}^{\lambda k}{}_{(\tau)}{}^{(b)} + \hat{\Gamma}_\lambda{}^{(b)}{}_{(0)}\hat{\mathscr{P}}^{\lambda k}{}_{(0)}{}^{(0)}, \tag{B6c}$$

$$(\hat{\mathscr{E}}2)^k{}_{(a)}{}^{(b)} = \hat{h}^k{}_{(a)}{}^{(b)} + \hat{\mathscr{U}}^{k(b)}{}_{(a)} - \hat{\mathscr{V}}^{k(b)}{}_{(a)} + {}^3\mathscr{D}_0\hat{\mathscr{P}}^{0k}{}_{(a)}{}^{(b)} + ({}^3\mathscr{D}_s + \partial_s \ln N)\hat{\mathscr{P}}^{sk}{}_{(a)}{}^{(b)} - \hat{\Gamma}_\lambda{}^{(0)}{}_{(a)}\hat{\mathscr{P}}^{\lambda k}{}_{(0)}{}^{(b)} + \hat{\Gamma}_\lambda{}^{(b)}{}_{(0)}\hat{\mathscr{P}}^{\lambda k}{}_{(a)}{}^{(0)}, \tag{B6d}$$

$$(\hat{\mathscr{E}}3)^{(a)(b)} = {}_{sg}\hat{\mathscr{T}}^{(a)(b)} + {}_{sm}\hat{\mathscr{T}}^{(a)(b)} - {}^3\mathscr{D}_0\hat{\mathscr{V}}^{0(a)(b)} - ({}^3\mathscr{D}_s + \partial_s \ln N)\hat{\mathscr{V}}^{s(a)(b)} - \hat{\Gamma}_\lambda{}^{(a)}{}_{(0)}\hat{\mathscr{V}}^{\lambda(0)(b)} - \hat{\Gamma}_\lambda{}^{(b)}{}_{(0)}\hat{\mathscr{V}}^{\lambda(0)(a)}, \tag{B7}$$

$$(\hat{\mathscr{E}}\mathscr{M})_\Sigma = \hat{\mathscr{J}}_\Sigma - {}^3\mathscr{D}_0\hat{h}^0{}_\Sigma - ({}^3\mathscr{D}_s + \partial_s \ln N)\hat{h}^s{}_\Sigma + f_{(0)}{}^{(0)\Lambda}{}_\Sigma\hat{\Gamma}_\lambda{}^{(0)}{}_{(0)}\hat{h}^\lambda{}_\Lambda + f_{(a)}{}^{(0)\Lambda}{}_\Sigma\hat{\Gamma}_\lambda{}^{(a)}{}_{(0)}\hat{h}^\lambda{}_\Lambda + f_{(0)}{}^{(b)\Lambda}{}_\Sigma\hat{\Gamma}_\lambda{}^{(0)}{}_{(b)}\hat{h}^\lambda{}_\Lambda. \tag{B8}$$

It follows from the relations (A3) and (B4d) that the lhs of (B4a)–(B4c) are functions of the gravitational symplectic variables $m_{(\alpha)}$, $n^{(\alpha)}$, $m^{(b)}$, $n_{(b)}$, $\hat{\mathscr{U}}^{0k}{}_{(a)}$, $\hat{e}^{(a)}_k$, $\hat{\mathscr{P}}^{0k}{}_{(\alpha)}{}^{(\beta)}$, $\hat{\Gamma}_k{}^{(\alpha)}{}_{(\beta)}$, $\hat{\mathscr{V}}^{0(a)(b)}$, and $\hat{g}^{(a)(b)}$, the matter symplectic variables $\hat{h}^0{}_\Sigma$, and $\hat{\phi}^\Sigma$, as well as their spatial derivatives. Therefore, from (B4a)–(B4c) we can determine seven quantities $\hat{\mathscr{U}}^{0k}{}_{(0)}$, $\hat{\mathscr{V}}^{0(a)(0)}$ in terms of the symplectic variables and their spatial derivatives. By means of these relations we are able to express the rhs of (B3) in terms of symplectic variables and their x^k derivatives. We may say that (4.21a) are really symplectic constraints.

APPENDIX C: THE COVARIANT LIE DERIVATIVE

The covariant Lie derivative[31,36] of a tensor density of weight r on M with respect to a given connection $\zeta = (\zeta_\mu{}^{(\alpha)}{}_{(\beta)})$ is defined as

$${}_\zeta\mathscr{L}_Z\mathscr{F}^{\mu(\alpha)}_{\nu(\beta)} = Z^\lambda{}_\zeta\mathscr{D}_\lambda F^{\mu(\alpha)}_{\nu(\beta)} + r\mathscr{F}^{\mu(\alpha)}_{\nu(\beta)}\mathscr{D}_\lambda Z^\lambda - \mathscr{F}^{\tau(\alpha)}_{\nu(\beta)}\mathscr{D}_\tau Z^\mu + \mathscr{F}^{\mu(\alpha)}_{\tau(\beta)}\mathscr{D}_\nu Z^\tau, \tag{C1}$$

where ${}_\zeta\mathscr{D}$ denotes the covariant derivative (2.8) with respect to the connection ζ (in this paper ζ is a fixed background connection). It is natural to take the definition

$${}_\zeta\mathscr{L}_Z\zeta_\mu{}^{(\alpha)}{}_{(\beta)} = Z^\tau{}_\zeta R^{(\alpha)}{}_{(\beta)\tau\mu}, \tag{C2}$$

where ${}_\zeta R^{(\alpha)}{}_{(\beta)\mu\nu}$ is the curvature tensor of the connection ζ. If the conditions (4.19) are satisfied then we have the following (3 + 1) decomposition of the covariant Lie derivative for a GL(4,R)-tensor-valued covector field $\psi_\lambda{}^\Sigma$ and skew-symmetric two-covariant tensor density $\mathscr{F}^{\mu\nu}{}_\Sigma$ on space-time:

$$({}_\zeta\mathscr{L}_Z\psi_k{}^\Sigma)\hat{} = \bar{Z}^0\,{}^3_\zeta\mathscr{D}_0\hat{\psi}_k{}^\Sigma + {}^3_\zeta\mathscr{L}_Z\hat{\psi}_k{}^\Sigma \tag{C3}$$

and

$$({}_\zeta\mathscr{L}_Z\mathscr{F}^{0k}{}_\Sigma)\hat{} = \bar{Z}^0\,{}^3_\zeta\mathscr{D}_0\hat{\mathscr{F}}^{0k}{}_\Sigma + {}^3_\zeta\mathscr{L}_Z\hat{\mathscr{F}}^{0k}{}_\Sigma. \tag{C4}$$

Here ${}^3_\zeta\mathscr{L}_Z$ denotes the σ-intrinsic covariant Lie derivative for geometric objects on σ [with respect to the connection ${}^3\zeta = (\hat{\zeta}_k{}^{(a)}{}_{(b)})$ on σ, taken in the direction of the σ-parallel part ${}^{\parallel}\mathbf{Z} = \bar{Z}^k\,\partial/\partial x^k$ of $\mathbf{Z}$]. For $({}_\zeta\mathscr{L}_Z\psi_0{}^\Sigma)$ and $({}_\zeta\mathscr{L}_Z\mathscr{F}^{sk}{}_\Sigma)\hat{}$ we have more complicated and less elegant formulas[31] but fortunately such expressions do not appear in the Hamiltonian analysis.

APPENDIX D: THE TIME DERIVATIVES FOR $\hat{\gamma}$ AND $\mathfrak{g}$

For the anholonomic coefficients $\hat{\gamma}_k{}^{(a)}{}_{(b)}$ of the Riemannian connection on σ, we have

$$\hat{\gamma}_k{}^{(a)}{}_{(b)} = \bar{\gamma}_k{}^r{}_s\hat{e}^{(a)}_r\hat{e}^s_{(b)} - \hat{e}^s_{(b)}\,\partial_k\hat{e}^{(a)}_s, \tag{D1}$$

$${}^3\mathscr{D}_0\hat{\gamma}_k{}^{(a)}{}_{(b)} = \bar{\partial}_0\hat{\gamma}_k{}^{(a)}{}_{(b)} - \bar{\partial}_k\hat{\Gamma}_0{}^{(a)}{}_{(b)} + \hat{\Gamma}_0{}^{(a)}{}_{(c)}\hat{\gamma}_k{}^{(c)}{}_{(b)} - \hat{\Gamma}_0{}^{(d)}{}_{(b)}\hat{\gamma}_k{}^{(a)}{}_{(d)} - \partial_k \ln N\,\hat{\Gamma}_0{}^{(a)}{}_{(b)} - (1/N)\partial_k N^r\,\hat{\gamma}_r{}^{(a)}{}_{(b)}, \tag{D2}$$

or equivalently

$${}^3\mathscr{D}_0\gamma_k{}^{(a)}{}_{(b)} = \tfrac{1}{2}\bar{\mathfrak{g}}^{rp}\{({}^3\nabla_k + \partial_k \ln N)\,{}^3\mathscr{D}_0\bar{\mathfrak{g}}_{ps} + ({}^3\nabla_s + \partial_s \ln N)\,{}^3\mathscr{D}_0\bar{\mathfrak{g}}_{pk} - ({}^3\nabla_p + \partial_p \ln N)\,{}^3\mathscr{D}_0\bar{\mathfrak{g}}_{ks}\}\hat{e}^{(a)}_r\hat{e}^s_{(b)} - \hat{e}^s_{(b)}({}^3\nabla_k + \partial_k \ln N)\,{}^3\mathscr{D}_0\hat{e}^{(a)}_s. \tag{D3}$$

The time derivatives ${}^3\mathscr{D}_0\bar{\mathfrak{g}}_{rs}$ can be computed by means of the relations $\bar{\mathfrak{g}}_{rs} = \hat{e}^{(a)}_r\hat{e}^{(b)}_s\hat{\mathfrak{g}}_{(a)(b)}$ and formulas (5.8) and (5.13). Another method of computing of these quantities is based on the formula ${}^3\mathscr{D}_0\bar{\mathfrak{g}}_{pq} = 2\bar{\gamma}_p{}^0{}_q$ and the relations

$$\bar{\gamma}_p{}^0{}_q = \bar{\Gamma}_p{}^0{}_q - \bar{r}_p{}^0{}_q = \bar{\Gamma}_{(p}{}^0{}_{q)} - \bar{r}_{(p}{}^0{}_{q)} = \tfrac{1}{2}(\hat{\Gamma}_p{}^{(0)}{}_{(a)}\hat{e}^{(a)}_q + \hat{\Gamma}_q{}^{(0)}{}_{(a)}\hat{e}^{(a)}_p) - \bar{r}_{(p}{}^0{}_{q)},$$

$$\sqrt{\bar{\mathfrak{g}}}\,\bar{r}_p{}^0{}_q = -\tfrac{1}{4}\bar{\mathfrak{g}}_{pq}(\bar{h}^i{}_0{}^j\bar{\mathfrak{g}}_{ij} + \bar{h}^i{}_i{}^0) + \tfrac{1}{2}\bar{\mathfrak{g}}_{pi}\bar{h}^i{}_0{}^j\bar{\mathfrak{g}}_{jq} + \tfrac{1}{2}\bar{\mathfrak{g}}_{qi}\bar{h}^i{}_p{}^0 + \tfrac{1}{2}\bar{\mathfrak{g}}_{pj}\bar{h}^0{}_q{}^j. \tag{D4}$$

APPENDIX E: THE (3 + 1) DECOMPOSITION OF NONMETRICITY, TORSION, AND CURVATURE

We have

$$\hat{M}_{0(0)(0)} = \hat{\Gamma}_0{}^{(0)}{}_{(0)}, \quad \hat{M}_{0(a)(0)} = \tfrac{1}{2}(\hat{\Gamma}_0{}^{(0)}{}_{(a)} - \hat{\Gamma}_0{}^{(c)}{}_{(0)}\hat{\mathfrak{g}}_{(c)(a)}),$$
$$\hat{M}_{0(a)(b)} = \tfrac{1}{2}\,{}^3\mathscr{D}_0\hat{\mathfrak{g}}_{(a)(b)}, \quad \hat{M}_{k(a)(b)} = \tfrac{1}{2}\,{}^3\mathscr{D}_k\hat{\mathfrak{g}}_{(a)(b)} = {}^3M_{k(a)(b)},$$
$$\hat{M}_{k(a)(0)} = \tfrac{1}{2}(\hat{\Gamma}_k{}^{(0)}{}_{(a)} - \hat{\Gamma}_k{}^{(c)}{}_{(0)}\hat{\mathfrak{g}}_{(c)(a)}), \tag{E1}$$
$$\hat{M}_{k(0)(0)} = \hat{\Gamma}_k{}^{(0)}{}_{(0)};$$

$$\hat{Q}^{(0)}{}_{s0} = -\hat{\Gamma}_0{}^{(0)}{}_{(c)}\hat{e}^{(c)}_s + \partial_s \ln N + \hat{\Gamma}_s{}^{(0)}{}_{(0)},$$
$$\hat{Q}^{(0)}{}_{sr} = \hat{\Gamma}_s{}^{(0)}{}_{(c)}\hat{e}^{(c)}_r - \hat{\Gamma}_r{}^{(0)}{}_{(c)}\hat{e}^{(c)}_s,$$
$$\hat{Q}^{(a)}{}_{s0} = \hat{\Gamma}_s{}^{(a)}{}_{(0)} - {}^3\mathscr{D}_0\hat{e}^{(a)}{}_s, \tag{E2}$$
$$\hat{Q}^{(a)}{}_{sr} = {}^3Q^{(a)}{}_{sr} = {}^3\mathscr{D}_s\hat{e}^{(a)}_r - {}^3\mathscr{D}_r\hat{e}^{(a)}_s;$$

$$\hat{R}^{(0)}{}_{(0)s0} = ({}^3\mathscr{D}_s + \partial_s \ln N)\hat{\Gamma}_0{}^{(0)}{}_{(0)} - {}^3\mathscr{D}_0\hat{\Gamma}_s{}^{(0)}{}_{(0)} + \hat{\Gamma}_s{}^{(0)}{}_{(c)}\hat{\Gamma}_0{}^{(c)}{}_{(0)} - \hat{\Gamma}_0{}^{(0)}{}_{(c)}\hat{\Gamma}_s{}^{(c)}{}_{(0)},$$
$$\hat{R}^{(0)}{}_{(0)sr} = {}^3\mathscr{D}_s\hat{\Gamma}_r{}^{(0)}{}_{(0)} - {}^3\mathscr{D}_r\hat{\Gamma}_s{}^{(0)}{}_{(0)} + \hat{\Gamma}_s{}^{(0)}{}_{(c)}\hat{\Gamma}_r{}^{(c)}{}_{(0)} - \hat{\Gamma}_r{}^{(0)}{}_{(c)}\hat{\Gamma}_s{}^{(c)}{}_{(0)},$$

$$
\begin{aligned}
\hat{R}^{(a)}{}_{(0)s0} &= ({}^3\mathscr{D}_s + \partial_s \ln N)\hat{\Gamma}_0{}^{(a)}{}_{(0)} - {}^3\mathscr{D}_0 \hat{\Gamma}_s{}^{(a)}{}_{(0)} \\
&\quad + \hat{\Gamma}_s{}^{(a)}{}_{(0)}\hat{\Gamma}_0{}^{(0)}{}_{(0)} - \hat{\Gamma}_0{}^{(a)}{}_{(0)}\hat{\Gamma}_s{}^{(0)}{}_{(0)}, \\
\hat{R}^{(a)}{}_{(0)sr} &= {}^3\mathscr{D}_s \hat{\Gamma}_r{}^{(a)}{}_{(0)} - {}^3\mathscr{D}_r \hat{\Gamma}_s{}^{(a)}{}_{(0)} \\
&\quad + \hat{\Gamma}_s{}^{(a)}{}_{(0)}\hat{\Gamma}_r{}^{(0)}{}_{(0)} - \hat{\Gamma}_r{}^{(a)}{}_{(0)}\hat{\Gamma}_s{}^{(0)}{}_{(0)}, \\
\hat{R}^{(0)}{}_{(b)s0} &= ({}^3\mathscr{D}_s + \partial_s \ln N)\hat{\Gamma}_0{}^{(0)}{}_{(b)} - {}^3\mathscr{D}_0 \hat{\Gamma}_s{}^{(0)}{}_{(b)} \\
&\quad + \hat{\Gamma}_s{}^{(0)}{}_{(0)}\hat{\Gamma}_0{}^{(0)}{}_{(b)} - \hat{\Gamma}_0{}^{(0)}{}_{(0)}\hat{\Gamma}_s{}^{(0)}{}_{(b)}, \\
\hat{R}^{(0)}{}_{(b)sr} &= {}^3\mathscr{D}_s \hat{\Gamma}_r{}^{(0)}{}_{(b)} - {}^3\mathscr{D}_r \hat{\Gamma}_s{}^{(0)}{}_{(b)} \\
&\quad + \hat{\Gamma}_s{}^{(0)}{}_{(0)}\hat{\Gamma}_r{}^{(0)}{}_{(b)} - \hat{\Gamma}_r{}^{(0)}{}_{(0)}\hat{\Gamma}_s{}^{(0)}{}_{(b)}, \\
\hat{R}^{(a)}{}_{(b)s0} &= - {}^3\mathscr{D}_0 \hat{\Gamma}_s{}^{(a)}{}_{(b)} + \hat{\Gamma}_s{}^{(a)}{}_{(0)}\hat{\Gamma}_0{}^{(0)}{}_{(b)} - \hat{\Gamma}_0{}^{(a)}{}_{(0)}\hat{\Gamma}_s{}^{(0)}{}_{(b)}, \\
\hat{R}^{(a)}{}_{(b)sr} &= {}^3R^{(a)}{}_{(b)sr} + \hat{\Gamma}_s{}^{(a)}{}_{(0)}\hat{\Gamma}_r{}^{(0)}{}_{(b)} - \hat{\Gamma}_r{}^{(a)}{}_{(0)}\hat{\Gamma}_s{}^{(0)}{}_{(b)}.
\end{aligned} \tag{E3}
$$

APPENDIX F: THE BASIC GRAVITATIONAL GAUGE GROUP

Let $[L^{(\alpha)}{}_{(\beta)}(\cdot)]$ be an element of the Lie algebra of the local GL(4,R) group. Such an element generates infinitesimal transformations δ_L of field potentials (3.2). If $\mathbf{Z}$ is a vector field on space-time, that is, an element of the Lie algebra of the diffeomorphism group and $\zeta = (\zeta_\mu{}^{(\alpha)}{}_{(\beta)})$ is a fixed connection then the infinitesimal operators $\delta_Z = {}_\zeta\mathscr{L}_Z$ are defined by (C1) and (C2). These operators satisfy the following commutation relations:

$$
\begin{aligned}
&[\delta_{Z_1}, \delta_{Z_2}] = \delta_{[Z_1, Z_2]} + \delta_{L_3}, \\
&[\delta_{L_1}, \delta_{L_2}] = \delta_{[L_1, L_2]}, \ [\delta_L, \delta_Z] = \delta_{L_4},
\end{aligned} \tag{F1}
$$

where $[Z_1, Z_2]$ is the commutator of vector fields Z_1, Z_2; $[L_1, L_2]$ is the commutator of corresponding matrices (the right multiplication); and

$$
\begin{aligned}
L_3{}^{(\alpha)}{}_{(\beta)} &= - {}_\zeta R^{(\alpha)}{}_{(\beta)\mu\nu} Z_1^\mu Z_2^\nu, \\
L_4{}^{(\alpha)}{}_{(\beta)} &= - Z^\lambda {}_\zeta\mathscr{D}_\lambda L^{(\alpha)}{}_{(\beta)}.
\end{aligned}
$$

The relations (F1) satisfy the Jacobi identity and the set of operators δ_Z, δ_L carries the structure of an infinite-dimensional Lie algebra $\mathscr{g}$. Let us observe that the operators of infinitesimal rotations δ_L form an ideal in $\mathscr{g}$ but the operators of infinitesimal translations δ_Z even do not form a subalgebra. It is clear that if we integrate the algebra $\mathscr{g}$ to a group G then the operators δ_L generate local GL(4,R) rotations in the space of field potentials. Now we explain how to define GL(4,R)-covariant translations. First of all, we have to assume that there exists a one-to-one correspondence between a neighborhood $\mathscr{O}_0$ of zero in the (vector) space of vector fields on M and a neighborhood $\mathscr{U}_{\text{id}}$ of the identity in Diff M called the exponential mapping exp: $\mathscr{O}_0 \rightarrow \mathscr{U}_{\text{id}}$.

Let Z belong to $\mathscr{O}_0$ and $\Phi = \exp(Z)$. For the unit interval we have the curve $[0,1] \ni t \rightarrow \Phi_t = \exp(tZ)$ in Diff M joining id with Φ. Let $\mathbf{x}$ be a point in space-time. We have the curve $t \rightarrow \mathbf{x}(t) = \Phi_t(\mathbf{x})$ in M joining $\mathbf{x}$ and $\Phi(\mathbf{x})$. The covariant translation of geometric objects with purely holonomic indices (space-time objects) from the point $\mathbf{x}$ to the point $\Phi(\mathbf{x})$ is accomplished by the standard action of Diff M. The covariant translation of geometric objects corresponding to appropriate representations of GL(4,R) [GL(4,R) tensors or connection coefficients] consists in their ζ-parallel transport along the curve $t \rightarrow \mathbf{x}(t)$ from the point $\mathbf{x}$ to the point $\Phi(\mathbf{x})$. Now it is clear that if curvature of ζ is not trivial then the composition of two covariant translations is not a pure translation but can be obtained as the composition of a covariant translation and the local rotation corresponding to an element of the holonomy group of ζ [cf. (F1)].

The set $G_{\mathscr{U}_{\text{id}}}$ of pairs $(\mathscr{G}, \Phi)$ where $\mathscr{G}$ is the field of GL(4,R) matrices and $\Phi \in \mathscr{U}_{\text{id}} \subset$ Diff M carries the natural bundle structure over $\mathscr{U}_{\text{id}} \subset$ Diff M. The fiber over a fixed point of the base is isomorphic to the local GL(4,R) group. We write $G_{\mathscr{U}_{\text{id}}} = \text{loc GL}(4,R) \times_b \mathscr{U}_{\text{id}}$ (the bundle product) or formally $G = \text{loc GL}(4,R) \times_b \text{Diff } M$ [G is the basic gravitational gauge group for GL(4,R) theories and it is parametrized by 20 "functions" on space-time].

We would like to emphasize that the above given construction requires an appropriate choice of topology and the differentiable structure in the space of vector fields on M as well as in Diff M in order to assure the existence of an exponential mapping. Necessary mathematical foundations are given in Ref. 44.

APPENDIX G: THE ANALYSIS OF THE ZERO EQUATION IN THE SET (6.12)

In the (3 + 1)-variables we get, from (6.12),

$$
\begin{aligned}
(\mathscr{D}_0(\mathscr{D}_\lambda h^\lambda{}_{(\alpha)}{}^{(\alpha)})) &= \sqrt{\bar{g}}\, {}^3\mathscr{D}_0((1/\sqrt{\bar{g}})(\hat{\jmath}_{(a)}\hat{f}^{(a)} + \hat{\jmath}_{(0)}\hat{f}^{(0)})) - (2/A\sqrt{\bar{g}})\{\hat{\mathscr{P}}^0{}_{(0)}(-{}^3\nabla_s \hat{\mathscr{P}}^s{}_{(0)} + \bar{\gamma}_s{}^s{}_0 \hat{\mathscr{P}}^0{}_{(0)} - \hat{\Gamma}_s{}^{(\alpha)}{}_{(0)}\hat{\mathscr{P}}^s{}_{(\alpha)} - \hat{\jmath}_{(0)}) \\
&\quad - \hat{\mathscr{P}}^s{}_{(0)}({}^3\nabla_s \hat{\mathscr{P}}^0{}_{(0)} + \bar{\gamma}_s{}^0{}_r \hat{\mathscr{P}}^r{}_{(0)} - \hat{\Gamma}_s{}^{(\alpha)}{}_{(0)}\hat{\mathscr{P}}^0{}_{(\alpha)}) - \hat{\mathscr{P}}^{0(a)}({}^3\mathscr{D}_s \hat{\mathscr{P}}^s{}_{(a)} - \bar{\gamma}_s{}^s{}_0 \hat{\mathscr{P}}^0{}_{(a)} - \hat{\Gamma}_s{}^{(0)}{}_{(a)}\hat{\mathscr{P}}^s{}_{(0)} - \hat{\jmath}_{(a)}) \\
&\quad + \hat{\mathscr{P}}^{s(a)}({}^3\mathscr{D}_s \hat{\mathscr{P}}^0{}_{(a)} + \bar{\gamma}_s{}^0{}_r \hat{\mathscr{P}}^r{}_{(a)} - \hat{\Gamma}_s{}^{(0)}{}_{(a)}\hat{\mathscr{P}}^0{}_{(0)}) + \hat{\Gamma}_0{}^{(0)}{}_{(0)}(-(\hat{\mathscr{P}}^0{}_{(0)})^2 - \hat{\mathscr{P}}^s{}_{(0)}\hat{\mathscr{P}}^r{}_{(0)}\bar{g}_{sr} + \hat{\mathscr{P}}^0{}_{(a)}\hat{\mathscr{P}}^{0(a)} \\
&\quad + \hat{\mathscr{P}}^s{}_{(a)}\hat{\mathscr{P}}^{r(a)}\bar{g}_{sr}) + \hat{\mathscr{P}}^0{}_{(a)}\hat{\mathscr{P}}^0{}_{(b)}(\tfrac{1}{2}\,{}^3\mathscr{D}_0 \hat{g}^{(a)(b)} - \hat{g}^{(a)(b)}\hat{\Gamma}_0{}^{(0)}{}_{(0)}) + \hat{\mathscr{P}}^s{}_{(a)}\hat{\mathscr{P}}^r{}_{(b)}\bar{g}_{sr}(\tfrac{1}{2}\,{}^3\mathscr{D}_0 \hat{g}^{(a)(b)} - \hat{g}^{(a)(b)}\hat{\Gamma}_0{}^{(0)}{}_{(0)}) \\
&\quad + (\hat{\mathscr{P}}^0{}_{(0)}\hat{\mathscr{P}}^0{}_{(a)} + \hat{\mathscr{P}}^s{}_{(a)}\hat{\mathscr{P}}^r{}_{(0)}\bar{g}_{sr})(\hat{\Gamma}_0{}^{(0)(a)} - \hat{\Gamma}_0{}^{(a)}{}_{(0)}) + (\hat{\mathscr{P}}^0{}_{(0)}\hat{\mathscr{P}}^s{}_{(a)} + \hat{\mathscr{P}}^0{}_{(a)}\hat{\mathscr{P}}^s{}_{(0)})(\hat{\Gamma}_s{}^{(0)(a)} - \hat{\Gamma}_s{}^{(a)}{}_{(0)}) \\
&\quad - 2\hat{\Gamma}_s{}^{(0)}{}_{(0)}\hat{\mathscr{P}}^0{}_{(0)}\hat{\mathscr{P}}^s{}_{(0)}\} + 2\hat{R}^{(\alpha)}{}_{(\beta)0s}\hat{f}^{(\beta)}\hat{\mathscr{P}}^s{}_{(\alpha)} = 0.
\end{aligned} \tag{G1}
$$

Applying the dynamical equations (5.13b) and relations for $\bar{\gamma}_s{}^0{}_r$ (D4) we obtain a linear equation for $\hat{\Gamma}_0{}^{(0)}{}_{(0)}$. If we make use of relations (E3) we see that the term $\hat{R}^{(\alpha)}{}_{(\beta)0s}\hat{f}^{(\beta)}\hat{\mathscr{P}}^s{}_{(\alpha)}$ contains not only $\hat{\Gamma}_0{}^{(0)}{}_{(0)}$ but also $\nabla_s \hat{\Gamma}_0{}^{(0)}{}_{(0)}$. Therefore, at first glance, the Eq. (G1) is a linear first-order differential equation for $\hat{\Gamma}_0{}^{(0)}{}_{(0)}$. In order to prove that (G1) is, in fact, an algebraic equation for $\hat{\Gamma}_0{}^{(0)}{}_{(0)}$ we make the following observations.

(i) In virtue of formulas (6.5b), (D2), (E3), $\hat{R}^{(\alpha)}{}_{(\beta)0s}\hat{f}^{(\beta)}\hat{\mathscr{P}}^{s}{}_{(\alpha)}$ is a function of symplectic variables, $\partial_s \ln N$, spatial derivatives of these variables, as well as of ${}^3\mathscr{D}_0(\hat{r}_s{}^{(a)}{}_{(b)} - \hat{\Gamma}_s{}^{(0)}{}_{(0)}\delta^{(a)}{}_{(b)})$, ${}^3\mathscr{D}_0\hat{\Gamma}_{[(a)}{}^{(0)}{}_{(b)]}$, and $({}^3\mathscr{D}_0\hat{\Gamma}_s{}^{(0)}{}_{(0)} - {}^3\nabla_s\hat{\Gamma}_0{}^{(0)}{}_{(0)})$.

(ii) On the other hand, Eq. (5.12c), (5.14), and (6.7b) can be rewritten in terms of the symplectic variables as well as of $({}^3\nabla_s\hat{f}^{(0)} + \hat{\Gamma}_s{}^{(0)}{}_{(0)}\hat{f}^{(0)})$, $({}^3\nabla_s\hat{f}^{(a)} + \hat{\Gamma}_s{}^{(0)}{}_{(0)}\hat{f}^{(a)})$, $\hat{\Gamma}_{[(a)}{}^{(0)}{}_{(b)]}$, and $(\hat{r}_s{}^{(a)}{}_{(b)} - \hat{\Gamma}_s{}^{(0)}{}_{(0)}\delta^{(a)}{}_{(b)})$. Moreover, the dependence of these equations on the variables belonging to the last four groups is linear.

(iii) We take ${}^3\mathscr{D}_0$ derivatives of the rewritten equations (5.12c), (5.14), and (6.7b) and make use of the relations

$$
\begin{aligned}
&{}^3\mathscr{D}_0({}^3\nabla_s\hat{f}^{(0)} + \hat{\Gamma}_s{}^{(0)}{}_{(0)}\hat{f}^{(0)}) \\
&\quad = ({}^3\mathscr{D}_0\hat{\Gamma}_s{}^{(0)}{}_{(0)} - {}^3\nabla_s\hat{\Gamma}_0{}^{(0)}{}_{(0)})\hat{f}^{(0)} + C, \\
&{}^3\mathscr{D}_0({}^3\nabla_s\hat{f}^{(a)} + \hat{\Gamma}_s{}^{(0)}{}_{(0)}\hat{f}^{(a)}) \\
&\quad = ({}^3\mathscr{D}_0\hat{\Gamma}_s{}^{(0)}{}_{(0)} - {}^3\nabla_s\hat{\Gamma}_0{}^{(0)}{}_{(0)})\hat{f}^{(a)} + D,
\end{aligned}
\tag{G2}
$$

where C and D denote terms depending on the symplectic variables, $\partial_s \ln N$, first and second spatial derivatives of these quantities, and linearly on $\hat{\Gamma}_0{}^{(0)}{}_{(0)}$.

(iv) We obtain a system of linear equations for unknowns

$$
\begin{aligned}
&{}^3\mathscr{D}_0(\hat{r}_s{}^{(a)}{}_{(b)} - \hat{\Gamma}_s{}^{(0)}{}_{(0)}\delta^{(a)}{}_{(b)}), \\
&\quad {}^3\mathscr{D}_0\hat{\Gamma}_{[(a)}{}^{(0)}{}_{(b)]}, \quad ({}^3\mathscr{D}_0\hat{\Gamma}_s{}^{(0)}{}_{(0)} - {}^3\nabla_s\hat{\Gamma}_0{}^{(0)}{}_{(0)}).
\end{aligned}
\tag{G3}
$$

If the regularity conditions (6.8) hold then this system can be solved in a similar way as was solved the system (5.12c), (5.14), and (6.7b) for $\hat{\Gamma}_s{}^{(a)}{}_{(b)}$, $\hat{\Gamma}_{[(a)}{}^{(0)}{}_{(b)]}$, and $\hat{\Gamma}_s{}^{(0)}{}_{(0)}$. Thus, the unknowns (G3) are functions of symplectic variables, $\partial_s \ln N$, first and second spatial derivatives of these quantities, as well as of $\hat{\Gamma}_0{}^{(0)}{}_{(0)}$ (with the linear dependence on $\hat{\Gamma}_0{}^{(0)}{}_{(0)}$).

(v) In virtue of (i) and (iv) the quantity $\hat{R}^{(\alpha)}{}_{(\beta)0s}\hat{f}^{(\beta)}\hat{\mathscr{P}}^{(s)}{}_{(\alpha)}$ depends on $\hat{\Gamma}_0{}^{(0)}{}_{(0)}$ linearly and it does not depend on $\nabla_s\hat{\Gamma}_0{}^{(0)}{}_{(0)}$!

We see that (G1) is a linear algebraic equation for $\hat{\Gamma}_0{}^{(0)}{}_{(0)}$ and, eventually, $\hat{\Gamma}_0{}^{(0)}{}_{(0)}$ is a rational function of symplectic variables, $\partial_s \ln N$ as well as of first and second spatial derivatives of these quantities.

[1] A. Salam, Rev. Mod. Phys. **52**, 525 (1980).

[2] S. W. Hawking and W. Israel, in *General Relativity–An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U. P., Cambridge, 1979).

[3] J. Kijowski, Gen. Relativ. Gravit. **9**, 857 (1978); M. Ferraris and J. Kijowski, Gen. Relativ. Gravit. **14**, 165 (1982).

[4] F. W. Hehl, G. D. Kerlick, and P. von der Heyde, Z. Naturforsch. A **31**, 111, 524, 823 (1976); Phys. Lett. B **63**, 446 (1976); F. W. Hehl and G. D. Kerlick, Gen. Relativ. Gravit. **9**, 691 (1978); F. W. Hehl, E. A. Lord, and L. L. Smalley, Gen. Relativ. Gravit. **13**, 1037 (1981).

[5] W. Szczyrba, (a) Lett. Math. Phys. **2**, 265 (1978); (b) Commun. Math. Phys. **60**, 215 (1978).

[6] H. Weyl, Z. Phys. **56**, 330 (1929).

[7] R. Geroch, J. Math. Phys. **9**, 1739 (1968); **11**, 343 (1970).

[8] D. W. Sciama, in *Recent Developments in General Relativity* (Pergamon–PWN, Oxford–Warsaw, 1964); T. W. Kibble, J. Math. Phys. **2**, 212 (1961).

[9] R. Utiyama, Phys. Rev. **101**, 1597 (1956).

[10] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

[11] A. Trautman, Symp. Math. **12**, 139 (1973).

[12] A. Trautman, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980); Acta Phys. Austriaca **23**, Suppl. 401 (1981); in *Lecture Notes in Mathematics 926* (Springer, Berlin, 1982).

[13] Y. Ne'eman, in *Lecture Notes in Mathematics 676* (Springer, Berlin, 1978).

[14] Y. Ne'eman, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980).

[15] J. Hennig and J. Nitsch, Gen. Relativ. Gravit. **13**, 947 (1981).

[16] F. W. Hehl, P. von der Heyde, G. D. Kerlick, and J. M. Nester, Rev. Mod. Phys. **48**, 393 (1976).

[17] F. W. Hehl, in *Cosmology and Gravitation*, edited by P. Bergman and V. DeSabbata (Plenum, New York, 1980).

[18] F. W. Hehl, J. Nitch, and P. von der Heyde, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 1.

[19] S. W. MacDowell and F. Mansouri, Phys. Rev. Lett. **38**, 739 (1977).

[20] L. N. Chang and F. Mansouri, Phys. Rev. D **17**, 3168 (1978).

[21] Y. Ne'eman and T. Regge, Rev. Nuovo Cimento **1**, 1 (1978); Y. Ne'eman, in *Cosmology and Gravitation*, edited by P. Bergmann and V. De Sabbata (Plenum, New York, 1980).

[22] A. Tseytlin, Phys. Rev. D. **26**, 3327 (1982); L. Castellani, P. Fré, and P. van Nieuwenhuizen, Ann. Phys. (N Y) **136**, 398 (1981).

[23] J. M. Nester, Ph. D. dissertation, University of Maryland, 1977 (unpublished).

[24] C. N. Yang, Phys. Rev. Lett. **33**, 445 (1974).

[25] E. E. Fairchild, Jr., Phys. Rev. D **14**, 384 (1976); Phys. Rev. D **16**, 2438 (1977).

[26] R. Rauch and H. T. Nieh, Phys. Rev. D **24**, 2029 (1981); R. Rauch, J. C. Shaw, and H. T. Nieh, Gen. Relativ. Gravit. **14**, 331 (1982); S. Rawaswamy and P. B. Yasskin, Phys. Rev. D **19**, 2264 (1979).

[27] P. B. Yasskin, Ph. D. dissertation, University of Maryland, 1979 (unpublished).

[28] P. Baekler, F. W. Hehl, and E. W. Mielke, in *Proceedings of the Second Marcel Grossmann Meeting on General Relativity*, edited by R. Ruffini (North-Holland, Amsterdam, 1982); P. Baekler, F. W. Hehl, and H. J. Lentzen, to appear in *Proceedings of the Third Marcel Grossmann Meeting on the Recent Developments in General Relativity*, Shanghai, 1982; P. Baekler and F. W. Hehl, Phys. Lett. A **100**, 392 (1984); J. D. McCrea, Phys. Lett. A**100**, 397 (1984); J. Phys. A **16**, 997 (1983).

[29] E. Sezgin and P. van Nieuwenhuizen, Phys. Rev. D **21**, 3269 (1980); K. Hayashi and T. Shirafuji, Prog. Theor. Phys. **64**, 866, 883, 1435, 2222 (1980).

[30] W. Szczyrba, Phys. Rev. D **25**, 2548 (1982).

[31] (a) W. Szczyrba, Ann. Phys. (NY) **158**, 320 (1984); (b) "Tensor matter fields in the ECSK theory," preprint, Warsaw, 1983.

[32] W. Szczyrba, "The canonical dynamics of the Yang theory of gravity," preprint, Warsaw 1984.

[33] M. Blagojević and I. Nikolić, Phys. Rev. D **28**, 2455 (1983).

[34] F. W. Hehl, E. A. Lord, and Y. Ne'eman, Phys. Rev. D **17**, 428 (1978); Phys. Lett. B **71**, 432 (1977).

[35] W. Szczyrba, J. Math. Phys. **22**, 1926 (1981).

[36] M. Antonowicz and W. Szczyrba, "Spinor matter fields in SL (2,C)-gauge theories of gravity," Phys. Rev. D (in press).

[37] Y. Ne'eman and Dj. Sijački, Ann. Phys. (NY) **120**, 292 (1979).

[38] J. Mickelsson, Comm. Math. Phys. **88**, 551 (1983).

[39] J. A. Schouten, *Ricci Calculus* (Springer, Berlin, 1954).

[40] L. Rosenfeld, Acad. R. Belgique **18**, (1940); F. J. Belinfante, Physica **6**, 887 (1939); W. Pauli, Rev. Mod. Phys. **13**, 203 (1941).

[41] E. A. Lord, Phys. Lett. A **65**, 1 (1978).

[42] J. Kijowski and W. Szczyrba, Commun. Math. Phys. **46**, 183 (1976).

[43] V. D. Sandberg, Phys. Rev. D **12**, 3013 (1975).

[44] D. G. Ebin and J. E. Marsden, Ann. Math. **92**, 102 (1970); H. Omori, in *Lecture Notes in Mathematics 427* (Springer, Berlin, 1974).

A physical interpretation of the Kerr solution

Brian Punsly
Department of Physics, University of California, Los Angeles, California 90024

(Received 25 July 1984; accepted for publication 15 March 1985)

The pseudo-Riemannian nature of the Kerr solution is exploited to obtain a special relativistic analysis. Lorentz contraction and time dilation factors are found that represent the metric coefficients. The special relativistic treatment is extended to the study of the hypersurface at $r = 0$. A nontrivial topology is revealed and new extensions of the metric are found.

I. INTRODUCTION

One of the unsatisfying features of the general theory of relativity is that it admits numerous spurious solutions. Any problem is inevitably solved by keeping those solutions that correspond to realistic stress energy tensors and eliminating the vast majority of solutions that do not. Every symmetric metric can be used to solve Einstein's equations. Hence, just the mere existence of Einstein's divergenceless tensor is not a complete theory, as it gives no criteria for culling through the countless possible solutions.

Special relativity, on the other hand, is the most accurately tested of physical theories. One of the strongest motivations for the general theory is that "locally" special relativity should hold throughout space-time. This form of the equivalence principle is insured by choosing a pseudo-Riemannian manifold to represent space-time.

The only solution of Einstein's equations that has been experimentally verified (even though, only in an asymptotic sense) is the Schwarzschild solution for the external gravitational field of a nonrotating mass. For both the Schwarschild and Kerr metrics, the right-hand side of Einstein's equations vanishes. The equations take on the degenerate form $R^{\mu\nu} = 0$. Alternatively, one might view these vacuum solutions as those that inject a minimal amount of curvature (i.e., Ricci $= 0$ and Riemann $\neq 0$) and are asymptotically flat as well.

Perhaps, equations are not applicable to all physical problems. Certain geometries might be fundamental to nature due to their symmetries. In order to investigate this possibility, the pseudo-Riemannian nature of the Kerr geometry will be analyzed in this article. In particular, locally flat coordinate patches (flat at one point) will be examined. The motivation for this is to define a description of the Kerr solution that is independent of general relativity, but relies heavily on special relativity. In an effort to raise the level of this article above that of a pedagogical exercise, it will be shown using the methods developed in the text of the paper, that the geometry of the hypersurface at $r = 0$, in the Kerr space-time, has many interesting properties. It is not just a disklike obstruction as is assumed in the maximal analytic extension of Boyer and Lindquist.[1] The fascinating geometry of this surface implies that it is a source for the external geometry. This might be the reason that the Kerr solution is physically important, regardless of any theory that "predicts" this solution.

Our method will be as follows. The Kerr solution will be broken down into locally flat regions that are realized by orthonormal bases that are carried by freely falling observers. At each point of space-time, there exist such flat frames in which special relativity holds (at one point). By looking at time dilation and Lorentz contraction in the appropriate sense, the global Kerr solution will be reconstructed (the feasibility of such a procedure is suggested in the book by Misner, Thorne, and Wheeler).[2]

In order to do this, one needs a globally well-defined, continuous frame field which can be achieved momentarily by a Lorentz boost from the freely falling frames at each point of space-time (momentarily, since in general there will be accelerations and precessions between the two frames). This global frame should give a natural $3 + 1$ splitting of space-time. Such a splitting is possible if three of the four vector fields of the tetrad span a completely integrable distribution in the four-dimensional space-time. An equivalent condition for this is that the fourth leg of the tetrad is a hypersurface orthogonal vector field. Thus, the image of the freely falling frames under Lorentz transformation in the completely integrable frame will locally define the tangent space to a manifold at each point of space-time. Consequently, the continuity of the global frame field combined with the fact that the three-dimensional submanifolds foliate space-time (actually, when horizons are involved, one has to split space-time into various regions) allows the application of the Gauss–Codacci relations to the whole manifold to get back the Kerr metric.

The frame field that can be chosen to do this is not unique. However, certain ones lend themselves to physical interpretation. For example, in the limit of infinite r, in Boyer–Lindquist coordinates, the zero angular momentum frames become the stationary frames at infinity. Also, it will turn out that the local boost that realizes the zero angular momentum frame (to be denoted as the 0-L frame from now on) is characterized by a velocity less than the speed of light all the way from asymptotic infinity to the event horizon in the case $a < m$ and all the way to $r = 0$, when $a > m$. For these reasons, the 0-L frames will be used in this article.

It should be noted that the ideas in this paper are coordinate independent, since they depend only on orthonormal frame fields. However, in order to do many of the calculations one must analyze the geometry in some coordinate system. The Boyer–Lindquist coordinates are chosen, since they lend themselves to physical interpretation, as they are asymptotically spherical coordinates in the usual sense.

II. THE SCHWARZSCHILD SOLUTION

As a simple example of the methods of this paper, one can analyze the Schwarzschild solution outside of the event

 0022-2488/85/071728-12$02.50

horizon. In Schwarzschild coordinates, the metric is

$$ds^2 = -(1-2m/r)dt^2 + r^2\sin^2\theta\, d\phi^2 + r^2\, d\theta^2 + (1-2m/r)^{-1}\, dr^2. \tag{2.1}$$

It was implied in the introductory remarks that certain frames hold a more fundamental role in the geometry than others. In particular, the freely falling frames are the most fundamental as they "move" with the local geometry. For the special relativistic treatment to follow, these frames will be associated with the lab frame as they are at rest with respect to the local geometry. For a freely falling observer that is released from rest at infinity, his tetrad will be labeled by e_α, with $e_\alpha \cdot e_\beta = \eta_{\alpha\beta}$ and $(e_\alpha \cdot e_\beta)_{,\nu} = 0$ at all points along his world-line (the connection coefficients vanish in this frame along his world-line). Thus, special relativity holds in this frame on his world-line.

The choice of a frame that can be integrated to foliate space-time is a very natural one in these coordinates, since $\partial/\partial t$ is a hypersurface orthogonal vector field. The orthonormal frames with four-velocity $\partial/\partial t$ are the static frames. The legs of this tetrad will be labeled as e'_α.

At each point of the world-line of a freely falling observer, there exists a boost to the static frame at that point. Using the relation for the four-velocity, $u \cdot u = -1$, it can be shown that the radial velocity of a particle released from rest at infinity is

$$v^r = u^r/u^0 = (1-2m/r)(2m/r)^{1/2},$$

as measured from infinity.[3] The locally measured velocity between the static frame at r and the freely falling frame at that point can be found using clocks that flow at the local rate $d\hat{t} = (1-2m/r)^{1/2}\, dt$ and the locally measured distance element $d\hat{r} = (1-2m/r)^{-1/2}\, dr$. These relations give the locally measured velocity between the two frames at r (the boost velocity)[3]

$$\frac{d\hat{r}}{d\hat{t}} = \left(\frac{1-2m}{r}\right)^{-1}\frac{dr}{dt} = \left(\frac{2m}{r}\right)^{1/2}. \tag{2.2}$$

It is interesting that this is the precise value predicted by Newtonian theory.

The boost velocity allows one to determine the image of the freely falling frame in the basis of the static frame at the coordinate value of r by means of the transformation laws of special relativity

$$e'_0 = \gamma e_0 + \gamma v e_r, \tag{2.3a}$$

$$e'_r = \gamma v e_0 + \gamma e_r. \tag{2.3b}$$

Each tetrad is composed of four vector fields. Two of the vector fields are unaltered by the boost, but the "0" and "r" fields are affected as in (2.3). The flow of time in the inertial frame is characterized by a rate corresponding to the length of the vector e_0. If time flows at a unit rate in the inertial frame, it appears to flow at a rate γ^{-1} in the static frame. This corresponds to the projection on the timelike vector field e_0, $\gamma^{-1}e'_0$, with all other $e_i = 0$ (time dilation).

To measure the length of the image of the vector field e_r in the static frame, the base and tip of the vector e_r must be simultaneous in the frame of measurement. Consequently, one must use the inverse transformations

$$e_0 = \gamma e'_0 - \gamma v e'_r, \tag{2.4a}$$

$$e_r = -\gamma v e'_0 + \gamma e'_r, \tag{2.4b}$$

with $e'_0 = 0$ being the condition of simultaneity. A unit flow along the e_r direction in the freely falling frame corresponds to a flow $\gamma e'_r$ in the static frame. The image of the freely falling tetrad as viewed in the static frame can be pictured as in Fig. 1. It is convenient to label the legs of this image by E_α (i.e., $E_0 = \gamma^{-1}e'_0$ and $E_r = \gamma e'_r$).

The images of the three spacelike legs of the freely falling tetrad represent the tangent space to a manifold or equivalently, the best linear approximation to the manifold in a neighborhood of each point. The image of the fourth leg is the normal vector to the manifold and represents the rate of flow of proper time. The whole manifold can be pieced together by using the Gauss–Codacci relations. In order to implement these equations, one needs to know the metric of the three-dimensional submanifold that is generated by the completely integrable distribution. First, using (2.2)

$$\gamma^2 = (1-2m/r)^{-1}. \tag{2.5}$$

Then, the metric of the three-dimensional submanifold is

$$g^{(3)}_{ik} = E_i \cdot E_k = r^2\sin^2\theta\, d\phi^2 + r^2\, d\theta^2 + \gamma^2\, dr^2. \tag{2.6}$$

The normal vector to this three-surface is $E_0 = \gamma^{-1}e'_0$. If one writes this vector as a unit vector in Schwarzschild coordinates

$$n^\alpha = N^{-1}(1, -N^m),$$

then the Gauss–Codacci relations give a four-metric[2]

$$g^{(4)}_{\mu\nu} = \begin{vmatrix} N^sN_s - N^2 & N_s \\ N_s & g^{(3)}_{ik} \end{vmatrix} = \begin{vmatrix} -(1-2m/r) & 0 \\ 0 & g^{(3)}_{ik} \end{vmatrix}. \tag{2.7}$$

Equation (2.7) is the same as (2.1).

This procedure is no longer defined at the event horizon, where the boost velocity approaches c. This is what one would expect as this is the static limit.

III. STATIONARY FRAMES

One might initially feel that since the observers with four-velocity $\partial/\partial t$ worked in the treatment of the Schwarzschild metric that they would work in the Kerr solution. However, the stationary frames (with four-velocity $\partial/\partial t$) do not play the same role in the Kerr geometry, since $\partial/\partial t$ is not a hypersurface orthogonal vector field. As an example of

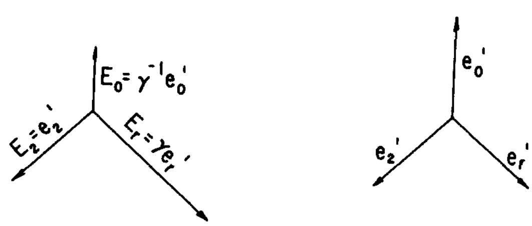


FIG. 1. The image of the freely falling tetrad (left) in the static basis (right), with one dimension suppressed.

why this fails, consider the case of the rotating disk as first discussed by Einstein.[4] Zel'dovich and Novikov show that the metric becomes[5]

$$ds^2 = (c^2 - \Omega^2 r'^2)dt^2 - 2\Omega r'^2 \, d\phi' \, dt$$
$$- dz'^2 - r'^2 \, d\phi'^2 - dr'^2.$$

If one tries to interpret the disk as being a spacelike submanifold that is orthogonal to the timelike vector field $\partial/\partial t$ then one encounters the ambiguity that clocks cannot be synchronized on the disk. The reason for this is that the three legs of the tetrad that are orthogonal to $\partial/\partial t$ do not lie in the tangent space to a manifold, since $\partial/\partial t$ is not hypersurface orthogonal. In particular X_1, X_2, and X_3 do not close under commutation, $[X_i, X_j] \neq C^k_{ij} X_k$, where each of the i, j, or k can run over the spacelike indices 1, 2, or 3.

As a consequence, one would assume that the method of this article applied to these frames is a fruitless enterprise. Actually, it is not. Utilizing the results of Appendix A, one can derive Carter's equations of motion for freely falling frames. However, no global analysis can be deduced from the stationary frames.

In a space-time with a metric containing off-diagonal terms, the spatial distance element is given by[6]

$$dl^2 = [g_{\alpha\beta} - (g_{0\alpha} g_{0\beta})/g_{00}] dx^\alpha \, dx^\beta. \tag{3.1}$$

The Kerr metric is given in Boyer–Lindquist coordinates by the line element

$$ds^2 = \rho^2 (dr^2/\Delta + d\theta^2)$$
$$+ [(r^2 + a^2)\sin^2\theta + (2mra^2/\rho^2)\sin^4\theta] d\phi^2$$
$$- (4mra \sin^2\theta/\rho^2) d\phi \, dt - (1 - 2mr/\rho^2) dt^2, \tag{3.2}$$

where $\rho^2 = r^2 + a^2 \cos^2\theta$, a is the angular momentum per unit mass, and $\Delta = r^2 - 2mr + a^2$. Consequently, azimuthal displacements are found by combining (3.1) and (3.2)

$$dl^2 = [(r^2 + a^2 - 2mr)/(1 - 2mr/\rho^2)] d\phi^2, \tag{3.3}$$

$$dl^2 = g_{rr}^{-1} g_{tt}^{-1} \rho^2 \sin^2\theta \, d\phi^2 \equiv L^2_\phi \, d\phi^2. \tag{3.4}$$

In order to interpret relation (3.4), one can call on a fact about orthogonal boosts from special relativity.[3] If $\gamma_1^{-2} = (1 - \beta_1^2)$ and $\gamma_2^{-2} = (1 - \beta_2^2)$, then the combined orthogonal boost Lorentz contraction factor is given by

$$\gamma^{-2} = \gamma_1^{-2} \gamma_2^{-2}. \tag{3.5}$$

So if $g_{tt}^{-1} = \gamma^2$ and $g_{rr}^{-1} = \gamma_r^{-2}$ as in the Schwarzschild case and if (3.5) is written as $\gamma^2 = \gamma_\phi^2 \gamma_r^2$, then

$$dl^2 = \gamma_\phi^2 \rho^2 \sin^2\theta \, d\phi^2. \tag{3.6}$$

This looks like the Lorentz contraction due to rotational motion with a lever arm length of $\rho \sin\theta$.

One can obtain Carter's equations of motion for a freely falling particle by using Eq. (A3) of Appendix A. These Wigner rotated velocities are appropriate since Carter's equations essentially piece together local coordinate patches along the world-line of the freely falling particle as viewed in the global frame of the stationary observers. The Wigner rotation is important for labeling coordinates and directions because each coordinate patch is Wigner rotated by a different amount with respect to the stationary frames.

In order to implement (A3), the boost velocities must be written in terms of a local coordinate basis. Using the definition of L_ϕ in (3.4), the obvious choice of basis is

$$e_\phi = L_\phi^{-1} \frac{\partial}{\partial\phi} = (\gamma_\phi \rho \sin\theta)^{-1} \frac{\partial}{\partial\phi} \text{ and } e_r = \gamma_r^{-1} \frac{\partial}{\partial r}, \tag{3.7}$$

where γ_ϕ and γ_r are taken to be constant in the local coordinate neighborhood. Then (3.6), (3.5), and (A3) (with $\tilde{v}_1 = \tilde{v}_r$ and $\tilde{v}_2 = \tilde{v}_\phi$) combine to give the four-velocity of a freely falling particle

$$u^r = \frac{(2mr - a^2 \sin^2\theta)^{1/2}}{\Delta^{1/2}} \gamma_r^{-1} \frac{\partial}{\partial r}$$
$$= \frac{(2mr - a^2 \sin^2\theta)^{1/2}}{\rho} \frac{\partial}{\partial r}, \tag{3.8a}$$

$$u^\phi = \left(\frac{1 - 2mr}{\rho^2}\right)^{-1/2} \frac{a \sin\theta}{\Delta^{1/2}} L_\phi^{-1} \frac{\partial}{\partial\phi} = \frac{a}{\Delta} \frac{\partial}{\partial\phi}. \tag{3.8b}$$

These are the radial and azimuthal components of Carter's equations of motion.

There are four constants that must be determined if one desires to solve Carter's equations. They are M, the mass of the particle, the energy $E = M$, since the particle is released from rest, L_z, the angular momentum of the particle about the symmetry axis, and $\mathscr{Q}$, which can be found from the other three, since $p_\theta = 0$ (Ref. 2). To find L_z, one need only to evaluate "near" infinity, since it is a conserved quantity. At asymptotic infinity, the dynamics are Newtonian for small rotational velocities. One has $L_z = M v_\phi r_{\text{effective}}$. The value of $r_{\text{effective}}$ is the lever arm length $\rho \sin\theta$ found in (3.6). The rotational velocity is determined from Eq. (A2). One is not interested in Wigner rotated velocities in this case. The azimuthal velocity of interest is the velocity that the freely falling frame is observed to have in the stationary frames "near" infinity. No matter how much the tetrad of the freely falling observer rotates, his azimuthal velocity will be the same as measured in the stationary frames

$$v_\phi = a \sin\theta/\rho, \qquad \text{therefore } L_z = Ma \sin^2\theta. \tag{3.9}$$

Notice that L_z is not equal to zero for a particle that is released from rest at infinity even though $v_\phi \to 0$. The fourth constant is

$$\mathscr{Q} = M^2 a^2 \cos^2\theta \sin^2\theta. \tag{3.10}$$

Carter's equations are[2]

$$\frac{d\theta}{d\lambda} = 0, \tag{3.11a}$$

$$\frac{dt}{d\lambda} = \frac{M(r^2 + a^2)}{\Delta}, \tag{3.11b}$$

$$\frac{dr}{d\lambda} = \frac{M(2mr - a^2 \sin^2\theta)^{1/2}}{\rho}, \tag{3.11c}$$

$$\frac{d\phi}{d\lambda} = \frac{Ma}{\Delta}. \tag{3.11d}$$

Equations (3.11c) and (3.11d) reproduce (3.8a) and (3.8b) if one defines the affine parameter as $d\tau = M \, d\lambda$.

This treatment is well defined as long as the boost velocity is less than the speed of light. From Eq. (A4), $v = c$ at the stationary limit as expected.

IV. THE ZERO ANGULAR MOMENTUM FRAMES

From Carter's equations of motion and azimuthal symmetry, one knows that the freely falling frames spiral down a surface of revolution toward the event horizon. At each point along this trajectory, there is a boost from the Riemann normal coordinates of the freely falling frame to the zero angular momentum frame (0-L frame) at that point (Fig. 2). These boosts will correspond to a relative velocity between the frames that has both a radial and azimuthal component. In general, the boost can only realize the 0-L frame to zeroth order (at the point), since there will be an acceleration and precession between the two frames in addition to a relative velocity. As long as there is a unique boost for every point along the world-line of the freely falling observer, there is no ambiguity in neglecting the first-order effects of acceleration and precession.

As in the Schwarzschild example, one wants to find the image of the inertial tetrad e_α in the frame with a hypersurface orthogonal four-velocity (with legs e'_α). Thus, one has the special relativistic transformations for two combined orthogonal boosts[3]

$$e'_0 = \gamma e_0 + \gamma_1 v_1 e_1 + \gamma_2 v_2 e_2, \tag{4.1a}$$

$$e'_1 = \gamma v_1 e_0 + \gamma_1 e_1 + \gamma v_1 v_2 e_2, \tag{4.1b}$$

$$e'_2 = \gamma v_2 e_0 + \gamma v_1 v_2 e_1 + \gamma_2 e_2, \tag{4.1c}$$

$$e'_3 = e_3, \tag{4.1d}$$

from which one can deduce as in the discussion following (2.3) that the clocks of the inertial observer appear time dilated in the 0-L frame, with time dilation factor (3.5), $\gamma^{-1} = \gamma_1^{-1}\gamma_2^{-1}$.

The inverse equations are

$$e_0 = \gamma e'_0 - \gamma_1 v_1 e'_1 - \gamma_2 v_2 e'_2, \tag{4.2a}$$

$$e_1 = -\gamma v_1 e'_0 + \gamma_1 e'_1 + \gamma v_1 v_2 e'_2, \tag{4.2b}$$

$$e_2 = -\gamma v_2 e'_0 + \gamma v_1 v_2 e'_1 + \gamma_2 e'_2, \tag{4.2c}$$

$$e_3 = e'_3. \tag{4.2d}$$

In analogy to the discussion following (2.4), if one associates the first boost with the "radial" direction and the second boost with the azimuthal direction, then a meter stick placed along the azimuthal direction in the inertial frame appears longer in the 0-L frame by a factor γ_ϕ. Similarly, a meter stick aligned with the radial direction in the freely falling frame appears longer by a factor γ_r in the 0-L frame.

The legs of the image of the freely falling tetrad as they appear in the 0-L frame are labeled as E_α. Pictorially, this image is given in Fig. 3.

The boost velocities can be deduced using the results of Sec. III and the coordinate transformation from the stationary frames to the 0-L frames [3]

$$\begin{bmatrix} e'_0 \\ e'_\phi \end{bmatrix} = \begin{bmatrix} |g_{tt} - w^2 g_{\phi\phi}|^{-1/2} & w|g_{tt} - w^2 g_{\phi\phi}|^{-1/2} \\ 0 & g_{\phi\phi}^{-1/2} \end{bmatrix} \begin{bmatrix} \tilde{e}_0 \\ \tilde{e}_\phi \end{bmatrix}, \tag{4.3}$$

$$e'_r = (\Delta^{1/2}/\rho)\tilde{e}_r \quad e'_\theta = \rho^{-1/2}\tilde{e}_\theta,$$

where e'_0 is the four-velocity of the 0-L frame, $\tilde{e}_\alpha$ are the basis vectors of the stationary frame, and $w \equiv -g_{\phi t}/g_{\phi\phi}$. The corresponding transformation for the coframes is given at the beginning of Appendix B. Using the natural isomorphism between vectors and covectors, $\langle \hat{\omega}_\phi, \hat{e}_\phi \rangle = 1$, the orthonormal covectors of the stationary frame at infinity are

$$\hat{\omega}_\phi = r \sin\theta\, d\phi, \quad \hat{\omega}_0 = dt. \tag{4.4}$$

These relations can be used to rewrite the transformation of the covectors in Appendix B in terms of orthonormal coframes

$$\begin{bmatrix} w'_0 \\ w'_\phi \end{bmatrix} = \begin{bmatrix} |g_{tt} - w^2 g_{\phi\phi}|^{1/2} & 0 \\ -w g_{\phi\phi}^{1/2} & g_{\phi\phi}^{1/2} \end{bmatrix} \begin{bmatrix} \hat{\omega}_0 \\ \hat{\omega}_\phi (r\sin\theta)^{-1} \end{bmatrix}. \tag{4.5}$$

Equation (4.5) can be used to find the velocity between the freely falling frames and the 0-L frames. One can make the approximation that w'_ϕ and the other basis forms are not exact differentials, but infinitesimal displacements

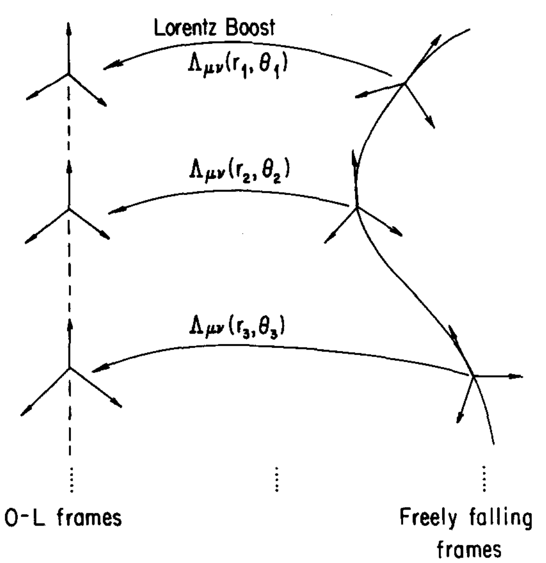


FIG. 2. There exists a Lorentz boost from the freely falling observer (right) to a 0-L frame (left) at each point along his world-line.

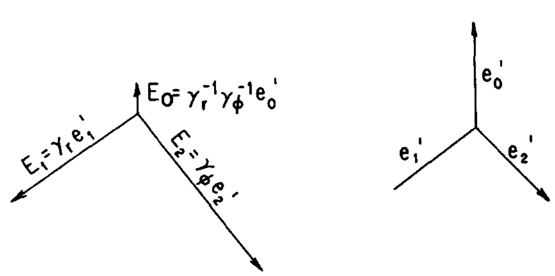


FIG. 3. The image of the freely falling tetrad (left) in the 0-L basis (right), with one dimension suppressed.

w'_ϕ ~ an infinitesimal *azimuthal* displacement in the 0-L frames,

$d\phi$ ~ an infinitesimal *angular* displacement in the stationary frames at asymptotic infinity.

The magnitudes of these displacements are obtained by taking the inner product of the forms with themselves

$$\langle w'_\phi, w'_\phi\rangle = w^2 g_{\phi\phi}\langle \hat{w}_0,\hat{w}_0\rangle - 2wg_{\phi\phi}\langle \hat{w}_0,\hat{w}\phi\rangle + g_{\phi\phi}(r\sin\theta)^{-1}\langle \hat{w}_\phi,\hat{w}_\phi\rangle.$$

The metric coefficients in the orthonormal basis at asymptotic infinity are $g^{\hat{0}\hat{0}} = -1$, $g^{\hat{\phi}\hat{\phi}} = 1$, and $g^{\hat{\phi}\hat{0}} = 0$. Thus,

$$|w'_\phi|^2 = w^2 g_{\phi\phi}\,dt^2 + g_{\phi\phi}\,d\phi^2. \tag{4.6}$$

Define the magnitude of the azimuthal displacement in the 0-L frame as $\Delta x_\phi \approx w_\phi$. Then, rewriting (4.6) in approximate form

$$(\Delta x_\phi)^2 = -w^2 g_{\phi\phi}(\Delta t)^2 + g_{\phi\phi}(\Delta\phi)^2, \tag{4.7}$$

$$\left(\frac{\Delta x_\phi}{\Delta t}\right)^2 = -w^2 g_{\phi\phi} + g_{\phi\phi}\left(\frac{\Delta\phi}{\Delta t}\right)^2. \tag{4.8}$$

The quantity $v'^2_\phi = g_{\phi\phi}(\Delta\phi/\Delta t)^2$ can be recognized as the azimuthal velocity between the stationary frames and the freely falling frames. This is a pure velocity with no Wigner rotation as in (A2) and (3.9). Consequently, the azimuthal velocity [in the sense of (A2)] between the freely falling frame and the 0-L frame at a point of space-time is

$$v^2_\phi = -w^2 g_{\phi\phi} + v'^2_\phi. \tag{4.9}$$

This equation makes sense, since if $v_\phi = 0$, then $v'_\phi = wg^{1/2}_{\phi\phi}$, which is the velocity of the 0-L frames as viewed from asymptotic infinity.

Equations (3.9) and (3.2) can be inserted into (4.9) to give

$$v^2_\phi = \frac{(1+2mr/\rho^2)(r^2+a^2-2mr)}{\rho^2(r^2+a^2)+2mra^2\sin^2\theta}\,a^2\sin^2\theta. \tag{4.10}$$

Remembering that $v_\phi = \tilde{v}_\phi(1-\tilde{v}^2_r)^{1/2}$, Eq. (A2) implies

$$\gamma^2_\phi = [(r^2+a^2)+(2mra^2/\rho^2)\sin^2\theta]/\rho^2. \tag{4.11}$$

Equivalently,

$$\gamma^2_\phi\rho^2\sin^2\theta = g_{\phi\phi}. \tag{4.12}$$

The lever arm length $\rho\sin\theta$ of (3.6) appears in Eq. (4.12) as well. The reason for this is that the relative motion between the 0-L frames and the stationary frames is entirely in the azimuthal direction (both are defined at fixed values of r and θ). An azimuthal boost is perpendicular to the lever arm. Thus, this length would not be Lorentz contracted under a transformation from one frame to another.

If one is to understand Eq. (4.12), then knowledge of the lever arm is essential. This is because Lorentz transformations are made between vectors with dimensions of length. But, the 0-L frames are defined with angular variables (a curvilinear basis). To give the dimensions of length to an angular displacement usually involves the introduction of a lever arm distance in the form e'_1 = (the effective lever arm distance) e'_ϕ, where e'_ϕ is a unit vector. Thus, one can define a rectangular basis locally as

$$e'_1 = e'_r,\quad e'_2 = \rho\sin\theta\, e'_\phi,\quad e'_3 = \rho e'_\theta, \tag{4.13}$$

and $e'_0 = u$, the four-velocity of the 0-L frame.

Combining Eqs. (4.1), (4.2), and (4.13), the image of the freely falling tetrad in the 0-L frame is (Fig. 3)

$$E_0 = \gamma_r^{-1}\gamma_\phi^{-1}e'_0, \tag{4.14a}$$

$$E_r = \gamma_r e'_r, \tag{4.14b}$$

$$E_\phi = \gamma_\phi\rho\sin\theta e'_\phi, \tag{4.14c}$$

$$E_\theta = \rho e'_\theta. \tag{4.14d}$$

These can be evaluated explicitly using the fact that the 0-L frames are defined at a fixed value of r. Thus, γ_r will be the same as the value found for the stationary frames (3.5). Using this and (4.12), the completely integrable distribution spans the tangent space to a three-dimensional submanifold with a metric, as in (2.6), $g^{(3)}_{ik} = E_i\cdot E_k$

$$g^{(3)}_{rr} = \gamma^{(2)}_r = \rho^2/\Delta, \tag{4.15a}$$

$$g^{(3)}_{\phi\phi} = \gamma^2_\phi\rho^2\sin^2\theta = (r^2+a^2)\sin^2\theta + (2mra^2/\rho^2)\sin^4\theta, \tag{4.15b}$$

$$g^{(3)}_{\theta\theta} = \rho^2. \tag{4.15c}$$

Employing (3.5), the time dilation factor is

$$\gamma^{-1} = \left[\frac{r^2+a^2-2mr}{(r^2+a^2)+(2mra^2/\rho^2)\sin^2\theta}\right]^{1/2}. \tag{4.16}$$

Not coincidentally, the length of the vector $\tilde{e}_0 + w\tilde{e}_\phi$ is γ^{-1} as well. From (4.3) the four-velocity of the 0-L frame is

$$e'_0 = \frac{\tilde{e}_0 + w\tilde{e}_\phi}{|\tilde{e}_0 + w\tilde{e}_\phi|}. \tag{4.17}$$

Thus, one can conclude that γ^{-1} is the time dilation associated with the proper time that flows along e'_0 in space-time orthogonal to the three-dimensional spacelike manifold. If one writes this normal vector as a unit vector in Boyer–Lindquist coordinates

$$n^\alpha = N^{-1}(1, -N^m),$$

then

$$N^\phi = -w,\quad N_\phi = g_{\phi t},\quad \text{and } N_\phi N^s - N^2 = g_{tt}. \tag{4.18}$$

The indices have been raised and lowered with the three-metric. Finally, one applies the Gauss–Codacci relations (2.7) in conjunction with the results (4.15) and (4.18) to get

$$g^{(4)}_{\mu\nu} = \begin{bmatrix} g_{tt} & g_{\phi t} \\ g_{\phi t} & g^{(3)}_{ik}\end{bmatrix} = \text{Kerr metric}.$$

Inspection of (4.16) shows that the boost velocity approaches c at the event horizon. Consequently, in order to study the space-time inside of the horizon, one must treat it as a separate space that asymptotically approaches the external space at the horizon.

V. BETWEEN THE TWO HORIZONS

There are two event horizons, at $r_\pm = m \pm (m^2-a^2)^{1/2}$. The region in between the horizons, $\{r/r_+ \geqslant r \geqslant r_-\}$, is considerably different from the external region that was discussed in the preceding sections. The coordinate r is a measure of time in this region and the four-velocity of the 0-L frames becomes a spacelike basis vector.

The result is that the region is dynamical. The spacelike part of this region, in the 0-L splitting of space-time, is actually an infinite cylinder with a cross section which is a two-dimensional spheroid of revolution (azimuthal symmetry). At r_+ and r_- this cross-sectional surface becomes lightlike. The two-dimensional spheroidal cross section varies with proper time r. The coordinate corresponding to $e_0' = u$ is the measure of length along the cylinder. This region is sometimes called an Einstein–Rosen bridge. The space is invariant under displacement along the u direction. Thus, frames at different points on the bridge are best labeled by values of the proper time r. The main difficulty in applying the formalism of this paper is that the boundaries of this space are both null. The result is that the boosts that are described in this section are very near the speed of light. An analysis along the lines of the previous sections will not be very enlightening, but is included for the sake of completeness.

The best that can be done in the framework of this paper is to consider the region $\{r/r_+ - \epsilon \geqslant r \geqslant r_- + \epsilon\}$ for ϵ an arbitrarily small positive number. The procedure for the most part is the same as for the external space-time. The hypersurface $r = r_+ - \epsilon$ becomes identified with asymptotic infinity. The 0-L frames can still be used, since e_0' remains hypersurface orthogonal, except now to a hypersurface of signature (2,1). The Gauss–Codacci relations still apply to such a splitting of space-time.

In the external space-time, the freely falling frames were released from rest. However, this is a very special case. Freely falling frames also exist with an initial velocity. This is the case of interest in the region between the horizons. In order to find this velocity, first note that the boost velocity from the freely falling frames to the 0-L frames approached c in the radial direction at $r = r_+$ (4.16). But, v_ϕ was less than c at the horizon. So, there is reason to choose $\tilde{v}_\phi$ (as defined in Appendix A) continuously across the horizon. In analogy to the external space-time, choose the boost from the freely falling frames to the 0-L frames to have no radial component $\tilde{v}_u = 0$ at infinity ($r_+ - \epsilon$). This defines what will be called the preferred freely falling frames. The boost from the "static" frames at infinity $r_+ - \epsilon$ that realize the preferred freely falling frames, is given by a velocity in the u direction

$$v_0^2 = 1 - \frac{2(m^2 - a^2)^{1/2}\epsilon}{(r_+^2 + a^2) + (2mr_+ a^2/\rho_+^2)\sin^2\theta}. \tag{5.1}$$

This is approximately c, since the static frames (those that move with the local geometry) at the actual infinity r_+ are the outgoing principal null congruences. Thus, an observer at a constant value of the u coordinate (i.e., any frame with a boost component $\tilde{v}_u = 0$ with respect to a 0-L frame) would be moving at the speed of light with respect to the local geometry as in (5.1). Of course, at $r_+ - \epsilon$, these statements are only approximate.

This equation has interesting implications for the preferred set of freely falling frames. One can define this Lorentz frame in terms of the coordinates of the static frames at infinity that "move" with the local geometry. Consequently, as before, the static frames at asymptotic infinity will play the role of the lab frame. The legs of the tetrad of the preferred freely falling observers, labeled E'_α can be written in terms of the clocks and meter sticks of the static frame at infinity, labeled $\hat{e}_\alpha$, as in (5.2)

$$E'_r = \gamma_0 \hat{e}_r, \text{ and } E'_u = \gamma_0^{-1}\hat{e}_u. \tag{5.2}$$

This is still an inertial frame; one can think of the transformation law as saying that the preferred freely falling observers carry a coordinate grid that is not cubical, but rectangular (actually the boundaries of extremely thin rectangular solids). The "natural" rate of time flow and distance measurement is given by the basis that is at rest with respect to the local geometry $\hat{e}_\alpha$. This basis is an orthonormal basis. But, these observers stay on the null surface of the horizon forever. Therefore, they are useless for the purposes of this paper. The preferred freely falling observers fall the entire length of the bridge. Hence, they can be assimilated into the formalism of this article. However, the preferred freely falling basis is necessarily an orthogonal basis in contrast to the static frames. Since it is an inertial frame, these observers transport this same rectangular grid as they fall from r_+ to r_-.

The boost that realizes the 0-L frame in the preferred freely falling basis is characterized by

$$\tilde{v}_u^2 = 1 - [2(m^2 - a^2)^{1/2}\epsilon/(2mr - a^2 - r^2)] \times g_{\phi\phi}(r,\theta)/g_{\phi\phi}(r_+,\theta), \tag{5.3}$$

and by the continuity of $\tilde{v}_\phi$, γ_ϕ is given by (4.11).

Equations (4.1) and (4.2) give the image of the preferred freely falling tetrad in the 0-L frames, in analogy to (4.14). In addition, one must employ (4.13) and (5.2)

$$E_0 = \gamma_u \gamma_0^{-1} e_0', \tag{5.4a}$$

$$E_r = \gamma_u^{-1}\gamma_\phi^{-1}\gamma_0 e_r', \tag{5.4b}$$

$$E_\phi = \gamma_\phi \rho \sin\theta e_\phi', \tag{5.4c}$$

$$E_\theta = \rho e_\theta'. \tag{5.4d}$$

Then, using the explicit expressions (5.1), (5.3), (5.4), and (4.11), one can find the metric of the three-manifold of signature (2,1) that is generated by the completely integrable distribution

$$g_{rr}^{(3)} = E_r \cdot E_r = -\gamma_u^{-2}\gamma_\phi^{-2}\gamma_0^2 = \rho^2/\Delta, \tag{5.5a}$$

$$g_{\phi\phi}^{(3)} = E_\phi \cdot E_\phi = \gamma_\phi^2 \rho^2 \sin^2\theta = (4.15\text{b}), \tag{5.5b}$$

$$g_{\theta\theta}^{(3)} = E_\theta \cdot E_\theta = \rho^2. \tag{5.5c}$$

The length of the normal vector to this three-manifold is

$$\gamma_u \gamma_0^{-1} = \left[\frac{r^2 + a^2 - 2mr}{(r^2 + a^2) + 2mra^2/\rho^2)\sin^2\theta}\right]^{1/2}. \tag{5.6}$$

Implementation of the Gauss–Codacci relations (2.7) proceeds exactly as in Sec. IV. Application of the explicit values in (5.6) and (5.5) results in the Kerr metric in this region.

VI. INSIDE OF THE INNER HORIZON

The region $\{r/0 \leqslant r \leqslant r_-\}$ is very similar to the external space-time except that the situation is somewhat inverted. Both regions are bounded by a null hypersurface. However, they are distinct in character. Inside of the horizon, the null hypersurface plays the role of asymptotic infinity. The similarity is that the boost from the preferred freely falling frames to the 0-L frames is characterized by a radial velocity that approaches c near the horizon, as is the case in the external space-time.

Inside of r_-, r is again a spacelike coordinate and e_0' is timelike. In analogy to Sec. V, one defines a set of preferred freely falling frames so that $\tilde{v}_\phi$ is continuous across the horizon as in (5.1) and $\tilde{v}_r \to 1$ at the horizon as in the external space-time. The boost from the static frames at infinity that realizes the preferred freely falling frames is in the radial direction and is given by the four-velocity

$$u = |\cos\theta|^{-1}(1, \sin\theta, 0, 0). \tag{6.1}$$

These frames are special because as they fall freely, they come to rest on the hypersurface at $r = 0$. They simply rotate around the surface, azimuthally, at the end of their fall. As in (5.2), the legs of the tetrad of the preferred freely falling frame E'_α can be written in terms of the natural clocks and meter sticks of the static frames at infinity $\hat{e}_\alpha$

$$E'_r = \gamma_0^{-1}\hat{e}_r, \text{ and } E'_0 = \gamma_0\hat{e}_0. \tag{6.2}$$

The boost from the preferred freely falling frame to the 0-L frame is given by

$$\tilde{v}_r^2 = (r^2\sin^2\theta + 2mr\cos^2\theta)/\rho^2, \tag{6.3}$$

and γ_ϕ is given by (4.11), since $\tilde{v}_\phi$ is continuous.

As in (5.4), the image of the preferred freely falling frame in the 0-L frame is

$$E_0 = \gamma_r^{-1}\gamma_\phi^{-1}\gamma_0 e_0', \tag{6.4a}$$

$$E_r = \gamma_r\gamma_0^{-1}e_r', \tag{6.4b}$$

$$E_\phi = \gamma_\phi\rho\sin\theta e_\phi', \tag{6.4c}$$

$$E_\theta = \rho e_\theta'. \tag{6.4d}$$

The metric of the three-manifold that is spanned by the completely integrable distribution is obtained by using (4.11), (6.1), (6.3), and (6.4)

$$g_{rr}^{(3)} = E_r \cdot E_r = \gamma_r^2\gamma_0^{-2} = \rho^2/\Delta, \tag{6.5a}$$

$$g_{\phi\phi}^{(3)} = E_\phi \cdot E_\phi = \gamma_\phi^2\rho^2\sin^2\theta = (4.15\text{b}), \tag{6.5b}$$

$$g_{\theta\theta}^{(3)} = E_\theta \cdot E_\theta = \rho^2. \tag{6.5c}$$

The length of the normal vector or time dilation is

$$\gamma_r^{-1}\gamma_\phi^{-1}\gamma_0 = (4.16). \tag{6.6}$$

Using (6.5) and (6.6), the calculations in the Gauss–Codacci relations proceed in exactly the same way as in Sec. IV to regain the Kerr metric inside of the inner horizon.

VII. THE KERR SOLUTION FOR $a^2 > m^2$

The region inside of the inner event horizon is the same type of region, in the context of this paper, as the entire space-time for the case $a^2 > m^2$. The difference between the two regions is that asymptotic infinity is no longer a null hypersurface. In this situation, it is the limit as $r \to \infty$ that is characterized by the stationary frames as in Sec. IV. The preferred freely falling frames have the same initial boost velocity from those frames that move with the local geometry at infinity as in (6.1). Also, the boost from the preferred freely falling frames to the 0-L frames is the same as in the last section, namely (6.3) and (4.11). The only difference is that r need not be less than r_-, but can extend to infinity.

In this section, it will be shown that these boost velocities in the previous paragraph are consistent with Carter's equations of motion. In Sec. III, the boost from the freely falling frames to the stationary frames was used to deduce Carter's equations of motion. Then, in Sec. IV, a coordinate transformation was implemented in deriving the boost to the 0-L frames. This time, Carter's equations will be used to give the boost from the preferred freely falling frames to the stationary frames and then a coordinate transformation to the 0-L frames will result in a boost that is consistent with (6.3) and (4.11).

To start with, one must determine the four constants of motion that characterize the preferred freely falling frames. The mass M will be the same as in Sec. III. The energy picks up a Lorentz contraction factor from (6.1)

$$E = (\cos\theta)^{-1}M. \tag{7.1a}$$

Then, as in (3.9)

$$L_z = Ea\sin^2\theta = Ma\sin^2\theta/\cos\theta, \tag{7.1b}$$

and since $p_\theta = 0$,

$$\mathscr{Q} = 0. \tag{7.1c}$$

Carter's equations of free fall for a particle with the given initial conditions are

$$\frac{dr}{d\lambda} = \frac{M(r^2\sin^2\theta + 2mr\cos^2\theta)^{1/2}}{\rho\cos\theta}, \tag{7.2a}$$

$$\frac{d\phi}{d\lambda} = \frac{Ma}{\Delta\cos\theta}, \tag{7.2b}$$

$$\frac{dt}{d\lambda} = \frac{M(r^2+a^2)}{\Delta\cos\theta}, \tag{7.2c}$$

$$\frac{d\theta}{d\lambda} = 0. \tag{7.2d}$$

These expressions can be related to a boost from the stationary frames by starting with (3.4), which is transcribed as (7.3)

$$dl^2 = g_{rr}^{-1}g_{tt}^{-1}\rho^2\sin^2\theta\, d\phi^2 \equiv L_\phi^2\, d\phi^2. \tag{7.3}$$

Then, (6.5) says

$$g_{rr}^{-1} = \bar{\gamma}_r^{-1}\gamma_0, \tag{7.4a}$$

$$g_{tt}^{-1} = \bar{\gamma}_r\bar{\gamma}_\phi\gamma_0^{-1}. \tag{7.4b}$$

As in (3.7), a local coordinate basis is chosen

$$e_\phi = (L_\phi)^{-1}\frac{\partial}{\partial\phi} = (\bar{\gamma}_\phi\rho\sin\theta)^{-1}\frac{\partial}{\partial\phi}, \tag{7.5a}$$

$$e_r = \frac{\gamma_0}{\bar{\gamma}_r}\frac{\partial}{\partial r}. \tag{7.5b}$$

Equation (7.2) can be expressed in analogy to (3.8) as

$$u^r = \frac{(r^2\sin^2\theta + 2mr\cos^2\theta)^{1/2}}{\Delta^{1/2}\cos\theta}e_r, \tag{7.6a}$$

$$u^\phi = \frac{a\sin\theta}{(1-2mr/\rho^2)^{1/2}\Delta^{1/2}\cos\theta}e_\phi. \tag{7.6b}$$

Utilizing Eqs. (A3) and (7.4), the boost parameters corresponding to (7.6) are

$$\tilde{v}_r^2 = (r^2\sin^2\theta + 2mr\cos^2\theta)/\rho^2, \tag{7.7a}$$

$$\tilde{v}_\phi'^2 = a^2\sin^2\theta/\Delta. \tag{7.7b}$$

The analysis leading to (4.8) goes through in the same way except there is an extra time dilation factor from the initial velocity of the preferred freely falling frames

$$\left(\frac{\Delta x_\phi}{\gamma_0^{-1}\Delta t}\right)^2 = -w^2 g_{\phi\phi} + g_{\phi\phi}\left(\frac{\Delta\phi}{\gamma_0^{-1}\Delta t}\right)^2. \tag{7.8}$$

Equation (4.9) becomes

$$v_\phi^2 = -w^2 g_{\phi\phi}\cos^2\theta + v_\phi'^2. \tag{7.9}$$

Recall a fact from (A2) that as in (4.9), $v_\phi = \tilde{v}_\phi(1-\tilde{v}_r^2)^{1/2}$. Then, (7.9) can be used to give the azimuthal boost parameter from the preferred freely falling frame to the 0-L frame

$$\tilde{v}_\phi^2 = \frac{(1+2mr/\rho^2)}{(r^2+a^2)(2mra^2/\rho^2)\sin^2\theta}\,a^2\sin^2\theta. \tag{7.10}$$

Equation (7.10) is identical to (4.11) and since both the stationary frames and the 0-L frames are defined at fixed values of r, (7.7a) is equivalent to (6.3). One can conclude that Carter's equation's of motion are consistent with the local analysis involving the preferred freely falling frames.

VIII. THE TWO-SURFACE AT $r = 0$

The metric (3.2), restricted to $r = 0$, is

$$ds^2 = -dt^2 + \cos^2\theta\, dr^2 + a^2\cos^2\theta d\theta^2 + a^2\sin^2\theta\, d\phi^2. \tag{8.1}$$

The metric of the hypersurface at $r = 0$ is the same as (8.1) with $dr^2 = 0$. The set $r = 0$ is a three-dimensional cylinder with a constant time cross section that is a two-dimensional spacelike surface. The techniques that have been developed in the previous sections can be used to obtain a physical understanding of the metric coefficients in (8.1). The ultimate nature of the hypersurface at $r = 0$ depends on the extension of the metric through the ring singularity. It is commonly assumed that the spacelike surface at $r = 0$ is an arbitrary separation of the space-time that results from the choice of coordinate patches that are used to cover the manifold. The two-surface at $r = 0$ is treated, trivially, as an obstructive membrane that has no physical significance and can simply be extended through by choosing negative r values.[1] It is the aim of this section to show that this surface is a fundamentally important structure in the space-time. Furthermore, there exist natural extensions that endow the two-surface with a topology that is not homotopically trivial, as is commonly assumed.

It is possible to gain insight into the global properties of this two-surface by using special relativity in the manner that has been elaborated on in the previous sections. First of all, using (4.11) and (6.3) or (7.7), the four-velocity that realizes the 0-L frames in the preferred freely falling basis at $r = 0$ is azimuthal

$$u = |\cos\theta|^{-1}(1,0,\sin\theta,0). \tag{8.2}$$

Furthermore, at $r = 0$, $e_0' = u$ (the four-velocity of the 0-L frames) reduces to the vector field $\partial/\partial t$. Consequently, (8.2) represents a boost to the stationary frames at $r = 0$. One should recall the remark made in Sec. VI that the preferred freely falling frames are special as they come to rest on this two-surface and simply rotate around azimuthally.

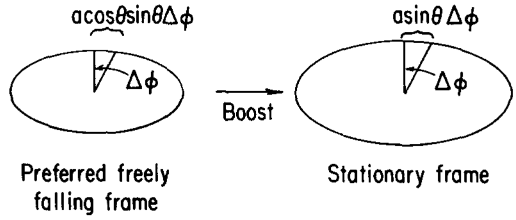


FIG. 4. The induced mapping of azimuthal rings as viewed by the preferred freely falling observers to their image in the stationary frames.

Now, some amazing results occur that allow one to piece together the entire two-surface. To see how this works, first notice that the four-velocity of the 0-L frames is hypersurface orthogonal everywhere on the surface. Then, by (8.2) and azimuthal symmetry, there is an induced map of azimuthal rings as measured by the locally inertial observers to azimuthal rings as measured by stationary observers (Fig. 4). The next thing to notice is that in Appendix C, (C7), it is shown that the stationary frames do not precess with respect to the preferred freely falling basis at $r = 0$. The only first-order effect is a radial acceleration that can be attributed to the rotational boost between the frames about the symmetry axis, (C8). As a result of Appendix B, (B3), it is also known that the surface has zero intrinsic or Gaussian curvature. Consequently, there is nothing to prevent one from placing one ring on top of another to get the induced mapping of global two-surfaces as pictured in Fig. 5.

In the stationary frame, one has a spherical shell corresponding to a radius of length a. However, this surface is not geometrically a sphere, since the meridians turn out to be twisted. This point will be expounded on in the next few paragraphs. This shell appears to the stationary observers to be rotating at the maximum velocity that is permitted by special relativity, as $v_\phi = c$ at the equator

$$v_\phi = wa\sin\theta = c\sin\theta.$$

The flat geometry of the inertial frames implies that the geometry of the two-surface, as viewed in this basis, is what one would expect by embedding the surface in a Euclidean three-space. The surface is two spheres of radius $\frac{1}{2}a$ that are tangent at a point on the symmetry axis. The surface is parametrized by

$$r(\theta,\phi) = a|\cos\theta|(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta). \tag{8.3}$$

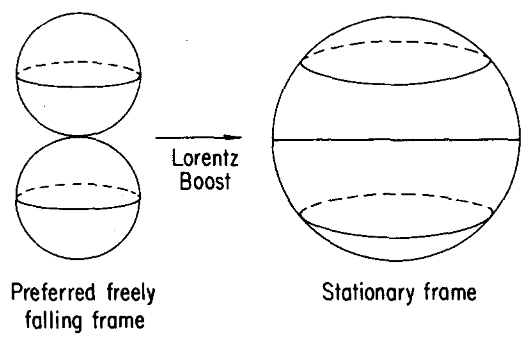


FIG. 5. The global map of the two-surface as viewed by locally inertial observers to its image in the stationary frames.

The metric coefficients are

$$\tilde{g}_{\phi\phi} = \left\langle \frac{\partial}{\partial\phi}, \frac{\partial}{\partial\phi} \right\rangle = a^2 \cos^2\theta \sin^2\theta, \tag{8.4a}$$

$$\tilde{g}_{\theta\theta} = \left\langle \frac{\partial}{\partial\theta}, \frac{\partial}{\partial\theta} \right\rangle = a^2. \tag{8.4b}$$

In order to explain the metric coefficients in (8.1), one must look at the image of this two-surface in the stationary frame. The $g_{\phi\phi}$ coefficient is an immediate consequence of the preceeding discussion of Lorentz contraction in this article applied to (8.2) and (8.4a). The $g_{\theta\theta}$ coefficient in (8.1) can be understood by analyzing the headlight effect on null rays on these surfaces by means of the Doppler formula

$$\cos\psi' = (\cos\psi - \beta)/(1 - \beta\cos\psi), \tag{8.5}$$

where ψ' and ψ are the angles that the null rays make with the direction of relative motion (azimuthal direction) in the rocket (stationary) and lab frames (freely falling), respectively. First, look at the Doppler formula for a photon along the direction $\partial/\partial\theta$ (meridian) in the two-surface that is embedded in the Euclidean three-space of the inertial frames. Thus, $\cos\psi = 0$. Since $\beta = |\sin\theta|$, $\cos\psi' = -|\sin\theta|$. Therefore, locally, one has the situation that is pictured in Fig. 6. Piecing the local geometries together, globally, the transformed null trajectories appear on the spherical shell as in Fig. 7.

The projection of the transformed null meridians of the inertial surface on the latitudinal direction of the "spherical" stationary surface can be found by looking at local coordinate patches in the stationary frame (Fig. 8). The arc length of the meridian of the surface in Euclidean space was found in (8.4b) to be $a\,d\theta$. Since there is no relative motion between the frames in the θ direction, the image of the null meridian in the stationary frame will also have length $a\,d\theta$. From Fig. 8, the component of this arc length along a meridian of the stationary surface is $a\cos\theta\,d\theta$. In conclusion, $g_{\theta\theta}$ as measured in the stationary frames is $a^2\cos^2\theta$ in agreement with (8.1).

The length of the bottom leg of the triangle in Fig. 8 is $a\sin\theta\,d\theta$. The metric coefficient $g_{\phi\phi}$ implies that $d\theta = \pm d\phi$. Thus, the vector field that generates the transformed meridians is $\partial/\partial\theta + \partial/\partial\phi$ or $-\partial/\partial\theta + \partial/\partial\phi$. By construction, these twisted meridians are the natural meridians of the surface (geodesics).

It is amazing that special relativity can be used to describe the geometry of the two-surface. The proof that it does

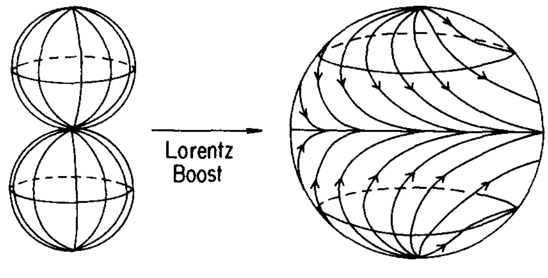


FIG. 7. The global headlight effect in the stationary frames.

is given by an explicit solution for the geodesic vector fields in the Kerr geometry restricted to $r = 0$. In Appendix B, (B4), it is shown that the obstruction to a trivial normal bundle (normal Euler characteristic) is zero. The hypersurface at $r = 0$ is a trivial bundle. Thus, the spatial projections of the null geodesics in the hypersurface are the same as the geodesics of the two-surface. If X is a geodesic field, then $\nabla_X X = 0$. To solve this equation, expand X in the orthonormal 0-L basis.

$$X = a_\phi(\sin\theta)^{-1}\frac{\partial}{\partial\phi} + a_\theta(\cos\theta)^{-1}\frac{\partial}{\partial\theta}. \tag{8.6}$$

All of the covariant derivatives involving 0-L basis vector fields can be found from the connection coefficients of Appendix B, (B2). Applying azimuthal symmetry and time translation invariance, the geodesic equation reduces to two coupled equations

$$a_\theta(a\cos\theta)^{-1}\frac{\partial}{\partial\theta}(a_\theta) + a_\phi^2(-a\sin\theta)^{-1} = 0, \tag{8.7a}$$

$$a_\theta a_\phi(a\sin\theta)^{-1} + a_\theta(a\cos\theta)^{-1}\frac{\partial}{\partial\theta}(a_\phi) = 0. \tag{8.7b}$$

There are four null vector fields that are solutions in the hypersurface

$$\frac{\partial}{\partial\theta} + \frac{\partial}{\partial\phi} + \frac{\partial}{\partial t}, \tag{8.8a}$$

$$\frac{\partial}{\partial\theta} - \frac{\partial}{\partial\phi} + \frac{\partial}{\partial t}, \tag{8.8b}$$

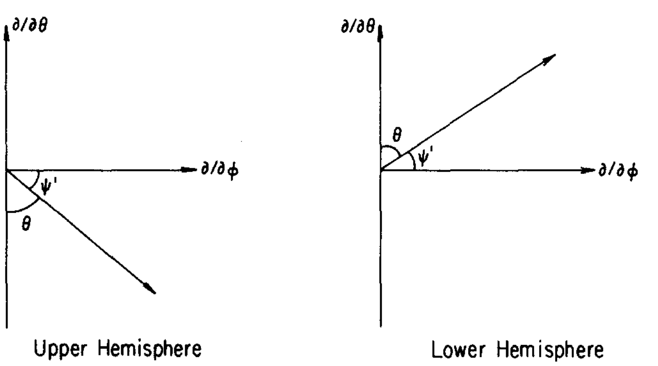


FIG. 6. The headlight effect as observed in a local coordinate patch in the stationary frame applied to the null meridians of the inertial two-surface.

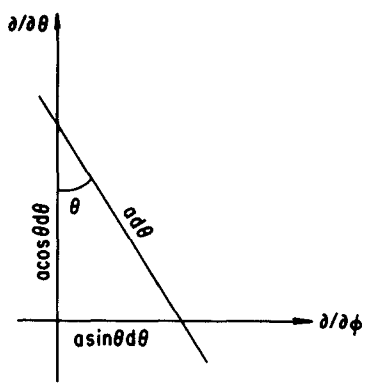


FIG. 8. The projection of a transformed null meridian on the basis vectors $\partial/\partial\theta$ and $\partial/\partial\phi$ in a local coordinate patch of the stationary frame.

$$\frac{\partial}{\partial\theta}+\frac{\partial}{\partial\phi}-\frac{\partial}{\partial t}, \quad (8.8c)$$

$$\frac{\partial}{\partial\theta}-\frac{\partial}{\partial\phi}-\frac{\partial}{\partial t}. \quad (8.8d)$$

It should be noted that these are geodesics in the hypersurface, but not in the bigger four-dimensional manifold. The solutions (8.8) are the same as those found from the special relativistic treatment.

With these models of the hypersurface, one can make some remarks on the global topology of the two-surface at $r=0$, by exploring the global geodesic structure. The twisted spherical shell that the stationary observers detect is topologically equivalent to the real projective space RP^2.

A geodesic beginning at the north pole of the twisted spherical shell and ending at the equator is given by (8.8a). Call such a geodesic with an initial azimuth, $\phi=0$, $X(\widetilde{\theta})$

$$X(\widetilde{\theta})=|\cos\widetilde{\theta}\,|\,e'_\theta+\sin\widetilde{\theta}e'_\phi \quad 0\leqslant\widetilde{\theta}\leqslant\tfrac{1}{2}\pi. \quad (8.9)$$

This geodesic reaches the equator at the point $\theta=\frac{1}{2}\pi$, $\phi=\frac{1}{2}\pi$.

Define another geodesic that starts at $\theta=\frac{1}{2}\pi$, $\phi=\frac{1}{2}\pi$, that is like solution (8.8d).

$$Y(\widetilde{\theta})=-|\cos\widetilde{\theta}\,|\,e'_\theta+|\sin\widetilde{\theta}\,|\,e'_\phi, \quad 0\leqslant\widetilde{\theta}\leqslant\tfrac{1}{2}\pi. \quad (8.10)$$

At $\widetilde{\theta}=\frac{1}{2}\pi$, $Y(\widetilde{\theta})$ reaches the north pole. Thus, both $X(\widetilde{\theta})$ and $Y(\widetilde{\theta})$ are geodesics that connect the same point on the equator to the north pole (Fig. 9). Conjugate points are a parameter distance $\frac{1}{2}\pi$ from each other. This is exactly what happens under the natural map of $S^2\rightarrow RP^2$.

Alternatively, if one continues a geodesic smoothly through the singularity at $\theta=\frac{1}{2}\pi$ to the lower hemisphere and then back to the north pole, one gets a global closed continuous geodesic which is always being dragged in the $+\boldsymbol{\phi}$ direction by the headlight effect. Remembering that the coordinate θ is bounded from above by π, there is a unique geodesic path $Z(\widetilde{\theta})$ that is drawn in Fig. 10

$$Z(\widetilde{\theta})=\begin{cases}X(\widetilde{\theta}) & 0\leqslant\widetilde{\theta}\leqslant\pi,\\ Y(\widetilde{\theta}) & \pi\leqslant\widetilde{\theta}\leqslant 2\pi.\end{cases} \quad (8.11)$$

If one propagates along $Z(\widetilde{\theta})$ from the north pole in the $+\boldsymbol{\theta}$ direction for a parameter distance $\frac{1}{2}\pi$, then one has a geodesic starting at $\phi=0$, $\theta=0$ that ends at the point $\phi=\frac{1}{2}\pi$, $\theta=\frac{1}{2}\pi$ on the equator. If one propagates along $Z(\widetilde{\theta})$ from the north pole along the $-\boldsymbol{\theta}$ direction for a parameter distance $\frac{1}{2}\pi$, then one has a geodesic starting at $\phi=\pi$, $\theta=0$ that ends at the point $\phi=\frac{1}{2}\pi$, $\theta=\frac{1}{2}\pi$ on the equator, as before. The global geodesic structure identifies antipodal points on the equator. Consequently, the equator is topologically RP^1. A sphere with antipodal points identified on the equator is topologically equivalent to a disk with antipodal points on the boundary identified. This is a standard model of RP^2. In terms of the previous discussion of conjugate points, the cut locus is RP^1, which is characteristic of RP^2.

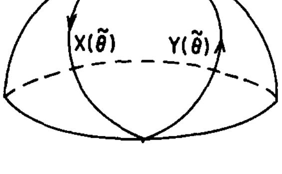


FIG. 9. Two null geodesics $X(\widetilde{\theta})$ and $Y(\widetilde{\theta})$ connect the same point on the equator to the north pole on the surface as viewed in the stationary frames.

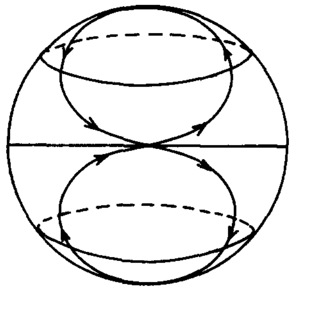


FIG. 10. A continuous closed geodesic path on the two-surface in the stationary frames.

Clearly, this is not the complete story. In Appendix B, it was found that $K=0$. The only compact two-surfaces that can be embedded with $K=0$ are the Klein bottle and the torus (RP^2 is compact). Thus, knowledge of the extension through the ring singularity is necessary for the determination of the global structure of the surface. If one assumes that the surface is compact, then the most natural choice is the Klein bottle. The reason for this is the elementary topological fact that the Klein bottle is $RP^2\ \#\ RP^2$, where # denotes the connected sum.[7] The connected sum is obtained by cutting out a region that is homotopically equivalent to a disk out of each RP^2 and then gluing them together along the boundary of the cut.

The most general Kerr–Schild solution depends on a complex function.[8] For the Kerr metric, this function has a quadratic branch point at $r=0$, $\theta=\frac{1}{2}\pi$. The Klein bottle description is consistent with this. Thus, the manifold can be thought of as two sheets, with each RP^2 lying on a different sheet. Since $r=0$, $\theta=\frac{1}{2}\pi$ is a quadratic branch point, one must pass through the ring singularity twice in the same direction to get back to the original sheet. This is precisely what happens with a Klein bottle. As one travels around the "outside," after one revolution, one ends up on the "inside." It takes two revolutions to get back to the outside.

The most natural choice for the second RP^2 is obtained by the discrete isometry of the Kerr metric $\phi\rightarrow-\phi$ and $t\rightarrow-t$. One can employ (8.8a) and (8.8d) to get the global closed geodesics on this RP^2, in analogy to (8.11). This RP^2 is distinct from the first. Even though RP^2 is not orientable, RP^1 is. As stated before, a model for RP^2 is a disk with RP^1 as a boundary. Thus, under the discrete isometry, the boundaries of the two disk models of RP^2 have opposite orientations.

This choice of the extension has the property of interchanging the roles of the ingoing and outgoing principal null congruences on each sheet. Explicitly, under discrete isometry

$$k_{\rm in}\rightarrow -k_{\rm out} \quad\text{and}\quad k_{\rm out}\rightarrow -k_{\rm in}. \quad (8.12)$$

This interchange of principal null congruences was a motivation for the maximal analytic extension of Boyer and Lindquist.[1] Clearly, a different RP^2 could be chosen for the connected sum, but this one is particularly natural and the metric is analytic.

Once these properties of the two-surface have been obtained, it is of interest to know if this is just a random slicing of space-time or is it physically significant? To investigate this point, the role of the principal null congruences on this surface will be explored. One can use the Doppler shift discussion that explained the metric coefficient $g_{\theta\theta}$ to interpret g_{rr} in (8.1). The surface as viewed by the locally inertial observers is treated as if it is embedded in a Euclidean three-space. Thus, $\tilde{g}_{rr} = 1$. As with the null meridians, one can shoot light rays normal to this surface. The headlight effect is clearly exactly the same as before, so $g_{rr} = \cos^2\theta$. The transformed null rays as viewed in the stationary frame are the outgoing principle null geodesics. At $r = 0$, $k_{\text{out}} = (1, 1, a^{-1}, 0)$ and $k_{\text{in}} = (1, -1, a^{-1}, 0)$ (Ref. 1). Thus, by inverting the argument, the spatial part of the outgoing principal null congruence is the normal vector field of the inertial two-surface. Similarly, as in (8.12), the ingoing principal null congruence corresponds to the normal vector field of the two-surface on the other sheet of space-time.

APPENDIX A: COMBINING ORTHOGONAL LORENTZ BOOSTS

Define two Lorentz boosts as $L(\tilde{v}_1)$ and $L(\tilde{v}_2)$, where $\tilde{v}_1 = \tanh^{-1}(v_1^*)$, $\tilde{v}_2 = \tanh^{-1}(v_2^*)$. The quantities v_1^* and v_2^* are the velocity parameters of $L(\tilde{v}_1)$ and $L(\tilde{v}_2)$, respectively. The corresponding Lorentz contraction factors are $\gamma_1 = (1 - \tilde{v}_1^2)^{-1/2}$ and $\gamma_2 = (1 - \tilde{v}_2^2)^{-1/2}$. For a single boost, $\tilde{v}_1$ would be the three-velocity of the boost. In the case of combined boosts, the component of the three-velocity in the direction that is labeled by 1 depends on both $\tilde{v}_1$ and $\tilde{v}_2$. This will be defined explicitly.

In this paper, $\tilde{v}_1$ corresponds to a radial boost and $\tilde{v}_2$ denotes an azimuthal boost. Since boosts do not commute, the order is important. In the Lorentz algebra, the product of two boosts is another boost composed with a Wigner rotation

$$L(\tilde{v}_1)L(\tilde{v}_2) = L(\theta\hat{n})L(\mathbf{v}_3).$$

The explicit results are derived for orthogonal boosts. $L(\theta\hat{n})$ is a rotation about the unit vector $\hat{n}$, which is normal to the plane spanned by $\tilde{v}_1$ and $\tilde{v}_2$. $L(\mathbf{v}_3)$ is a boost that is characterized by the four-velocity

$$u = (1 - \tilde{v}_1^2)^{-1/2}(1 - \tilde{v}_2^2)^{-1/2}(1,\tilde{v}_1,\tilde{v}_2(1 - \tilde{v}_1^2)^{1/2},0). \tag{A1}$$

The three-velocity is given by $v^i = u^i/u^0$.

$$\mathbf{v}_3 = (\tilde{v}_1,\tilde{v}_2(1 - \tilde{v}_1^2)^{1/2},0). \tag{A2}$$

Since the Wigner rotation is in the plane spanned by $\tilde{v}_1$ and $\tilde{v}_2$, sometimes one might be interested in $\mathbf{v}_3$ in a Wigner rotated basis. The combined effects of the rotation and the boost are defined by the velocities below:

$$u = (1 - \tilde{v}_1^2)^{-1/2}(1 - \tilde{v}_2^2)^{-1/2}(1,\tilde{v}_1(1 - \tilde{v}_2^2)^{1/2},\tilde{v}_2,0), \tag{A3}$$

$$\mathbf{v} = (\tilde{v}_1(1 - \tilde{v}_2^2)^{1/2},\tilde{v}_2,0). \tag{A4}$$

APPENDIX B: THE STRUCTURE EQUATIONS AT $r = 0$

The structure equations will be solved in the orthonormal coframe of the 0-L observers

$$w^t = \left[\frac{r^2 + a^2 - 2mr}{(r^2 + a^2) + (2mra^2/\rho^2)\sin^2\theta}\right]^{1/2} dt,$$

$$w^r = \left[\frac{r^2 + a^2\cos^2\theta}{r^2 + a^2 - 2mr}\right]^{1/2} dr,$$

$$w^\theta = (r^2 + a^2\cos^2\theta)^{1/2}\, d\theta,$$

$$w^\phi = \sin\theta\,[(r^2 + a^2) + (2mra^2/\rho^2)\sin^2\theta\,]^{1/2} \times\left[d\phi - \frac{2mra}{\rho^2(r^2 + a^2) + 2mra^2\sin^2\theta}dt\right].$$

The structure equations are $dw^\alpha = w^\alpha_\gamma \wedge w^\gamma$. The solutions to these equations are very complicated at an arbitrary point of space-time. But, when these solutions are restricted to the hypersurface $r = 0$, $w^r = 0$, one gets particularly simple expressions

$$\begin{aligned} &w^r{}_\theta = 0, \quad w^t{}_\theta = 0, \quad w^\phi{}_t = 0, \quad w^\phi{}_\theta = d\phi, \\ &w^t{}_r = (m/a^2\cos^3\theta)dt - (m\sin^2\theta/a\cos^3\theta)d\phi, \\ &w^\phi{}_r = (m\sin\theta/a^2\cos^3\theta)dt + (m\sin^3\theta/a\cos^3\theta)d\phi. \end{aligned} \tag{B1}$$

The first results that can be found from these solutions are the connection coefficients in the 0-L frame

$$w_{\mu\nu} = \Gamma_{\mu\nu\alpha}w^\alpha. \tag{B2}$$

Secondly, one can derive the second-order structure equations from (B1), which give the topological invariants of the two-surface at $r = 0$

$$R^\mu{}_\nu = dw^\mu{}_\nu + w^\mu{}_\alpha \wedge w^\alpha{}_\nu,$$

where $R^\mu{}_\nu$ is the curvature two-form, which is equivalent to the Riemann tensor. By direct calculation, when their values are restricted to $r = 0$, $dw^\phi{}_\theta = 0$ and so does $w^\phi{}_\alpha \wedge w^\alpha{}_\theta$. Thus, at $r = 0$, $R^\phi{}_\theta = 0$. Recall the relation from Cartan's method of moving frames

$$dw^\phi{}_\theta = -K\,dA + R^\phi{}_\theta, \tag{B3}$$

where K is the Gaussian curvature of the spacelike two-surface at $r = 0$ and $dA = w^\theta \wedge w^\phi$ is the element of surface area. One can conclude that $K = 0$ on this surface.

The obstruction to the existence of a trivial normal bundle (the normal Euler characteristic), in curved space, is the integral over the surface of the two-form $dw^r{}_t$

$$dw^r{}_t = -(m/a^3\cos^5\theta)(2 + \sin^2\theta)w^\theta \wedge w^\phi, \tag{B4}$$

$$\int_\Omega dw^r{}_t = 0. \tag{B5}$$

APPENDIX C: THE ACCELERATION AND PRECESSION BETWEEN FRAMES

The 0-L frame is not an inertial frame, since by the results of Appendix B, the connection coefficients do not vanish. However, at any one instant, there exists an inertial frame that is momentarily at rest with respect to the 0-L frame at any point of space-time. This frame will not remain at rest, since in general there is a precession and acceleration between the two frames. This inertial frame is not one of the preferred freely falling frames, but can be obtained from them by a Lorentz boost. Consequently, in order to find the precession and acceleration of the preferred freely falling frames with respect to the 0-L frames, one proceeds in two steps. First, find the acceleration and precession of the local inertial frame that is momentarily at rest with respect to the 0-L frame at a point. Second, Lorentz boost this result, as a second-rank tensor, into the preferred freely falling frame at the point.

The acceleration and precession of the first step can be determined from the connection coefficients defined in (B2), as follows.[2] Instantaneously, one can choose both frames to have the legs of their tetrads labeled e'_α. The four-velocity of the 0-L observers as measured in their own frames is $u = (1, 0, 0, 0)$. This can be used to calculate the components of the antisymmetric tensor $\Omega^{\alpha\beta}$ that is defined by the equation

$$\nabla_u e'_\alpha = \Omega^{\ \beta}_{\alpha} e'_\beta, \tag{C1}$$

or, since $u = (1, 0, 0, 0)$, $\nabla_0 e'_\alpha = \Omega^{\ \beta}_{\alpha} e'_\beta$. One can define an expression for the antisymmetric tensor involving accelerations and precessions

$$\Omega^{\mu\nu} = a^\mu u^\nu - u^\mu a^\nu + u_\alpha w_\beta \epsilon^{\alpha\beta\mu\nu}.$$

This allows one to identify components as $\Omega_r^{\ 0} = \Omega_0^{\ r} = \hat{g}_{00}\Omega^{0r} = a_r$, the radial acceleration between frames [note that the metric expressed in terms of the 0-L basis $\hat{g}^{\alpha\beta}$ is diag$(-1, 1, 1, 1)$]. Similarly, $\Omega_\phi^{\ r} = -\Omega_r^{\ \phi} = w$, the precession of the 0-L frame. From (C1) and (B2), one concludes that $\Gamma^{\ \beta}_{\alpha 0} = \Omega^{\ \beta}_{\alpha}$. Then, from the expression for $\Omega^{\mu\nu}$

$$\Gamma_{jk0} = -w^i \epsilon_{0ijk} \quad \text{and} \quad \Gamma_{j00} = a_j.$$

Thus, at $r = 0$

$$\mathbf{a} = -(m/a \cos^3\theta) e'_r, \tag{C2}$$

$$\mathbf{w} = -(m \sin\theta / a^2 \cos^3\theta) e'_\theta. \tag{C3}$$

The second part is more complicated. The preferred freely falling frame is not at rest with respect to the 0-L frame, even in a momentary sense. There is a boost to the inertial frame that is momentarily at rest. This boost has been found in (8.2) to be azimuthal with boost velocity $c \sin\theta$. Let $\Lambda^{\alpha\beta}$ be the generator of this boost. Then, rewriting (C1), one can get the acceleration and precession with respect to the preferred freely falling tetrad e_α

$$\nabla_u e_\gamma = \nabla_u \Lambda^{\ \alpha}_{\gamma} e'_\alpha = \Lambda^{\ \alpha}_{\gamma} \nabla_u e'_\alpha$$
$$= \Lambda^{\ \alpha}_{\gamma} \Omega^{\ \beta}_{\alpha} e'_\beta = \Lambda^{\ \alpha}_{\gamma} \Omega^{\ \beta}_{\alpha} (\Lambda^{-1})^{\delta}_{\beta} e_\delta.$$

The relative acceleration between the two frames is given by

$$\nabla_u e_0 = \Lambda^{\ \alpha}_{0} \Omega^{\ \beta}_{\alpha} (\Lambda^{-1})^{\delta}_{\beta} e_\delta, \tag{C4}$$

and the precession is given by

$$\nabla_u e_i = \Lambda^{\ \alpha}_{i} \Omega^{\ \beta}_{\alpha} (\Lambda^{-1})^{\delta}_{\beta} e_\delta. \tag{C5}$$

Explicitly, one has

$$\nabla_u e_\phi = (a_r \beta\gamma + \gamma w) e_r, \tag{C6a}$$

$$\nabla_u e_r = -(\beta\gamma a_r + \gamma w) e_\phi + (\gamma a_r + \beta\gamma w) e_0, \tag{C6b}$$

$$\nabla_u e_0 = (\gamma a_r + w\beta\gamma) e_r. \tag{C6c}$$

Since the boost goes in the opposite way of (8.2), $\beta = -c \sin\theta$ and $\gamma = (\cos\theta)^{-1}$. Thus

$$\nabla_u e_\phi = 0, \tag{C7a}$$

$$\nabla_u e_r = -(m/a^2 \cos^2\theta) e_0, \tag{C7b}$$

$$\nabla_u e_0 = -(m/a^2 \cos^2\theta) e_r. \tag{C7c}$$

There is no precession between the 0-L frame and the preferred freely falling frame at any point on the two-surface at $r = 0$. The acceleration vector is

$$\mathbf{a} = -(m/a^2 \cos^2\theta) e_r. \tag{C8}$$

[1] R. H. Boyer and R. W. Lindquist, J. Math. Phys. **8**, 265 (1967).

[2] C. Misner, K. Thorne, and J. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[3] A. Lightman, W. Press, R. Price, and S. Teukolsky, *Problem Book in Relativity and Gravitation* (Princeton U. P., Princeton, NJ, 1979).

[4] A. Einstein, *The Meaning of Relativity* (Princeton U. P., Princeton, NJ, 1950).

[5] Ya. B. Zel'dovich and I. D. Novikov, *Relativistic Astrophysics* (Univ. Chicago, Chicago, 1978), Vol. I.

[6] L. D. Landau and E. M. Lifschitz, *The Classical Theory of Fields* (Addison-Wesley, Reading, MA, 1971).

[7] W. S. Massey, *Algebraic Topology: An Introduction* (Harcourt, New York, 1967).

[8] R. P. Kerr and A. Schild, "Some algebraically degenerate solutions of Einstein's gravitational field equations," presented at the American Mathematical Society Symposium, New York, 1964.

Separation of variables for the Rarita–Schwinger equation on all type D vacuum backgrounds

N. Kamran[a)]
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

(Received 22 August 1984; accepted for publication 18 January 1985)

We present a separable master equation governing Rarita–Schwinger spin-$\frac{3}{2}$ fields, valid in the whole class of type D vacuum backgrounds.

I. INTRODUCTION

One of the most striking uses of separation of variables as a tool for studying the properties of a space-time was made by Carter,[1] who, while seeking to determine the geodesics of the Kerr solution in order to show that the analytic extension he had constructed for it was maximal, was the first to realize that the Hamilton–Jacobi equation for the non-null geodesics and the Klein–Gordon equation for a massive spin-zero field are solvable by separation of variables when the Kerr metric is expressed in advanced null coordinates. Teukolsky[2] subsequently extended Carter's result to higher spin wave equations when he studied gravitational perturbations and test electromagnetic and neutrino fields on the Kerr background. Assuming an algebraically special Petrov type D vacuum background (the class of which we shall denote by $\mathscr{D}_0$ in what follows) he obtained a set of decoupled equations for the components of maximal spin weight $s = \pm 2, \pm 1, \pm \frac{1}{2}$ of these fields. He then showed that these equations, when expressed for the Kerr solution using Boyer–Lindquist[3] coordinates and a Kinnersley[4] tetrad, can be unified into a single separable master equation. The case of spin-$\frac{3}{2}$ fields was subsequently studied by Güven,[5] who considered linear perturbations of a Kerr black hole by a Rarita–Schwinger field in his proof of the no-hair conjecture for the uncharged black holes of supergravity theory.[5–6] He succeeded under the assumption of a type D vacuum background to obtain decoupled equations for two scalars which completely determine those solutions of the Rarita–Schwinger equations which cannot be generated from a vacuum by supersymmetry transformations. He then showed that these equations are separable when expressed for the Kerr solution using Boyer–Lindquist coordinates and a Kinnersley tetrad.

In view of the validity of Teukolsky's and Güven's decoupling procedure for every solution in $\mathscr{D}_0$, one is lead to conjecture that the separation of Teukolsky's and Güven's equations can be performed for the whole class $\mathscr{D}_0$. This conjecture has been proved for Teukolsky's equations by Kamran and McLenaghan,[8] who, thanks to the exhaustive integration and the single expression recently obtained by Debever, Kamran, and McLenaghan[9–12] for the general solution of the type D vacuum and electrovac field equations with cosmological constant, combined Teukolsky's equations expressed for the $\mathscr{D}_0$ solutions as presented in Ref. 11 (hereafter referred to as DKM) into a single separable master equation valid for gravitational, electromagnetic, and neutrino field perturbations. They also showed[13–14] that Chandrasekhar's separation procedure[15] for the Dirac equation in the Kerr solution could be extended to the *whole* class $\mathscr{D}_0$ in the massless case.

In the present paper, we are able to prove the above-mentioned conjecture for Güven's equations by showing that they can be combined into a single equation which is solvable by separation of variables in the whole class $\mathscr{D}_0$ of type D vacuum solutions. In the proof, we use the symmetric tetrad and coordinates of DKM for the $\mathscr{D}_0$ solutions, thus showing that a Kinnersley tetrad is not necessary for separation—a fact first pointed out for Teukolsky's equations in the Kerr case by Carter and McLenaghan.[16]

II. SEPARATION OF VARIABLES FOR THE RARITA-SCHWINGER EQUATION IN THE CLASS $\mathscr{D}_0$

Using Debever's complex vectorial formalism,[17] the Rarita–Schwinger equation on a curved background space-time may be written as

$$\theta^2 \wedge H_1 - \theta^3 \wedge H_2 = 0, \tag{2.1a}$$

$$\theta^1 \wedge H_2 - \theta^4 \wedge H_1 = 0, \tag{2.1b}$$

where the two forms H_1 and H_2 are defined in terms of the components of the Majorana spinor-valued one-form ψ representing the spin-$\frac{3}{2}$ field

$$\psi = \begin{pmatrix} F_1 \\ F_2 \\ \bar{F}_2 \\ -\bar{F}_1 \end{pmatrix}, \tag{2.1c}$$

by

$$H_1 = dF_1 + \sigma_2 \wedge F_1 + \sigma_3 \wedge F_2, \tag{2.2a}$$

$$H_2 = dF_2 - \sigma_2 \wedge F_2 - \sigma_1 \wedge F_1, \tag{2.2b}$$

and where in Eqs. (2.2a) and (2.2b), the one-forms σ_1, σ_2, and σ_3 are the complex vectorial components[18] of the matrix ω_{ab} of connection one-forms.

Güven shows in Ref. 5 that if the background space-time has a Petrov type D Weyl tensor and satisfies the Einstein vacuum field equations, there exists a null tetrad of basis one-forms and a supersymmetry gauge in which

$$H_1 = a_1 \Theta_1, \quad H_2 = b_2 \Theta_2, \tag{2.3a}$$

where a_1 and b_2 are scalars of Geroch, Held, and Penrose (GHP)[19] types $\{-3, 0\}$ and $\{3, 0\}$, respectively, and where, denoting by Ψ_2 the only nonvanishing GHP component of the Petrov type D Weyl tensor,

$$\Theta_1 := \Psi_2 \theta^2 \wedge \theta^4, \quad \Theta_2 := \Psi_2 \theta^1 \wedge \theta^3. \tag{2.3b}$$

The scalars a_1 and b_2, which are invariant under supersymmetry transformations and gauge invariant in the sense of

[a)] Present address: CRMA, Université de Montréal, Case Postale 6128, Succarsale "A," Montréal, P. Q. H3C 3J7, Canada.

 0022-2488/85/071740-03$02.50

Stewart and Walker,[20] satisfy the following euqations:

$$[(\text{Þ}'-\bar{\rho}')(\text{Þ}+2\rho)-(\eth'-\bar{\tau})(\eth+2\tau)-\Psi_2]a_1=0, \quad (2.4a)$$

$$[(\text{Þ}-\bar{\rho})(\text{Þ}'+2\rho')-(\eth-\bar{\tau}')(\eth'+2\tau')-\Psi_2]b_2=0. \quad (2.4b)$$

The components F_1 and F_2 of the Rarita–Schwinger field are completely determined by the scalars a_1 and b_2 through the relations (25)–(34) of Ref. 5, which indicate that a_1 and b_2 may be interpreted as scalar Debye potentials for the Rarita–Schwinger field, analogous to those introduced by Cohen and Kegeles[21] for neutrino, electromagnetic, and gravitational perturbations of space-times.

Güven then shows that Eqs. (2.4a) and (2.4b) are solvable by separation of variables when expressed in the Kerr solution using Boyer–Lindquist coordinates and a Kinnersley tetrad.

We now consider Eqs. (2.4a) and (2.4b) written in the single expression given in Theorems 1 and 2 of DKM for the general solution of Einstein's field equations for the class $\mathscr{D}_0$ of Petrov type D vacuum metrics, using the symmetric null tetrad given by Eqs. (3.37)–(3.39) of DKM. The separability result of Güven can be generalized through the following theorem, which is stated using the notations of DKM for the $\mathscr{D}_0$ solutions.

Theorem: Güven's equations (2.4a) and (2.4b) when expressed in the coordinates and symmetric null tetrad given in Theorems 1,2 and Eqs. (3.37)–(3.39) of DKM admit, for all solutions in the class $\mathscr{D}_0$ and R-separable,[22] a solution of the form

$$\Phi_{\pm 3/2}(u,v,w,x)=e^{i(ru+qv)}T(w,x)^{-1/2}Z(w,x)^{3/4}\times e^{-i3/4\mathscr{B}(w,x)}\Psi_{\pm 3/2}(w,x), \quad (2.5a)$$

where

$$\Phi_{3/2}:=b_2,\quad \Phi_{-3/2}:=a_1,$$

$$\Psi_{\pm 3/2}(w,x)=G_{\pm 3/2}(w)H_{\pm 3/2}(x), \quad (2.5b)$$

$$d\mathscr{B}=Z^{-1}(\epsilon_1 m'\,dw+\epsilon_2 p'\,dx), \quad (2.5c)$$

$$Z:=\epsilon_1 p(w)-\epsilon_2 m(x), \quad (2.5d)$$

and where r and q are arbitrary real constants.

The integrability condition for Eq. (2.5c) is equivalent[23] to the condition that the metric defined by DKM Eqs. (2.5) be of Petrov type D. We prove the above theorem by showing that $\Phi_{\pm 3/2}$ satisfies Güven's equations (2.4a) and (2.4b) if and only if $\Psi_{\pm 3/2}$ satisfies the equation

$$(\mathring{D}_{ws}\mathring{L}_{ws}-\mathring{D}_{xs}\mathring{L}_{xs}+F_s(w,x))\Psi_s(w,x)=0, \quad (2.6a)$$

where s is a discrete parameter taking the values $\frac{3}{2}$ and $-\frac{3}{2}$ and where

$$\mathring{L}_{ws}:=-fW\frac{\partial}{\partial w}+\frac{2}{1+f^2}\left[\frac{1}{2}\left(1+\frac{|s|}{s}\right)-\frac{1}{2}\left(1-\frac{|s|}{s}\right)f^2\right]W^{-1}(pir-\epsilon_2 iq)-|s|fW', \quad (2.6b)$$

$$\mathring{L}_{xs}:=iX\frac{\partial}{\partial x}+\frac{|s|}{s}X^{-1}(\epsilon_1 iq-mir)+i|s|X', \quad (2.6c)$$

$$\mathring{D}_{ws}:=W\frac{\partial}{\partial w}+\frac{2}{1+f^2}\left(\frac{1}{2}\left(1+\frac{|s|}{s}\right)f-\frac{1}{2}\left(1-\frac{|s|}{s}\right)f^{-1}\right)W^{-1}(pir-\epsilon_2 iq)+W'(1-|s|), \quad (2.6d)$$

$$\mathring{D}_{xs}:-iX\frac{\partial}{\partial x}+\frac{s}{|s|}X^{-1}(\epsilon_1 iq-mir)+iX'(|s|-1), \quad (2.6e)$$

$$\begin{aligned}F_s:=&-2\epsilon_1 Z^{-1}(-p'+im')(|s|/s)(pir-\epsilon_2 iq)+\tfrac{1}{2}f(W^2)'\epsilon_1 Z^{-1}(-p'+im')\\&+fW^2\epsilon_1 Z^{-2}[\epsilon_1 p'(-p'+im')+p''Z+\epsilon_1(-p'+im')^2]+2\epsilon_2 Z^{-1}(p'-im')(|s|/s)(\epsilon_1 iq-mir)\\&+\tfrac{1}{2}i(X^2)'\epsilon_2 Z^{-1}(p'-im')+X^2\epsilon_2 Z^{-2}[-i\epsilon_2 m'(p'-im')-m''Z-\epsilon_2(p'-im')^2]-2ZT^{-2}\Psi_2,\end{aligned} \quad (2.6f)$$

and where $W(w)$ and $X(x)$ are related to the DKM metric functions $U(w)$ and $V(x)$ by $W=U^{1/2}$ and $X=V^{1/2}$.

We shall prove that for all the solutions in $\mathscr{D}_0$ and for $s=-\frac{3}{2},\frac{3}{2}$, the function F_s defined by Eq. (2.6f) is of the form

$$F_s(w,x)=f_s(w)+g_s(x), \quad (2.7a)$$

in which case Eq. (2.6a) separates into the pair of *decoupled ordinary differential equations*

$$(\mathring{D}_{ws}\mathring{L}_{ws}+f_s(w)-\lambda_s)G_s(w)=0, \quad (2.7b)$$

$$(\mathring{D}_{xs}\mathring{L}_{xs}-g_s(x)-\lambda_s)H_s(x)=0, \quad (2.7c)$$

where λ_s is the separation constant. It should be noted that although the first-order operator $\mathring{D}_{w-3/2}$ is not defined for $f=0$, the second-order operator $\mathring{D}_{w-3/2}\mathring{L}_{w-3/2}$ reduces to a well-defined operator in this case.

We now proceed with the proof of our theorem. By Eqs. (2.5b), (2.5c), and (2.5d) we have for $s=\frac{3}{2}$

$$(\text{Þ}'+2\rho')\Phi_{3/2}=e^{i(ru+qv)}2^{-1/2}T^{1/2}Z^{1/4}e^{-i3/4B}\times[\mathring{L}_{w3/2}-fW\epsilon_1 Z^{-1}(-p'+im')]\Psi_{3/2}, \quad (2.8a)$$

$$(\eth'+2\tau')\Phi_{3/2}=e^{i(ru+qv)}2^{-1/2}T^{1/2}Z^{1/4}e^{-i3/4B}\times[\mathring{L}_{x3/2}-X\epsilon_2 Z^{-1}(p'-im')]\Psi_{3/2}, \quad (2.8b)$$

and for $s=-\frac{3}{2}$

$$(\text{Þ}+2\rho)\Phi_{-3/2}=-f^{-1}e^{i(ru+qv)}2^{-1}T^{1/2}Z^{1/4}e^{-i3/4B}\times[\mathring{L}_{w-3/2}-fW\epsilon_1 Z^{-1}(-p'+im')]\Psi_{-3/2}, \quad (2.8c)$$

$$(\eth+2\tau)\Phi_{-3/2}=-e^{i(ru+qv)}2^{-1/2}T^{1/2}Z^{1/4}e^{-i3/4B}\times[\mathring{L}_{x-3/2}-X\epsilon_2 Z^{-1}(p'-im')]\Psi_{-3/2}. \quad (2.8d)$$

TABLE I. Explicit form of $F_{3/2}$.

Case	$F_{3/2}$
A^*	$4\sqrt{2}(w+ix)ir+2a^2f_0(w^2+x^2)-a(f_1x+g_1w)$
B^0_-	$4ir(k+ix)+i(X^2)'(k-ix)^{-1}-2X^2(k-ix)^{-2}-\frac{1}{3}(k+ix)(k-ix)^{-2}(-6kf_2+3ig_1)$, where $X^2 := -f_2x^2+g_1x+k^2f_2$
B^0_+	$4ir(w+il)-f(W^2)'(w-il)^{-1}+2fW^2(w-il)^{-2}-\frac{1}{3}(w+il)(w-il)^{-2}(-6ilg_2+3f_1)$, where $fW^2 := -g_2w^2+f_1w+l^2g_2$
C^*	$-\sqrt{2}xr-x^{-1}(X^2)'+2x^{-2}X^2-(w+x^{-1})f_3$, where $X^2 := -f_0x^4+f_1x^3-f_2x^2+f_3x$
C^{00}	0

Moreover, if we define the functions

$$K_w := T2^{-1/2}Z^{-1/2}[W\epsilon_1 Z^{-1}(\tfrac{3}{4}p'-(i/4)m') - 2^{-1}WT^{-1}T_{,w}], \quad (2.9a)$$

$$K_x := T2^{-1/2}Z^{-1/2}[X\epsilon_2 Z^{-1}(-\tfrac{1}{4}p'+i\tfrac{3}{4}m') + 2^{-1}iXT^{-1}T_{,x}], \quad (2.9b)$$

we have

$$\text{Þ}-\bar{\rho} = 2^{-1/2}Z^{-1/2}TD_{w3/2}+K_w,$$
$$\eth-\bar{\tau}' = 2^{,1/2}Z^{-1/2}TD_{x3/2}+K_x, \quad (2.9c)$$

$$\text{Þ}'-\bar{\rho}' = -f(2^{-1/2}Z^{-1/2}TD_{w-3/2}+K_w),$$
$$\eth'-\bar{\tau} = -(2^{-1/2}TD_{x-3/2}+K_x), \quad (2.9d)$$

where the operators $D_{w\pm3/2}$ and $D_{x\pm3/2}$ are, respectively, obtained from the operators $\mathring{D}_{w\pm3/2}$ and $\mathring{D}_{x\pm3/2}$ by performing the substitutions $ir\to\partial/\partial u$ and $iq\to\partial/\partial v$.

Now, by Eqs. (2.8) and (2.3), we can combine Eqs. (2.4a) and (2.4b) into the following single equation valid for both the positive and negative values of s being considered; that is, for $s=-\frac{3}{2}$ and $\frac{3}{2}$:

$$\{[2^{-1/2}Z^{-1/2}T\mathring{D}_{w\pm3/2}+K_w][2^{-1/2}Z^{1/4}T^{1/2}e^{-i3/4\mathscr{B}}(\mathring{L}_{w\pm3/2}-fW\epsilon_1Z^{-1}(-p'+im'))]$$
$$-[2^{-1/2}Z^{-1/2}T\mathring{D}_{x\pm3/2}+K_x]\times[2^{-1/2}Z^{1/4}T^{1/2}e^{-i3/4\mathscr{B}}(\mathring{L}_{x\pm3/2}-X\epsilon_2Z^{-1}(p'-im'))]$$
$$-\Psi_2T^{-1/2}Z^{3/4}e^{-i3/4\mathscr{B}}\}\Psi_{\pm3/2}=0. \quad (2.10)$$

It is then straightforward to show that Eq. (2.10) is identical to Eq. (2.6a). To complete the proof of the theorem, one has to show that F_s as defined by Eq. (2.6f) splits into the sum of a function of w and a function of x for every solution in $\mathscr{D}_0$ and for $s=\frac{3}{2}, -\frac{3}{2}$. In order to do this, we calculate for each of the above values of s the expression of F_s using the expressions given after Theorem 2 of DKM for the metric functions p, m, W, X, and T in the class $\mathscr{D}_0$. The explicit form of F_s for the vacuum case of each element of the exhaustive set $\{A^*, B^0_-, B^0_+, C^*, C^{00}\}$ of cases for the class $\mathscr{D}_0$ enumerated after Theorem 2 of DKM is given in Table I. This is done for $s=\frac{3}{2}$ only since the substitution $s\to -s$ in F_s amounts simply to the replacement of r and q by $-r$ and $-q$ therein.

From Table I, we see that F_s is indeed of the form given by Eq. (2.7a). This completes the proof of our result.

ACKNOWLEDGMENT

It is a pleasure to thank Professor R. G. McLenaghan for his countinued interest and encouragement in this research and for financial support through his Natural Sciences and Engineering Research Council of Canada (NSERC) operating grant.

[1] B. Carter, Phys. Rev. **174**, 1559 (1968).
[2] S. Teukolsky, Astrophys. J. **185**, 635 (1973).
[3] R. H. Boyer and R. W. Lindquist, J. Math. Phys. **8**, 265 (1967).
[4] W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).
[5] R. Güven, Phys. Rev. D **22**, 2327 (1980).
[6] D. Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, Phys. Rev. D **13**, 3214 (1976).
[7] S. Deser and B. Zumino, Phys. Lett B **62**, 335 (1976).
[8] N. Kamran and R. G. McLenaghan, to appear in *The Ivor Robinson Festschift*, edited by A. Trautman and W. Rindler.
[9] R. Debever, N. Kamran, and R. G. McLenaghan, Phys. Lett. A **93**, 399 (1983).
[10] R. Debever, N. Kamran, and R. G. McLenaghan, Bull Cl. Sci. Acad. R. Belg. **68**, 592 (1982).
[11] R. Debever, N. Kamran, and R. G. McLenaghan, J. Math. Phys. **25**, 1955 (1984).
[12] R. Debever, N. Kamran, and R. G. McLenaghan, Ann. Inst. H. Poincaré **41**, 2 (1984).
[13] N. Kamran and R. G. McLenaghan, Lett. Math. Phys. **7**, 381 (1983).
[14] N. Kamran and R. G. McLenaghan, J. Math. Phys. **25**, 1019 (1984).
[15] S. Chandrasekhar, Proc. R. Soc. London Ser. A **349**, 571 (1976).
[16] B. Carter and R. G. McLenaghan, in *Proceedings of the Second Marcel Grossman Meeting on Recent Developments of General Relativity Trieste, 1979*, edited by R. Ruffini (North-Holland, Amsterdam, 1982).
[17] R. Debever, Cah. Phys. **168–169**, 303 (1964).
[18] Reference 17.
[19] R. Geroch, A. Held, and R. Penrose, J. Math. Phys. **14**, 874 (1973).
[20] J. Stewart and M. Walker, Proc. R. Soc. London Ser. A **341**, 49 (1974).
[21] J. M. Cohen and L. S. Kegeles, Phys. Lett. A **54**, 5 (1975).
[22] P. Moon and D. E. Spencer, *Field Theory Handbook* (Springer, New York, 1971), 2nd ed., p. 96.
[23] Reference 11, Eq. (4.2).

On a certain class of solutions of the Einstein–Maxwell equations

Jerzy Klemens Kowalczyński
Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland

(Received 17 July 1984; accepted for publication 14 December 1984)

This paper deals with exact solutions of the Einstein–Maxwell equations. The solutions belong to a metric form which is a special case of the Robinson–Trautman form. The chosen metric form admits positive, zero, and negative Gaussian curvature (labeled $\epsilon = 1, 0, -1$, respectively) of its two-dimensional angularlike part. Some general properties of the solutions are given for all the values of ϵ. All the explicit solutions for $\epsilon = \pm 1$ are presented. They are of the Petrov type II. The physical meaning of ϵ is discussed.

I. INTRODUCTION

The present paper deals with the exact solutions of the Einstein–Maxwell equations (without currents)

$$G_{\mu\nu} = -\lambda g_{\mu\nu} + 2(F_{\rho\mu}F_{\nu}{}^{\rho} + \tfrac{1}{4}g_{\mu\nu}F_{\rho\tau}F^{\rho\tau}), \tag{1.1a}$$

$$F_{[\mu\nu,\rho]} = 0, \quad F^{\mu\nu}{}_{;\nu} = 0. \tag{1.1b}$$

The solutions are limited to a special case of the Robinson–Trautman metric form,[1] namely

$$ds^2 = ds_0^2 + \Omega\, dq^2, \tag{1.2a}$$

where

$$ds_0^2 := 4p^2(1+\epsilon Y\overline{Y})^{-2} dY\, d\overline{Y} + 2dp\, dq - \epsilon\, dq^2, \tag{1.2b}$$

$$\Omega := p^{-2}(2mp + f). \tag{1.2c}$$

No additional limitations are assumed.

The meanings of the symbols are as follows: $G_{\mu\nu}$, $g_{\mu\nu}$, and $F_{\mu\nu}$ are the Einstein, metric, and electromagnetic field tensors, respectively; λ is the cosmological constant; Y is a complex coordinate; p and q are real coordinates; m is an arbitrary real constant; f is a disposable real function that belongs to class C^2 and depends on all four coordinates Y, $\overline{Y}$, p, and q; and ϵ is a discrete parameter equal to 1, 0, or -1. Here and below, every symbol with an overbar means a complex conjugate quantity of the given symbol.

The signature is $+++-$ and the Ricci tensor is assumed as $R_{\mu\nu} := R^{\rho}{}_{\mu\nu\rho}$. Units are chosen such that $c = \kappa = 1$, where c is the speed of light and κ is the gravitational constant.

The form ds_0^2 is the flat part of form (1.2a), while Ω determines the curvatures of space-times (1.2). If $f = 0$, then Eqs. (1.2) are the well-known explicit solutions of Eqs. (1.1a) for $\lambda = F_{\mu\nu} = 0$ (i.e., of $R_{\mu\nu} = 0$), namely, that of Schwarzschild for $\epsilon = 1$, that of Levi-Civita for $\epsilon = 0$, and that of Levi-Civita (anti-Schwarzschild) for $\epsilon = -1$. Our aim is to find the explicit expressions of f that fulfill Eqs. (1.1).

In Sec. II some general properties of the solutions are briefly reviewed. In Sec. III all the explicit solutions for $\epsilon \neq 0$, i.e., for $\epsilon = \pm 1$, and their Petrov classification are given. All the explicit solutions for $\epsilon = 0$ have been presented in another paper.[2] In Sec. IV the physical interpretation of parameter ϵ is given.

The present paper should be considered as a whole with Ref. 2 to encompass all the explicit solutions of Eqs. (1.1) with limitation (1.2).

II. SOME GENERAL PROPERTIES

When solving Eqs. (1.1) under condition (1.2) the known results[3] that include explicit expressions ready for integration[4] were used. After easy integration of a part of those expressions one finds that

$$f = -\tfrac{1}{3}\lambda p^4 + pA - B\overline{B}. \tag{2.1}$$

From Eqs. (1.2) and (2.1) it is easy to see that function A must be real and that conditions $A = A' +$ real const and $A = A'$ are equivalent for every real expression A' since m is arbitrary.

If we assume $B \neq 0$, then from Eqs. (1.1), (1.2), and (2.1) we obtain

$$A_{,p} = B_{,p} = B_{,\overline{Y}} = 0, \tag{2.2a}$$

$$A_{,Y\overline{Y}}\overline{B} - A_{,Y}\overline{B}_{,\overline{Y}} + 4B_{,q}\overline{B}^2(1+\epsilon Y\overline{Y})^{-2} = 0, \tag{2.2b}$$

$$A_{,Y}A_{,\overline{Y}}(1+\epsilon Y\overline{Y})^2 - 8A_{,q}B\overline{B} = 0, \tag{2.2c}$$

and

$$\begin{aligned} &F_{Yp} = 0, \quad F_{Y\overline{Y}} = (B - \overline{B})(1+\epsilon Y\overline{Y})^{-2}, \\ &F_{pq} = -\tfrac{1}{2}p^{-2}(B + \overline{B}), \\ &F_{Yq} = \tfrac{1}{2}p^{-1}B_{,Y} - \tfrac{1}{4}A_{,Y}\overline{B}^{-1}. \end{aligned} \tag{2.3}$$

The electromagnetic field (2.3) vector potential V_μ (that fulfills the general equations $F_{\mu\nu} = V_{\nu,\mu} - V_{\mu,\nu}$ and is determined with an accuracy up to the gradient of an arbitrary function) is as follows:

$$V_Y = \int \overline{B}(1+\epsilon Y\overline{Y})^{-2}\, d\overline{Y}, \quad V_{\overline{Y}} = \int B(1+\epsilon Y\overline{Y})^{-2}\, dY,$$

$$V_p = 0, \quad V_q = -\tfrac{1}{4}A(B^{-1} + \overline{B}^{-1}) + \tfrac{1}{2}p^{-1}(B + \overline{B}). \tag{2.4}$$

It is well known that the introduction of λ into the Robinson–Trautman metrics is very simple.[5] This is the cause of the absence of λ in Eqs. (2.2). Thus Eqs. (2.2) and (2.3) are special cases of the more general equations that have been given by Robinson *et al.*[6,7]

As regards $B = 0$, it is easy to prove, making use of Eqs. (1.1b) and the explicit expressions mentioned at the beginning of this section, the following implication: if $\epsilon \neq 0$, then $A =$ const (i.e., $A = 0$). Thus if we assume $\epsilon \neq 0$ and $B = 0$, then we have no electromagnetic field and Eqs. (1.2) and (2.1) are the well-known Schwarzschild($\epsilon = 1$)–Levi-Civita ($\epsilon = -1$)–de Sitter solutions. If $\epsilon = 0$, then electromagnetic fields may exist for $B = 0$, namely, we then have $F_{Y\overline{Y}} = F_{Yp} = F_{pq} = 0$, $F_{Yq} = C$, $A_{,Y} = A_{,\overline{Y}} = A_{,p} = 0$, and

 0022-2488/85/071743-09$02.50

$A_{,q} = 2C\overline{C}$, where C is an arbitrary complex function of one real variable q only.[2]

The curvature and electromagnetic field invariants and pseudoinvariants of space-times (1.2) and (2.1) are

$$R_{\mu\nu\rho\tau}R^{\mu\nu\rho\tau} = 4p^{-8}[(2mp+f)^2 + (2mp+f-2B\overline{B})^2 + (2mp+f-B\overline{B}+\lambda p^4)^2], \tag{2.5a}$$

$$R_{\mu\nu\rho\tau}R^{\rho\tau\xi\sigma}R_{\xi\sigma}{}^{\mu\nu} = 4p^{-12}[2(2mp+f)^3 + 2(2mp+f-2B\overline{B})^3 - (2mp+f-B\overline{B}+\lambda p^4)^3], \tag{2.5b}$$

$$R_{\mu\nu\rho\tau}\check{R}^{\mu\nu\rho\tau} = 0, \quad R_{\mu\nu\rho\tau}R^{\rho\tau\xi\sigma}\check{R}_{\xi\sigma}{}^{\mu\nu} = 0, \tag{2.6}$$

$$F_{\mu\nu}F^{\mu\nu} = -p^{-4}(B^2 + \overline{B}^2), \tag{2.7}$$

$$F_{\mu\nu}\check{F}^{\mu\nu} = ip^{-4}(B^2 - \overline{B}^2), \tag{2.8}$$

where $\check{R}_{\mu\nu\rho\tau}$ and $\check{F}_{\mu\nu}$ are dual with respect to the $R_{\mu\nu\rho\tau}$ and $F_{\mu\nu}$ tensors. We also have

$$(F_{\mu\nu} + i\check{F}_{\mu\nu})(F^{\mu\nu} + i\check{F}^{\mu\nu}) = -4p^{-4}B^2. \tag{2.9}$$

Equations (2.5)–(2.9) hold for both $B \neq 0$ and $B = 0$, and for all the values of ϵ. Equation (2.9) means that our electromagnetic fields are null if and only if $B = 0$. This and the remarks given above imply that the electromagnetic field is either zero or non-null for $\epsilon \neq 0$ and may be nonzero and null for $\epsilon = 0$.

Vector $k^\mu := \delta^\mu_p$ is important for the space-times (1.2) in view of the Goldberg–Sachs theorem (generalized) and the concepts of principal null directions of the Weyl and electromagnetic field tensors. In fact, the following *conclusions* (cf. Sec. 3 in Ref. 3) are obtained from Eqs. (1.2): k^μ is null, geodetic, shear-free, rotation-free, its expansion does not vanish, and if a given space-time (1.2) is not conformally flat, then k^μ is a double Debever–Penrose vector. We also conclude from our $F_{\mu\nu}$ that $k_{[\mu}F_{\nu]\rho}k^\rho = 0$, i.e., k^μ is here a principal null vector of the electromagnetic fields. Since $dq = k_\mu dx^\mu$ and vector k^μ is null for space-times both curved (1.2a) and flat (1.2b), Eqs. (1.2) represent the Kerr–Schild metric form.

III. EXPLICIT SOLUTIONS FOR $\epsilon \neq 0$ AND THEIR PETROV CLASSIFICATION

In this section we only consider the cases $\epsilon = \pm 1$, since case $\epsilon = 0$ has been presented in another paper.[2] If the presence of an electromagnetic field is assumed, it suffices to solve Eqs. (2.2) only (see Sec. II). The explicit solutions of Eqs. (2.2) are the following:

$$A = 0, \quad B = \phi(Y), \tag{3.1}$$

where ϕ is an arbitrary analytic function of Y only, and

$$A = 2a\ln[(1-Y\overline{Y})(1+Y)^{-1}(1+\overline{Y})^{-1}] + \tfrac{1}{2}a\ln q,$$
$$B = (aq)^{1/2}e^{ib}, \quad aq \geqslant 0, \quad \epsilon = -1, \tag{3.2}$$

$$A = 2a\ln(1+\epsilon Y\overline{Y}) + \tfrac{1}{2}a\ln q, \quad B = (aq)^{1/2}Y, \quad aq \geqslant 0, \tag{3.3}$$

where a and b are arbitrary real constants.[8]

The limitations $aq \geqslant 0$ [the case $a = 0$ can be treated as an effect of the limiting transition $B \to 0$, which is admitted by Eqs. (2.2) and (2.3), see Eq. (A1) in Appendix A] result from Eq. (2.2c), which gives $A_{,q} \geqslant 0$, and can be reversed by the transformation $q = -q'$. The fact that the values of q (for every given $a \neq 0$) are of one sign agrees with an invariant property of solutions (3.2) and (3.3). Indeed, it results from Eqs. (2.1) and (2.5) that for these solutions the strongest (scalar) singularities occur for $q = 0$, and also for $p = 0$ and in the cases $\epsilon = -1$ for $Y\overline{Y} = 1$. Thus space-times (3.2) and (3.3) cannot be extended beyond the boundaries $p = 0$ or $q = 0$, and in the cases $\epsilon = -1$ beyond boundaries $Y\overline{Y} = 1$. Thus, each one of the solutions (3.2) and (3.3) represents more than one space-time for a given value of $\epsilon (\neq 0)$.

Solution (3.2) exists only for $\epsilon = -1$ and it does not have a counterpart in the Schwarzschild branch $\epsilon = 1$. This means that solution (3.2) is specific for the tachyon (see Sec. IV). Such a fact was probably unknown for the Robinson–Trautman-type solutions hitherto.[9]

Solutions (3.1)–(3.3), together with Eq. (2.1), represent *all* the exact solutions of Eqs. (1.1) under limitation (1.2) for $\epsilon \neq 0$. A proof of this theorem is given in Appendices A–E.

Solution (3.1) has been given by many authors.[3,6,10] Solutions (3.2) and (3.3) seem to be new.[11]

When determining the Petrov types we used Lemma 1 from Ref. 3, since the premise of that lemma holds for the metric form (1.2). The following conclusions have been drawn: solution (3.1) is of the Petrov type [2, 1, 1] (Penrose's notation) if and only if (iff for short) $\phi_{,Y} \neq 0$, it is of type [2, 2] iff $\phi_{,Y} = 0$ and at the same time $\phi \neq 0$ or $m \neq 0$, and it is conformally flat iff $\phi = m = 0$ (cf. p. 369 in Ref. 3); solutions (3.2) and (3.3) are of type [2, 1, 1] iff $a \neq 0$, they are of type [2, 2] iff $a = 0$ and $m \neq 0$, and they are conformally flat iff $a = m = 0$.

IV. ON THE PHYSICAL MEANING OF PARAMETER ϵ

Parameter ϵ, equal to $+1$ or 0 or -1, is geometrically the sign of the Gaussian curvature of every two-dimensional surface defined by the conditions $p = \text{const} \neq 0$ and $q = \text{const}$.[1,12]

Physically ϵ is of course a kinematic parameter and is commonly related to a speed of the singular source of the field that produces the curvature of a given space-time. Cases $\epsilon = 1$, since they belong to the Schwarzschild branch, have commonly and naturally been related to such sources moving slower than light (bradyons). Cases $\epsilon = 0$ and $\epsilon = -1$ have been related to such sources moving with the speeds equal to (luxons) and greater than (tachyons) the speed of light, respectively. These interpretations for $\epsilon = 0$, -1 have been given by many authors without[12,13] as well as with justifications. A short justification treating cases $\epsilon = 0$ as the limiting ones has been given on p. 470 of Ref. 1. Justifications based on the shapes of singularities have been given for cases $\epsilon = -1$ (see Refs. 14 and 15).

Here another justification will be presented.

Let us take the following coordinate transformations:

$$Y = [z' + \epsilon_1(x^2+y^2+z'^2)^{1/2}]^{-1}(x+iy),$$
$$p = \epsilon_2(x^2+y^2+z'^2)^{1/2}, \tag{4.1}$$
$$q = t' + \epsilon_2(x^2+y^2+z'^2)^{1/2},$$

for $\epsilon = 1$;

$$Y = (z'+t')^{-1}(x+iy),$$
$$p = \tfrac{1}{2}(z'+t'), \tag{4.2}$$
$$q = (z'+t')^{-1}(x^2+y^2+z'^2-t'^2),$$

for $\epsilon = 0$; and

$$Y = [t' + \epsilon_1(t'^2 - x^2 - y^2)^{1/2}]^{-1}(x + iy),$$
$$p = \epsilon_2(t'^2 - x^2 - y^2)^{1/2},$$
$$q = z' - \epsilon_2(t'^2 - x^2 - y^2)^{1/2},$$
$$t'^2 \geqslant x^2 + y^2, \tag{4.3}$$

for $\epsilon = -1$; where $\epsilon_1{}^2 = \epsilon_2{}^2 = 1$ and coordinates x, y, z', and t' are real. Coordinates z' and t' are additionally involved in the Lorentz transformations

$$z' = (1 - v^2)^{-1/2}(z - vt), \quad t' = (1 - v^2)^{-1/2}(t - vz),$$
$$0 \leqslant v^2 < 1. \tag{4.4}$$

These transformations give us by Eq. (1.2b) that

$$ds_0^2 = dx^2 + dy^2 + dz^2 - dt^2, \tag{4.5}$$

for all the values of ϵ.

Now let us change our choice of units (Sec. I) so as to bring to light the gravitational constant κ keeping $c = 1$. Let us also put $\lambda = 0$ and assume a non-null electromagnetic field, i.e., $B \neq 0$. Keeping Ω as in Eqs. (1.2c) and (2.1) we must then replace Ω by $\kappa\Omega$ in Eq. (1.2a). The justification consists in switching off the gravitational interactions by the limiting transition $\kappa \to 0$. Then from Eq. (1.2a) we obtain $ds^2 = ds_0^2$, thus we find ourselves in the flat space-time where the electromagnetic field still exists and Eqs. (2.7) and (2.8) hold. Using the pseudo-Cartesian coordinate system [Eq. (4.5)] we can clearly determine in the flat space-time every geometrical shape including the shapes of physical singularities. In our case the latter are the shapes of $p = 0$ regions. These regions are singular sources of the electromagnetic fields by virtue of Eqs. (2.7) and (2.8), since $B \neq 0$, in the space-time representation. From relations (4.1), (4.4), and (4.5) we see that the region $p = 0$ is a timelike line for $\epsilon = 1$ and from relations (4.2), (4.4), and (4.5) that the region $p = 0$ is a null hyperplane for $\epsilon = 0$ (see Ref. 16). Thus, in space we have the pointlike singular source moving with a speed less than that of light (pointlike bradyon) for $\epsilon = 1$, and the plane singular source (perpendicular to the z axis) moving with the velocity of light (plane luxon) in the direction $-z$ for $\epsilon = 0$.

The justification presented above has been given by Plebański.[17]

Let us apply this justification to the cases $\epsilon = -1$. Then from relations (4.3)–(4.5) we find that the region $p = 0$ is a null hypersurface.[18] The field of its normal vectors (that are null ones) is well behaved on the whole region $p = 0$ except for the spacelike line $x = y = t' = 0$. Thus this line is *geometrically distinguished* on the hypersurface $p = 0$. In fact, it is the edge of the $p = 0$ light wedge.[19] Thus it can be identified as a world-line of a pointlike tachyon. In space, in the reference frames characterized by $v \neq 0$ the singular source has the shape of a circular cone with the axis z. The cone expands along its normals (in space) with the speed of light and in consequence its vertex moves along the axis z with a velocity $V = v^{-1}$. Since $|V| > 1$ and $c = 1$, this looks as if a pointlike tachyonic singular source moved along the axis z and generated its "ballistic" conical shock surface, which is the singular source of the electromagnetic field. In the $v = 0$ frame the source of the field has the shape of a cylinder with radius $(x^2 + y^2)^{1/2}$ equal to t at every moment $t > 0$ and with axis z, i.e., it expands radially with the speed of light, and the pointlike tachyon has infinite velocity along axis z at moment $t = 0$ (see Ref. 19).

The above justification, consisting in the formal cancellation of the gravitational interactions ($\kappa \to 0$), seems to be reasonable if it is combined with the natural physical interpretation of the irremovable geometrical singularities of curved space-times. This natural interpretation consists of assuming that these singularities are a mathematical idealization and that in the physical reality they represent appropriately high but *finite* concentrations of physical fields (or/and matter) including, of course, the gravitational field (cf. Sec. 2 of Ref. 15).

APPENDIX A: INTRODUCTORY REMARKS AND THE FIRST STEP OF THE PROOF

To prove the theorem, saying that Eqs.(2.1) and (3.1)–(3.3) are all solutions of Eqs. (1.1) with limitation (1.2) for $\epsilon = \pm 1$, the way of direct calculation was used. The main problem consisted of finding a relatively short path among a large number of possibilities. The relatively short path presented below has been in fact a ghastly sequence of calculations. Therefore, we give here only its scheme, putting emphasis on the more important items so as to permit anybody interested to carry out easily the proof himself in detail.

Notation for all Appendices A–E: The $D, E, F, G, H,...$ are complex functions of two variables, Y and q, analytic in Y (or of $\overline{Y}$ and q analytic in $\overline{Y}$ for the symbol with an overbar). The $g, h, j, k, l,...$ are disposable complex functions of one real variable q. The a and b are arbitrary real constants. An integer subscript at any symbol does not change the above general meaning of the symbol. The γ, ζ, η, ϑ, φ, and ψ are integer indices (η, ϑ, φ, and ψ are defined in Appendix C).

To obtain new solutions, i.e., involving the electromagnetic fields, for $\epsilon \neq 0$ we must assume $B \neq 0$ (see Sec. II). Thus it suffices to solve Eqs. (2.2). Taking into account (2.2a) and integrating (2.2b) we get

$$A_{,Y} = 4\overline{B}\,[D + \epsilon B_{,q} Y^{-1}(1 + \epsilon Y\overline{Y})^{-1}], \tag{A1}$$

where D is a disposable function.

Assuming $B_{,q} = 0$ we easily obtain from (A1) and (2.2c) that $A = \text{const}$, i.e., solution (3.1).

Henceforth we put

$$B_{,q} \neq 0. \tag{A2}$$

Let us assume

$$B_{,Y} = 0. \tag{A3}$$

We substitute $A_{,Y}$ from (A1) into (2.2c) and calculate $A_{,qY}$ from such a (2.2c) and $A_{,Yq}$ from (A1). Equating the obtained expressions ($A_{,qY} = A_{,Yq}$ since $A \in C^2$ by the assumption $f \in C^2$) and applying the bar operation we get

$$YD\,(\overline{E}_0 + Y\overline{E}_1 + Y^2\overline{E}_2 + Y^3\overline{E}_3) = \overline{F}_0 + Y\overline{F}_1 + Y^2\overline{F}_2, \tag{A4}$$

where

$$E_0 := D_{,Y} - \epsilon Y^{-2} B_{,q}, \tag{A5a}$$
$$E_1 := \epsilon Y D_{,Y} + 2\epsilon (YD)_{,Y} - Y^{-1} B_{,q}, \tag{A5b}$$
$$E_2 := (Y^2 D)_{,Y} + 2Y(YD)_{,Y}, \tag{A5c}$$
$$E_3 := \epsilon Y (Y^2 D)_{,Y}, \tag{A5d}$$
$$F_0 := -\epsilon \overline{B}_{,q} D_{,Y} + Y^{-2} B_{,q} \overline{B}_{,q} = -\epsilon \overline{B}_{,q} E_0, \tag{A6a}$$
$$F_1 := 2(\overline{B}D)_{,q} - \overline{B}_{,q} D - 2Y\overline{B}_{,q} D_{,Y}$$

$$+ 2\epsilon Y^{-1}(B_{,q}\overline{B})_{,q} + \epsilon Y^{-1}B_{,q}\overline{B}_{,q}, \tag{A6b}$$

$$F_2 := 2\epsilon Y(\overline{B}D)_{,q} - \epsilon Y\overline{B}_{,q}(YD)_{,Y}. \tag{A6c}$$

Relations (A2)–(A5a), (A5d), and (A6a) give us

$$D \neq 0, \quad E_0 \neq 0, \quad E_3 \neq 0. \tag{A7}$$

Since $(YD)_{,\overline{Y}} = 0$ and $E_3 \neq 0$ we obtain from (A4) that

$$F_0 = g_1E_1 + g_2E_2 + g_3E_3, \tag{A8a}$$

$$F_1 = g_1E_2 + g_2E_3, \tag{A8b}$$

$$F_2 = g_1E_3. \tag{A8c}$$

Eliminating $(\overline{B}D)_{,q}$ from (A8b) and (A8c) we get a differential (with respect to Y) equation for D. Its general solution is

$$D = Y^{-1}(\overline{B}_{,q} + 2g_1Y + \epsilon g_2Y^2)^{-1} \times [-2\epsilon(B_{,q}\overline{B})_{,q} - \epsilon B_{,q}\overline{B}_{,q} + h_1Y]. \tag{A9}$$

Substituting such a D into (A4) we obtain by (A6a), (A7), and (A8c) that

$$h_1 = \epsilon\bar{g}_1g_2, \tag{A10}$$

$$(B_{,q}\overline{B})_{,q} = 0, \tag{A11}$$

and by (A7), (A8), (A10), and (A11) that

$$2\epsilon g_1\bar{g}_1 + g_2\bar{g}_2 + B_{,q}\overline{B}_{,q} = 0, \tag{A12}$$

$$2g_1\bar{g}_2 + \bar{g}_1\overline{B}_{,q} + \epsilon g_2\bar{g}_3 = 0. \tag{A13}$$

Relations (A2), (A12), and (A13) give us

$$\epsilon = -1, \quad g_1 \neq 0, \quad g_2 \neq 0. \tag{A14}$$

Next, substituting D from (A9) into (A8a) we get a new algebraic equation involving our functions of q. This equation and (A10)–(A14) enable us to eliminate $\bar{g}_2$, g_3, and $\bar{g}_3$. As a result we obtain the second-degree algebraic equation for g_2 with coefficients expressed by g_1, $\bar{g}_1$, $B_{,q}$, and $\overline{B}_{,q}$. Its two solutions are

$$g_{2(1)} = \bar{g}_1^{-2}B_{,q}(B_{,q}\overline{B}_{,q} - 2g_1\bar{g}_1), \qquad g_{2(2)} = -g_1^2(\overline{B}_{,q})^{-1}. \tag{A15}$$

Relations (A1)–(A3), $A_{,Y\overline{Y}} = \overline{A}_{,\overline{Y}Y}$, (A11), and the transformation $q = q' + a_1$ give us[20]

$$B = (aq)^{1/2}e^{ib}, \quad a \neq 0, \tag{A16}$$

i.e., B from (3.2). Relations (A12) and (A14)–(A16) give us

$$g_1 = \tfrac{1}{2}a^{1/2}q^{-1/2}e^{ih}, \qquad g_2 = -\tfrac{1}{2}a^{1/2}q^{-1/2}e^{i(b+2h)}, \quad h = \bar{h} \tag{A17}$$

(the same result is for both $g_{2(1)}$ and $g_{2(2)}$). Using (A16), (A17), and $\epsilon = -1$ in (A9) and substituting such D and $\overline{D}$ into (A4) we get $h_{,q} = 0$, i.e., $h =$ real const. Transforming $Y = Y'e^{-i(b+h)}$ we obtain[20] the following explicit form of (A1):

$$A_{,Y} = 2aY^{-1}[(1 + Y)^{-1} + (Y\overline{Y} - 1)^{-1}]. \tag{A18}$$

Integrating (A18) and using (A16) and (2.2c) we get solution (3.2), which is the one and only solution in the branch $B_{,q} \neq 0$ and $B_{,Y} = 0$.

Thus all possibilities are exhausted for $B_{,Y} = 0$ and henceforth we put

$$B_{,Y} \neq 0. \tag{A19}$$

Relations (A1), (A2), and (A19) will be valid in all the following Appendices B–E.

APPENDIX B: THE BEGINNING OF BRANCH $B_{,q}B_{,Y} \neq 0$

Let us define

$$G := (B_{,Y})^{-1}B \neq 0, \qquad H := (B_{,Y})^{-1}B_{,q} \neq 0, \qquad J := (B_{,Y})^{-1}D, \tag{B1}$$

that give us

$$G_{,q} = G_{,Y}H - H_{,Y}G. \tag{B2}$$

From (A1) and $A_{,Y\overline{Y}} = \overline{A}_{,\overline{Y}Y}$ we calculate, using (B1), J and from $J_{,\overline{Y}} = 0$ we get

$$H(\overline{K}_0 + Y\overline{K}_1) = G(\overline{L}_0 + Y\overline{L}_1) + \overline{M}_0 + Y\overline{M}_1 + Y^2\overline{M}_2 + Y^3\overline{M}_3, \tag{B3}$$

where

$$K_0 := -1 - G_{,Y}, \tag{B4a}$$

$$K_1 := 2\epsilon G - \epsilon Y(1 + G_{,Y}), \tag{B4b}$$

$$L_0 := -H_{,Y}, \tag{B5a}$$

$$L_1 := 2\epsilon H - \epsilon YH_{,Y}, \tag{B5b}$$

$$M_0 := \epsilon(Y^{-1}H)_{,Y} + J_{,Y}, \tag{B6a}$$

$$M_1 := -3Y^{-1}H + 2H_{,Y} + 3\epsilon YJ_{,Y}, \tag{B6b}$$

$$M_2 := -2\epsilon H + \epsilon YH_{,Y} + 3Y^2J_{,Y}, \tag{B6c}$$

$$M_3 := \epsilon Y^3J_{,Y}. \tag{B6d}$$

Assuming $L_0 = 0$ we have $H = j_1 \neq 0$. Then from $(B3)_{,\overline{Y}}$, (B6c), and (B6d) we obtain $J_{,Y} = 0$, and then from (B4), (B6a), and (B6b) we get $G = -\epsilon\bar{j}_1^{-1}j_1Y^{-1} + j_2 + Y$. Such a G contradicts $j_1 = H \neq 0$ by (B2). Thus we have

$$L_0 \neq 0. \tag{B7}$$

Assuming $L_1 = 0$ we have $H = j_3Y^2$, i.e., $j_3 \neq 0$. Then from $[\overline{L}_0^{-1}(B3)]_{,\overline{Y}}$, (B6a), and (B6b) we obtain $J_{,Y} = -\epsilon j_3$, and then from (B4), (B6c), and (B6d) we get $G = -Y + j_4Y^2 + \epsilon\bar{j}_3^{-1}j_3Y^3$. Such a G contradicts $j_3 \neq 0$ by (B2). Thus we have

$$L_1 \neq 0. \tag{B8}$$

Assumption (B9):

$$K_0 = 0. \tag{B9}$$

This means by (B4) that $G = l_1 - Y$ and $K_1 \neq 0$, which enables us to determine H, explicitly in terms of Y only, from (B3). Substituting such G and H into (B2) we obtain, among other things, that $J_{,Y} = 0$ and $\bar{l}_1(M_0 + \bar{l}_1L_0) = 0$. Assuming $l_1 \neq 0$ we get by (B5a) and (B6a) that $H = l_2Y(\epsilon - Y\bar{l}_1)^{-1}$, which substituted into (B3) gives us by $l_1 \neq 0$ that $l_2 = 0$, contradicting $H \neq 0$. Thus $l_1 = 0$ and from $G = -Y$, $J_{,Y} = 0$, (B1), (B3)–(B6), and (A1) we get

$$A_{,Y} = 4l_3Y^{-2}\overline{Y}^{-1}[l_4 + \epsilon l_{3,q}(1 + \epsilon Y\overline{Y})^{-1}], \qquad B = l_3Y^{-1}e^{ib}, \quad l_3 = \bar{l}_3 \neq 0. \tag{B10}$$

Integrating this $A_{,Y}$ and substituting the obtained A as well as $A_{,Y}$ and B from (B10) into (2.2c) we get, after short calculations of l_3 and l_4 and after the transformations $q = q' + a_1$ (see Ref. 20) and $Y = Y'^{-1}e^{ib}$ (see Ref. 20), solution (3.3).

Assumption (B11):

$$K_1 = 0. \tag{B11}$$

This means by (B4) that $G = Y + l_5Y^2$ and $K_0 \neq 0$. Acting analogously to the preceding assumption we obtain here

$l_5 = 0$ and $J_{,Y} = 0$ and then, after similar calculations, also solution (3.3).

Thus henceforth we can assume

$$K_0 K_1 \neq 0. \tag{B12}$$

Relations (B1)–(B8) and (B12) will be valid in all the following Appendices C–E, where we shall find no solutions.

APPENDIX C: THE BEGINNING OF ASSUMPTION $K_0K_1 \neq 0$

Now we begin the more complicated part of the proof. Using (A1) in $(2.2c)_{,\bar{Y}}$ we express the latter in terms of $G, \bar{G}, H, \bar{H}, J$, and $\bar{J}$ instead of $B, \bar{B}, D$, and $\bar{D}$. Next we calculate J from $A_{,Y\bar{Y}} = \bar{A}_{,\bar{Y}Y}$, using (A1) and (B1), and H from (B3)–(B6). Substituting such H and J into the modified $(2.2c)_{,\bar{Y}}$ we obtain, after a long calculation, that

$$G(\bar{N}_0 + Y\bar{N}_1 + Y^2\bar{N}_2) + \bar{P}_0 + Y\bar{P}_1 + Y^2\bar{P}_2 + Y^3\bar{P}_3 + Y^4\bar{P}_4 = 0, \tag{C1}$$

where N_γ and P_ζ are complicated expressions consisting of G, H, J, their derivatives, and Y. The explicit forms of these expressions will not be necessary in the following. It appears that $P_0 + \cdots + \bar{Y}^4 P_4$ can be presented as $\bar{Y}^2 Y^{-1} G H^2 + (1 + \epsilon Y\bar{Y}) \times$ three-degree polynomial of $\bar{Y}$, thus $P_0 + \cdots + \bar{Y}^4 P_4 \neq 0$ by $GH \neq 0$. This means that there exist γ and ζ such that $N_\gamma \neq 0$ and $P_\zeta \neq 0$, and that G can be determined, explicitly in terms of Y only, from (C1). Thus we can define the following integer indices.

η is the maximal γ such that $N_\gamma \neq 0$.
ϑ is the maximal ζ such that $P_\zeta \neq 0$.
φ is the minimal γ such that $N_\gamma \neq 0$.
ψ is the minimal ζ such that $P_\zeta \neq 0$.

These definitions and (C1) give us

$$N_\eta = n P_\vartheta \neq 0, \tag{C2}$$

$$N_\varphi = r P_\psi \neq 0. \tag{C3}$$

Substituting G from (C1) into (B3), from $H_{,\bar{Y}} = 0$ we get, applying the bar operation, that

$$\{(K_0 + \bar{Y}K_1)^{-1}(\bar{Y}^\varphi N_\varphi + \cdots + \bar{Y}^\eta N_\eta)^{-1}[(\bar{Y}^\varphi N_\varphi + \cdots + \bar{Y}^\eta N_\eta)(M_0 + \bar{Y}M_1 + \bar{Y}^2 M_2 + \bar{Y}^3 M_3) - (\bar{Y}^\psi P_\psi + \cdots + \bar{Y}^\vartheta P_\vartheta)(L_0 + \bar{Y}L_1)]\}_{,Y} = 0, \tag{C4}$$

where the superscripts at the $\bar{Y}$'s are powers (do not confuse the summation convention).

Assumption (C5):

$$J_{,Y} = 0. \tag{C5}$$

From (B5b), (B6c), (B6d), (B8), and (C5) we obtain

$$M_2 = -L_1 \neq 0, \quad M_3 = 0. \tag{C6}$$

From (B8), (B12), (C2), (C4), and (C6) we get

$$L_1 = n_1 K_1 \neq 0, \tag{C7}$$

for all the possible values of η and ϑ, except for the case when $\vartheta = \eta + 1$ and $n = -1$.

If $\vartheta = \eta + 1$ and $n = -1$, then from (C1), $G_{,\bar{Y}} = 0$, and (C2) we obtain $N_{\eta-1} + P_{\vartheta-1} = r_1 P_\vartheta$ for $\eta > 0$ and $P_0 = r_1 P_1$ for $\eta = 0$ (i.e., $\vartheta = 1$). Using this, (C2), the assumptions $\vartheta = \eta + 1$ and $n = -1$, and (C6) in (C4), we get, from such a (C4) and by (B12), that

$$M_1 + L_0 + r_1 L_1 = r_2 K_1. \tag{C8}$$

The separate cases (C7) and (C8) cover the whole domain of (C5).

Case (C7): Integrating Y^{-3}(C7) we get

$$H = n_1(G - Y) + n_2 Y^2. \tag{C9}$$

Substituting such an H into (B3) and using (C6) and (C7) we obtain

$$G[n_1\bar{K}_0 - \bar{L}_0 + Y(n_1 - \bar{n}_1)\bar{K}_1] = \bar{M}_0 + Y(n_1\bar{K}_0 + \bar{M}_1) + Y^2[-n_2\bar{K}_0 + (n_1 - \bar{n}_1)\bar{K}_1] - Y^3 n_2 \bar{K}_1. \tag{C10}$$

Let us assume $\bar{n}_1 K_0 - L_0 = 0$. Integrating this equation and eliminating H from the result and (C9) we get $G(n_1 - \bar{n}_1) = n_3 + (n_1 + \bar{n}_1)Y - n_2 Y^2$, which gives us $n_1 - \bar{n}_1 \neq 0$ by $n_1 \neq 0$. Substituting such a G into (C10) we obtain $M_0 = 0$, i.e., $H = n_4 Y$. This and (C9) give us $n_2 = n_3 = 0$ and then a contradiction by (C10) and by, e.g., (B12) and $n_1 \neq 0$.

Thus we have $n_1\bar{K}_0 - \bar{L}_0 \neq 0$ and can calculate G from (C10). From $G_{,\bar{Y}} = 0$ we get among other results that $M_0 = n_5(\bar{n}_1 K_0 - L_0)$. Integrating this equation and eliminating H from the result and (C9) we obtain G in terms of Y, n_1, n_2, n_5, and n_6. Substituting such a G into (C10) we get a polynomial of Y equal to zero. Its coefficients (equal to zero) give us $M_0 = 0$, i.e., $n_5 = 0$ since $\bar{n}_1 K_0 - L_0 \neq 0$, and then $n_2 = 0$ by (B12) and $n_1 \neq 0$. The final result is

$$G = n_1^{-1}(n_1 + n_6)Y, \quad H = n_6 Y,$$
$$n_1 = \bar{n}_1 \neq 0, \quad n_6 = \bar{n}_6 \neq 0, \quad n_2 = 0, \tag{C11}$$

which is the one and only solution of (C9) and (C10).

From (B2), (C11), and (B12) we get $G = aY$ and $a \neq \pm 1$.

The obtained results and (B1) enable us to express B and D in terms of Y and n_6. Using such B and D in (A1) and $(2.2c)_{,Y}$ we get $D = 0$ and then a contradiction with (A2) by $a \neq \pm 1$. This terminates case (C7).

Case (C8): Integrating Y^{-3} (C8) we get

$$H = (\epsilon - r_1 Y)^{-1}(-G r_2 Y + r_2 Y^2 + r_3 Y^3). \tag{C12}$$

Substituting such an H into (B3) and using (C6) and (C8) we obtain $G \times$ two-degree polynomial of $Y = Y^4 r_3 \bar{K}_1 +$ three-degree polynomial of Y. Substituting G from (C1) into this equation and using (C2), (B12), and the condition $\vartheta = \eta + 1$, which holds here, we get $r_3 = 0$ and then $r_2 \neq 0$ by (C12) and $H \neq 0$. Thus the explicit form of the considered equation is

$$G[\epsilon\bar{L}_0 + Y(r_2\bar{K}_0 - r_1\bar{L}_0 + \epsilon\bar{L}_1) + Y^2(r_2\bar{K}_1 - r_1\bar{L}_1)] = -\epsilon\bar{M}_0 + Y(-\epsilon\bar{r}_2\bar{K}_1 + \epsilon\bar{L}_0 + \epsilon\bar{r}_1\bar{L}_1 + r_1\bar{M}_0) + Y^2[r_2\bar{K}_0 + r_1\bar{r}_2\bar{K}_1 - r_1\bar{L}_0 + (\epsilon - r_1\bar{r}_1)\bar{L}_1] + Y^3(r_2\bar{K}_1 - r_1\bar{L}_1). \tag{C13}$$

From (B7), (C13), and $G_{,\bar{Y}} = 0$ we get among others that $M_0 = \epsilon r_4 L_0$. Integrating this equation and eliminating H from the result and (C12) we obtain G in terms of Y, r_1, r_4, and $r_5 \neq 0$. Substituting such a G into (C13) we get a polynomial of Y equal to zero. Its coefficients (equal to zero) give us, among others, $r_4 = \bar{r}_5$, and the final result is

$$G = Y + r_5(-\epsilon + r_1 Y)(1 + \bar{r}_5 Y)^{-1},$$
$$H = r_2 r_5 Y(1 + \bar{r}_5 Y)^{-1},$$
$$r_1 r_5 = \bar{r}_1\bar{r}_5, \quad r_2 r_5 = \bar{r}_2\bar{r}_5 \neq 0, \quad r_3 = 0. \tag{C14}$$

This is the one and only solution of (C12) and (C13).

Using (A1) in (2.2c) and substituting $\overline{G}$ and $\overline{H}$ from (C14) into $(2.2c)_{,Y}(\overline{B}_{,\overline{Y}})^{-1}$ we get a polynomial of $\overline{Y}$ equal to zero. We split an analysis of its coefficients (equal to zero) into the cases $J=0$ and $J\neq 0$, and then we easily obtain contradictions with, e.g., (A2). This terminates case (C8) and assumption (C5).

Assumption (C15):

$$J_{,Y}\neq 0. \tag{C15}$$

This and (B6d) give us

$$M_3\neq 0. \tag{C16}$$

We shall split the following considerations into the cases $\vartheta>\eta+2$, $\vartheta<\eta+2$, and $\vartheta=\eta+2$.

Case (C17):

$$\vartheta>\eta+2. \tag{C17}$$

From (B8), (B12), (C2), (C4), and (C17) we get

$$L_1=s_1K_1\neq 0. \tag{C18}$$

Integrating Y^{-3}(C18) we obtain

$$H=s_1(G-Y)+s_2Y^2. \tag{C19}$$

Substituting such an H into (B3) we get by (C18) that $G[s_1\overline{K}_0-\overline{L}_0+Y(s_1-\overline{s}_1)\overline{K}_1]=$ three-degree polynomial of Y. From (C1), (C2), (C17), and (C18) we then obtain that $s_1=\overline{s}_1$ and next from (C19) that $s_1\overline{K}_0-\overline{L}_0=-2s_1+2\overline{s}_2\overline{Y}\neq 0$. Thus G is a three-degree polynomial of Y. Substituting such a G and H from (C19) into (B2) we get $s_2=0$. The final result is

$$-2s_1G=\overline{M}_0+Y(s_1\overline{K}_0+\overline{M}_1)+Y^2(s_1\overline{K}_1+\overline{M}_2)+Y^3\overline{M}_3. \tag{C20}$$

Thus the coefficients of the polynomial of Y on the right-hand side of (C20) are functions of q only. This gives a system of four equations, which enables us, by (B4) and (B6), to eliminate $G_{,Y}$, $H_{,Y}$, and $J_{,Y}$. Then it appears that H is also eliminated. As a result we get $G=s_3Y^{-1}+s_4+s_5Y+s_6Y^2$, which contradicts (C16) and (C20). This terminates case (C17).

Case (C21):

$$\vartheta<\eta+2. \tag{C21}$$

From (B12), (C2), (C4), (C16), and (C21) we get

$$M_3=u_1K_1\neq 0. \tag{C22}$$

Integrating Y^{-3} (C22) we obtain

$$J=u_1(Y^{-1}-Y^{-2}G)+u_2. \tag{C23}$$

Using (A1) and (B1) in $A_{,Y\overline{Y}}=\overline{A}_{,\overline{Y}Y}$ and substituting J from (C23) into such an $A_{,Y\overline{Y}}=\overline{A}_{,\overline{Y}Y}$ we get H in terms of G, Y, u_1, u_2, their complex conjugates, and $\overline{H}$. Substituting such an H into (B3) we obtain, by (C22), that $G(u_1\overline{K}_0+Y\overline{Q}_1+Y^2\overline{Q}_2+Y^3\overline{Q}_3)=Y(u_1\overline{K}_0+\epsilon\overline{M}_0)+Y^2\overline{R}_2+Y^3\overline{R}_3+Y^4\overline{R}_4+Y^5(u_2-\overline{u}_2)\overline{Y}^2\overline{K}_1$. (The explicit forms of Q_γ and R_ζ will not be necessary in the following.) This, (B12), (C1), (C2), and (C21) give us $u_2=\overline{u}_2$. From (B12) and (C22) we have $u_1\overline{K}_0\neq 0$, thus G can be calculated (explicitly in terms of Y only). From $G_{,\overline{Y}}=0$ we get $\overline{u}_1K_0+\epsilon M_0=u_3K_0$. Integrating this equation and eliminating J from the result and (C23) we obtain

$$H=G[\epsilon u_1Y^{-1}+(\overline{u}_1-u_3)Y]-\epsilon u_1+u_4Y+(\overline{u}_1-u_3)Y^2. \tag{C24}$$

Substituting H from (C24) and J from (C23) into $A_{,Y\overline{Y}}=\overline{A}_{,\overline{Y}Y}$, where (A1) and (B1) are applied, and using $u_2=\overline{u}_2$ we get

$$G=(\overline{S}_0+Y\overline{S}_1+Y^2\overline{S}_2)^{-1}(Y\overline{T}_1+Y^2\overline{T}_2+Y^3\overline{T}_3), \tag{C25}$$

where

$$S_0:=-\epsilon\overline{u}_1Y(G+Y)\neq 0, \tag{C26a}$$

$$S_1:=G[\epsilon u_1+(\overline{u}_1-u_3)Y^2]-\epsilon\overline{u}_3Y+u_4Y^2-u_3Y^3, \tag{C26b}$$

$$S_2:=(u_1-\overline{u}_3)Y(Y-G), \tag{C26c}$$

$$T_1:=-\epsilon u_3Y(G+Y)+\epsilon(u_4-\overline{u}_4)Y, \tag{C27a}$$

$$T_2:=G[-\epsilon u_1+\overline{u}_4Y+(\overline{u}_1-u_3)Y^2]+\epsilon\overline{u}_3Y+(u_4-\overline{u}_4)Y^2-u_3Y^3, \tag{C27b}$$

$$T_3:=\overline{u}_3Y(Y-G). \tag{C27c}$$

Inequality $S_0\neq 0$ results from (B12) and (C22). From (C25), $G_{,\overline{Y}}=0$, (C26a), and (C27a) we get $u_4=\overline{u}_4$ by (B12). From (C1), (C2), (C21), (C22), (C25), (C26c), and (C27c) we get $S_2\neq 0$, i.e., $u_1-\overline{u}_3\neq 0$.

Substituting G from (C25) and H from (C24) into (B2) we obtain a polynomial of Y equal to zero. Its coefficient at the maximal power of Y is $(\overline{u}_1-u_3)(\overline{S}_2+\overline{T}_3)\overline{T}_3$ $(=0)$. This means, by (C22) and $\overline{u}_1-u_3\neq 0$, that $T_3=0$. This implies $u_3=0$ and then $T_1=0$ by $u_4=\overline{u}_4$. Then the coefficient at Y is $2\epsilon u_1\overline{S}_0\overline{T}_2$ $(=0)$, i.e., $T_2=0$. Thus we get $T_1=T_2=T_3=0$ that contradicts $G\neq 0$ by (C25). This terminates case (C21).

Case (C28):

$$\vartheta=\eta+2. \tag{C28}$$

From (B8), (B12), (C2), (C4), (C16), and (C28) we get

$$nM_3=L_1+sK_1. \tag{C29}$$

Let us assume

$$M_0=0. \tag{C30}$$

From (B5a), (B6a), (B6b), and (C30) we have

$$M_1=L_0. \tag{C31}$$

Eliminating $J_{,Y}$ from (C29) and (C30) and integrating the result we obtain $H=(\epsilon nY-1)^{-1}(sG-sY+v_1Y^2)$. Substituting such an H into (B3) we get $G(\cdots+\epsilon nY^2\overline{L}_1)=\cdots-\epsilon nY^4\overline{M}_3$. Substituting G from (C1) and using (C2) and (C28) we obtain $nM_3=L_1$. This means, by (B12) and (C29), that

$$s=0. \tag{C32}$$

The final result, taking into account (C29)–(C32), is

$$G=[-\overline{L}_0+Y(\epsilon n\overline{L}_0-\overline{L}_1)+\epsilon nY^2\overline{L}_1]^{-1}\times Y[\overline{L}_0+Y(v_1\overline{K}_0-\epsilon n\overline{L}_0+\overline{M}_2)+Y^2(v_1\overline{K}_1+\overline{n}^{-1}\overline{L}_1-\epsilon n\overline{M}_2)-\epsilon n\overline{n}^{-1}Y^3\overline{L}_1], \tag{C33}$$

$$H=v_1Y^2(\epsilon nY-1)^{-1},\quad J=v_2-\epsilon Y^{-1}H, \tag{C34}$$

i.e., $v_1\neq 0$ since $H\neq 0$. From (C33) and $G_{,\overline{Y}}=0$ we get a system of equations involving, after the bar operation, the expressions K_0, K_1, L_0, L_1, and M_2. From this system and (C34) we obtain

$$G=(\epsilon nY-1)^{-1}Y(1+\bar{n}^{-1}Y-\epsilon n\bar{n}^{-1}Y^2),$$

$$n\bar{v}_1=\bar{n}v_1. \tag{C35}$$

Equations (C34) and (C35) are the one and only solution of (C29)–(C33).

Using (A1) in (2.2c) and substituting $\overline{G}$, $\overline{H}$, and $\overline{J}$ from (C34) and (C35) into $(2.2\text{c})_{,Y}(\overline{B}_{,\overline{Y}})^{-1}$ we get a polynomial of $\overline{Y}$ equal to zero. We split an analysis of its coefficients (equal to zero) into the cases $v_2\neq 0$ and $v_2=0$. For $v_2\neq 0$ we obtain a contradiction with (C35), and for $v_2=0$ with, e.g., (A2) or (C15). This terminates subcase (C30).

Thus henceforth we put

$$M_0\neq 0. \tag{C36}$$

APPENDIX D: CONTINUATION OF CASE (C28) FOR $M_0\neq 0$

Relations (A1), (A2), (A19), (B1)–(B8), (B12), (C1)–(C4), (C15), (C16), (C28), (C29), and (C36) are valid in the following.

We shall split our considerations into the assumptions $\psi>\varphi$, $\psi<\varphi$, and $\psi=\varphi$ [see (C3) and (C4)].

Assumption (D1):

$$\psi>\varphi. \tag{D1}$$

From (B12), (C3), (C4), (C36), and (D1) we get

$$M_0=w_1K_0\neq 0. \tag{D2}$$

Eliminating $J_{,Y}$ from (C29) and (D2) and integrating the result we obtain $H=(\epsilon nY-1)^{-1}[G(s-nw_1Y^2)-sY+w_2Y^2-nw_1Y^3]$. Substituting such an H into (B3) we get $G=(\cdots-nw_1Y^3\overline{K}_1)^{-1}[\cdots+nY^4(w_1\overline{K}_1+\epsilon\overline{M}_3)]$. Substituting such a G into (C1) we obtain a contradiction with (B12), (C2), and (D2) by (C28). This terminates assumption (D1).

Assumption (D3):

$$\psi<\varphi. \tag{D3}$$

From (B7), (B12), (C3), (C4), and (D3) we get

$$L_0=w_3K_0\neq 0. \tag{D4}$$

Integrating (D4) we obtain

$$H=w_3(G+Y)+w_4. \tag{D5}$$

Substituting such an H into (B3) we get, by (C16) and (D4), that

$$\begin{aligned}G\,&[(w_3-\bar{w}_3)\overline{K}_0+Y(w_3\overline{K}_1-\overline{L}_1)]\\&=\overline{M}_0-w_4\overline{K}_0+Y(\overline{M}_1-w_3\overline{K}_0-w_4\overline{K}_1)\\&\quad+Y^2(\overline{M}_2-w_3\overline{K}_1)+Y^3\overline{M}_3\neq 0.\end{aligned} \tag{D6}$$

Substituting G from (D6) into (C1) and taking into account (B12), (C3), and (D3) we obtain $w_3=\bar{w}_3$, i.e., $w_3\overline{K}_1-\overline{L}_1\neq 0$ by (D6), and then $\overline{M}_0-w_4\overline{K}_0\neq 0$ by (C1), (C3), (D3), and (D6). Using these in (D6) and substituting then G from (D6) and H from (D5) into (B2) we get $w_4=0$. From (D5), $w_3=\overline{w}_3$, and $w_4=0$ we obtain that $w_3\overline{K}_1-\overline{L}_1=-2\epsilon\bar{w}_3\overline{Y}$ and then from (C16), (C36), and (D6) that $M_0=w_5Y\neq 0$ and $M_3=w_6Y\neq 0$. Eliminating $J_{,Y}$ from the two last relations and integrating the result we get $H=w_6+w_7Y+\frac{1}{2}\epsilon w_5Y^3$. Substituting such an H into (D5) we obtain G that contradicts, e.g., (C36) by (D6) and $w_3-\bar{w}_3=w_4=0$. This terminates assumption (D3).

Assumption (D7):

$$\psi=\varphi. \tag{D7}$$

From (B7), (B12), (C3), (C4), (C36), and (D7) we get

$$rM_0=L_0+kK_0. \tag{D8}$$

Eliminating $J_{,Y}$ from (C29) and (D8) [$nr\neq 0$ by (C2) and (C3)] and integrating the result we get H in terms of G, Y, and functions of q. Substituting such an H into (B3) we obtain an equation that enables us, by (C16) or (C36), to calculate G (explicitly in terms of Y only). Substituting such a G into (C1) and using (C2), (C28), and (C29) [or (C3), (D7), and (D8)] we get, by (B12), that

$$k=\bar{s}. \tag{D9}$$

Using (D9), the function H that has been obtained above is

$$\begin{aligned}H=(-1+\epsilon nY+nr^{-1}Y^2)^{-1}[G(s-nr^{-1}\bar{s}Y^2)\\-sY+nuY^2-nr^{-1}\bar{s}Y^3].\end{aligned} \tag{D10}$$

Integrating (D8) and using (D9) and (D10) we get

$$\begin{aligned}J=(-1+\epsilon nY+nr^{-1}Y^2)^{-1}\{G[-\epsilon sY^{-1}+r^{-1}(\bar{s}-s)]\\+\epsilon s+w+[r^{-1}(s+\bar{s})-\epsilon n(u+w)]Y\\-nr^{-1}(u+w)Y^2\}.\end{aligned} \tag{D11}$$

Substituting G from (C1) and H from (D10) into (B2) and using (C2), (C3), (C28), and (D7) we get a polynomial of Y equal to zero. Its coefficients (equal to zero) at the maximal and minimal powers of Y give us

$$n_{,q}=s(n+\epsilon\bar{r}),\quad r_{,q}=\bar{s}(\epsilon\bar{n}+r). \tag{D12}$$

Substituting H and $\overline{H}$ from (D10) and J and $\overline{J}$ from (D11) into $A_{,Y\overline{Y}}=\overline{A}_{,\overline{Y}Y}$, where (A1) and (B1) are used, we get

$$\mathscr{P}_1+\mathscr{P}_2G+\mathscr{P}_3\overline{G}+\mathscr{P}_4G\overline{G}=0, \tag{D13}$$

where the $\mathscr{P}_\gamma$'s are polynomials of Y and $\overline{Y}$ with coefficients consisting of n, r, s, u, w, and their complex conjugates. The explicit form of (D13) is very long.

Substituting G from (C1) into (D13) and using (C2), (C3), (C28), and (D7) we obtain a polynomial of Y equal to zero. Its coefficients (equal to zero) at the maximal and minimal powers of Y, after being divided by $\overline{P}_\vartheta$ and $\overline{P}_\psi$, respectively, are polynomials of $\overline{Y}$ since the expressions including $\overline{G}$ cancel themselves there. Coefficients (equal to zero) of these two polynomials of $\overline{Y}$ give us

$$u-\bar{u}=(s-\bar{s})[(n\bar{n})^{-1}+(r\bar{r})^{-1}], \tag{D14}$$

$$w-\bar{w}=(\bar{s}-s)(r\bar{r})^{-1}. \tag{D15}$$

Let us assume $s=0$. This means that $u\neq 0$ by (D10) and $H\neq 0$. The condition $s=0$ makes (D13) much simpler, e.g., $\mathscr{P}_4=0$. Using such a (D13) we can, by $u\neq 0$, express G as a quotient of polynomials of Y. Dividing these polynomials themselves we get $G=-\bar{n}^{-1}Y^2+Y(1-2\epsilon\bar{n}^{-1}\overline{Y}^{-1})+$ a quotient of two-degree polynomials of Y. This contradicts $G_{,\overline{Y}}=0$. Thus we have

$$s\neq 0. \tag{D16}$$

APPENDIX E: CONTINUATION OF ASSUMPTION (D7) FOR $s \neq 0$

Relations (A1), (A2), (A19), (B1)–(B8), (B12), (C1)–(C4), (C15), (C16), (C28), (C29), (C36), and (D7)–(D16) are valid in the following part of the proof where some new tricks have to be applied.

Using (A1) in (2.2c) and substituting $\overline{H}$ from (D10), $\overline{J}$ from (D11), and then $\overline{G}$ from (C1) into $(2.2c)_{,Y}(\overline{B}_{,\overline{Y}})^{-1}$ we get a polynomial of $\overline{Y}$ equal to zero. Its coefficients (equal to zero) at the maximal and minimal powers of $\overline{Y}$ give us, by (C2), (C3), (C28), (D7), (D14), and (D15), that

$$D[s+2n(u+w)Y]+D_{,Y}[-sY+n(u+w)Y^2] -2D_{,q}=0, \tag{E1}$$

$$3\bar{s}U+U_{,Y}(rw+\bar{s}Y)-2U_{,q}=0, \tag{E2a}$$

where

$$U:=\epsilon Y^{-1}B_{,q}+D\neq 0. \tag{E2b}$$

Inequality $U\neq 0$ results from (C36).

Using (B1) we express (E1) and (E2) in terms of G, H, and J instead of B and D. Substituting H from (D10), J from (D11), and then G from (C1) into the modified (E1) and (E2) we get two polynomials of Y equal to zero. Some their coefficients (equal to zero) enable us to find the equations involving our functions of q. The coefficient at the minimal power of Y in the polynomial generated by (E1) gives us

$$s_{,q}=2s^2, \tag{E3}$$

by (C3), (D7), and (D16). The coefficient at the power of Y following the minimal power in the same polynomial gives us, after a long calculation, that

$$2w_{,q}=-2\epsilon s^2-3n\bar{r}^{-1}s^2+\epsilon n\bar{r}^{-1}su+sw(1-\epsilon n\bar{r}^{-1}) +(r\bar{r})^{-1}(\bar{s}-s)(s+2\bar{s}-2\epsilon\bar{n}r^{-1}\bar{s}), \tag{E4}$$

by (C3), (D7), (D12), (E3), and by the fact that the terms including $\overline{N}_{\varphi+1}$ and $\overline{P}_{\psi+1}$ cancel themselves. The coefficient at the maximal power of Y in the polynomial generated by (E2) vanishes by (E3). The coefficient at the power of Y preceding the maximal power in the same polynomial gives us

$$2(u+w)_{,q}=-2\epsilon\bar{s}^2-3\bar{n}^{-1}r\bar{s}^2+\bar{s}u(1-2\epsilon\bar{n}^{-1}r) +\bar{s}w(1-\epsilon\bar{n}^{-1}r)+(n\bar{n})^{-1}(s-\bar{s}) \times(\bar{s}+2s-2\epsilon n^{-1}r\bar{s}), \tag{E5}$$

by (C2), (C28), (D12), and (E3) (the terms including $\overline{N}_{\eta-1}$ and $\overline{P}_{\vartheta-1}$ do not appear there).

The general solution of (E2a) with condition (E3) is

$$U=\bar{s}^{3/4}W(X), \tag{E6a}$$

$$X:=\bar{s}^{1/4}Y+\frac{1}{2}\int r\bar{s}^{1/4}w\,dq, \tag{E6b}$$

where W is an arbitrary, but different from zero by (E2b), analytic function of one complex variable X.

Using (E6b) as a coordinate transformation we now introduce a new coordinate system $X, \overline{X}, p, q$. This is possible by (D16).

Let us consider an expression $Z:=X^2W_{,X}W^{-1}$. In the new coordinate system we have $Z_{,q}=0$. Using (C1), (D10), (D11), and (E6b) we express G, H, and J explicitly in terms of X instead of Y. Next, using (B1), (E2b), the H and J just now obtained in terms of X, (C28), and (E6) we get, by (A19), Z as a quotient of polynomials of X. A degree of the polynomial in the numerator is greater by 1 than that in the denominator. In each one of these polynomials we explicitly calculate, using (C2), (D14), and (D15), only the coefficients at two maximal powers of X. Dividing these polynomials themselves we get $Z=X+t_1+$ the rest of the division, where $t_1:=\epsilon\bar{s}^{1/4}[-r+3\bar{n}\bar{s}^{-1}(\bar{u}+\bar{w})]+\frac{1}{2}\int r\bar{s}^{1/4}w\,dq$. We have $Z_{,q}=0$ here, thus t_1 has to be constant, i.e., $t_{1,q}=0$. This gives us, after the bar operation, that

$$6(u+w)_{,q}=2\epsilon s^2+3n^{-1}\bar{r}s^2+3su(1-2\epsilon n^{-1}\bar{r}) +sw(3-7\epsilon n^{-1}\bar{r})+\epsilon s(nr)^{-1}(\bar{s}-s). \tag{E7}$$

Using (D12), (D14), (D15), and (E3) we find that $(\bar{u}+\bar{w})_{,q}=(u+w)_{,q}+$ terms including n, r, s, and their complex conjugates (without derivatives). Substituting such a $(\bar{u}+\bar{w})_{,q}$ into $\overline{\text{(E5)}}$ and using there (D14) and (D15) with respect to $\bar{u}$ and $\bar{w}$ leaving $(u+w)_{,q}$ intact we get another equation determining the latter. Eliminating $(u+w)_{,q}$ from this equation and (E7) we obtain

$$w=s(3\epsilon+2n\bar{r}^{-1})+(r\bar{r})^{-1}(\bar{s}-s). \tag{E8}$$

Substituting such a w into (E4) and using (D12) and (E3) we get an equation determining u. Adding this equation and (E8) we obtain

$$u+w=15s(\epsilon+n^{-1}\bar{r}). \tag{E9}$$

Eliminating u and w by means of (E8) and (E9) we easily get from the appropriate preceding equations and (D16) that

$$n+\epsilon\bar{r}=0 \tag{E10a}$$

and then that

$$\begin{aligned} &s=\bar{s}, \quad u=\bar{u}=-\epsilon s,\\ &w=\bar{w}=\epsilon s, \quad u+w=0. \end{aligned} \tag{E10b}$$

From (D12) and (E10a) we have

$$n_{,q}=r_{,q}=0. \tag{E11}$$

Now restoring the old $Y, \overline{Y}, p, q$ coordinate system and using (E10) in (D13) we get after a calculation that

$$G=\epsilon n^{-1}+Y-\bar{n}^{-1}Y^2, \tag{E12a}$$

and then from (D10), (D11), and (E10) that

$$H=-sG, \quad J=n^{-1}sY^{-1}. \tag{E12b}$$

Using (A1) in $(2.2c)_{,Y}(B_{,Y}\overline{B}_{,\overline{Y}})^{-1}$ and expressing the latter in terms of G, H, J, their derivatives, $\overline{G}, \overline{H}$, and $\overline{J}$ we obtain, by (E12), a polynomial of Y and $\overline{Y}$ equal to zero. Its coefficients depend algebraically on n and $\bar{n}$ only, by (D16), (E3), and (E11), and are different from zero, which is a contradiction.

This terminates the proof since all possibilities have been exhausted, provided Providence has protected me from making a mistake in the calculations (which were carefully checked).

[1] I. Robinson and A. Trautman, Proc. R. Soc. London Ser. A **265**, 463 (1962).

[2] K. Bajer and J. K. Kowalczyński, J. Math. Phys. **26**, 1330 (1985).

[3] J. K. Kowalczyński and J. F. Plebański, Int. J. Theor. Phys. **16**, 357 (1977); Errata **17**, 387 (1978).

[4] Those expressions can be found in Sec. 2 of Ref. 3. They relate to a metric form more general than our (1.2) but they are limited by the assumption $\epsilon=\pm 1$. Our form (1.2) was obtained by putting in Ref.

3 $m=\bar{m}=\text{const}$, $\alpha=2^{-1/2}(1+\epsilon Y\bar{Y})$, and $\beta=\sigma=0$ (see Lemma 2 there). As regards $\epsilon=0$ those expressions can also be used (see Ref. 4 in Ref. 2).

[5]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (VEB DVW, Berlin, 1980), Theorem 24.7.

[6]I. Robinson, A. Schild, and H. Strauss, Int. J. Theor. Phys. **2**, 243 (1969).

[7]Equations more general than our (2.2) and (2.3) have also been given on p. 472 of Ref. 1 although a generalization of Eq. (2.2c) has been presented there in a context unnecessarily confined to a null electromagnetic field.

[8]The possible problems with the sign and dimension [q has dimension of length ($c=1$)] of the logarithm arguments are solvable in the standard way (cf. Ref. 11).

[9]P. A. Hogan and G. O'Brien, Phys. Lett. A **67**, 335 (1978), have presented a solution, belonging to a special case of the Robinson–Trautman metric form different from our (1.2), that belongs to the Schwarzschild branch and has no counterpart in the anti-Schwarzschild branch.

[10]G. C. Debney, R. P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).

[11]They have been signaled by J. K. Kowalczyński, Phys. Lett. A **65**, 269 (1978).

[12]I. Robinson and A. Trautman, Phys. Rev. Lett. **4**, 431 (1960).

[13]S. N. Pandey, Ind. J. Pure Appl. Math. **6**, 109 (1975); J. F. Plebański, Ann. Phys. (NY) **90**, 196 (1975), especially p. 234. A justification for $\epsilon=-1$ given by A. Peres, Phys. Lett. A **31**, 361 (1970), seems to be not good.

[14]P. C. Vaidya, Curr. Sci. **40**, 651 (1971).

[15]J. K. Kowalczyński, Phys. Lett. A **74**, 157 (1979).

[16]This does agree with the supposition on p. 470 of Ref. 1 for $\epsilon=1$, but does not (with respect to the shape of the region $p=0$) for $\epsilon=0$. A null path for the latter has been suggested there as a result of the limiting transition $\epsilon\to 0$, while we have a null hyperplane. This would mean that transition $\epsilon\to 0$ is not a good trick, especially that parameter ϵ is discrete [for our special case of the Robinson–Trautman metric form; as regards mathematical properties of ϵ (denoted by K in Ref. 1) for the general Robinson–Trautman metric form see Ref. 1].

[17]J. F. Plebański (private communication).

[18]The coordinate system $Y, \bar{Y}, p, q$ cannot be employed to examine the properties of hypersurface $p=0$ since the coordinate p is badly behaved in the region $p=0$ [see Eq. (1.2b)]. The coordinate system x, y, z, t, behaving in the best way everywhere [Eq. (4.5)], is the best for this but we cannot do this by means of the direct substitution from relations (4.3) since then the normal vectors of the hypersurface $p=0$ would be undetermined. Thus such an equivalent expression must be used which will make these vectors determined, e.g., $p^2=t'^2-x^2-y^2=0$. Then it appears that the normal vectors are null, i.e., the hypersurface $p=0$ is null.

[19]For a better visualization the reader can have a look at the shapes of the region $p=0$ (in space-time) and shock surfaces Σ_v^p (in space) in Figs. 1 and 2 in Ref. 15 where spatial cylindrical coordinates have been used.

[20]In the following we drop the prime at the new coordinate. Note that such a transformation does not change Eqs. (1.2) and (2.1)–(2.4).

Charged perfect fluid in rigid rotation

S. K. Chakraborty and N. Bandyopadhyaya
Physics Department, Presidency College, Calcutta, India

(Received 27 July 1984; accepted for publication 14 December 1984)

This paper is a continuation of the work of Bonnor and Raychaudhuri on a charged dust distribution in rigid rotation in general relativity. Here the authors are concerned with the similar problem of a charged perfect fluid with nonvanishing pressure. As in Raychaudhuri's work, symmetry-independent reduction of Einstein–Maxwell equations is undertaken. Certain assumptions in Raychaudhuri's paper regarding the inheritance by the electromagnetic field of the symmetry resulting from the rigidity of the motion are justified.

I. INTRODUCTION

In a recent paper Raychaudhuri[1] considered the problem of a rigidly rotating charged dust distribution in general relativity. He introduced no symmetry assumption and reduced the Einstein–Maxwell equations to a relatively simple set of equations on the assumption that the electromagnetic four-potential (A^μ) and the fluid velocity vector (v^μ) are everywhere coincident in direction, i.e., $A^\mu = kv^\mu$, where k is a scalar. In that paper Raychaudhuri did not mention any particular gauge, nor did he justify his assumptions regarding the inheritance of the symmetry consequent upon the rigidity of motion by the local electric and magnetic fields and the scalar k. These points have been clarified in the present paper (in the Appendix) and the corresponding situation in the perfect fluid case is considered. Particular attention is given to the two interesting and simple cases where either k is a constant or the fluid motion is geodesic.

II. REDUCTION OF THE EINSTEIN–MAXWELL FIELD EQUATIONS

From the condition of rigid motion one has

$$v^\mu = \lambda \xi^\mu,$$

with

$$\lambda_{,\mu}\xi^\mu = \sigma_{\alpha\beta} = \theta = 0$$

and $\qquad$ (2.1)

$$\dot v^\mu (\equiv v^\mu{}_{;\nu} v^\nu) = (\ln \lambda)_{;\mu},$$

where ξ^μ is a timelike Killing vector field, λ is a scalar field, and $\sigma_{\alpha\beta}$ and θ are the shear tensor and expansion, respectively. Defining the local electric and magnetic fields by

$$E^\mu = F^{\alpha\mu} v_\alpha,$$

$$B^\mu = \tfrac{1}{2}\eta^{\alpha\mu\nu\sigma} F_{\nu\sigma} v_\alpha,$$

one has, from Maxwell's equations (assuming convective current only)

$$E^\alpha{}_{;\alpha} + E^\alpha \dot v_\alpha - 2B^\alpha \omega_\alpha = 4\pi\sigma, \tag{2.2}$$

$$(E^\beta{}_{;\alpha} - E_\alpha{}^{;\beta}) v^\alpha - \eta^{\beta\alpha\lambda\sigma}(\dot v_\lambda v_\alpha B_\sigma + v_\lambda B_{\sigma;\alpha}) = 0, \tag{2.3}$$

$$B^\alpha{}_{;\alpha} + B^\mu \dot v_\mu + 2E_\alpha \omega^\alpha = 0, \tag{2.4}$$

$$(B^\mu{}_{;\alpha} - B_\alpha{}^{;\mu}) v^\alpha - \eta^{\lambda\alpha\mu\sigma}(\dot v_\lambda v_\sigma E_\alpha - v_\lambda E_{\sigma;\alpha}) = 0, \tag{2.5}$$

where σ is the charge density and ω^α is the vorticity vector. From the condition of the vanishing of $\underset{\xi^\alpha}{\mathscr{L}} E^\mu$ *and* $\underset{\xi^\alpha}{\mathscr{L}} B^\mu$ (see the Appendix) it follows that

$$v^\alpha (E_{\alpha;\mu} - E_{\mu;\alpha}) = 0 \tag{2.6}$$

and

$$v^\alpha (B_{\alpha;\mu} - B_{\mu;\alpha}) = 0. \tag{2.7}$$

Using (2.6) and (2.3), we have

$$B_\mu = \lambda \psi_{,\mu}, \tag{2.8}$$

where ψ is some scalar field.

Now from the relation $A^\mu = kv^\mu$, which is assumed to hold in the Lorentz gauge, it follows that $\underset{\xi^\alpha}{\mathscr{L}} k = 0$ (see the Appendix), and one has

$$E_\mu = -k\dot v_\mu + k_{,\mu} \tag{2.9}$$

and

$$B_\mu = -2k\omega_\mu. \tag{2.10}$$

Einstein's equations

$$\begin{aligned} G^\mu{}_\nu = 8\pi\big[&(\rho + p)v^\mu v_\nu - p\delta^\mu{}_\nu - (1/4\pi)(\tfrac{1}{2}\delta^\mu{}_\nu - v^\mu v_\nu) \\ &\times (E^2 + B^2) - (1/4\pi)(E^\mu E_\nu + B^\mu B_\nu) \\ &- (1/4\pi)(v^\mu S_\nu + v_\nu S^\mu)\big], \end{aligned}$$

where $S^\mu \equiv \eta^{\mu\nu\rho\sigma} E_\nu B_\rho v_\sigma$ is the Poynting vector, $B^2 = -B_\mu B^\mu$, $E^2 = -E_\mu E^\mu$, $\omega^2 = -\omega_\mu \omega^\mu$, and Maxwell's equations,

$$F^{\mu\nu}{}_{;\nu} = 4\pi J^\mu (\equiv 4\pi\sigma v^\mu),$$

$$*F^{\mu\nu}a_{;\nu} = 0,$$

in conjunction with the identity

$$v^\mu{}_{;\mu;\sigma} - v^\mu{}_{;\sigma;\mu} = R_{\sigma\alpha} v^\alpha,$$

now give

$$\dot v^\mu{}_{;\mu} + 2\omega^2 = 4\pi(\rho + 3p) + (E^2 + B^2), \tag{2.11}$$

$$2S^\gamma = -(\eta^{\mu\nu\beta\gamma}/k)(k\dot v_\beta + k_{,\beta})\omega_\mu v_\nu. \tag{2.12}$$

Taking the divergence of Einstein equations, one gets

 0022-2488/85/071752-03$02.50

$$\dot v^\nu = \frac{p_{,\mu}(g^{\mu\nu} - v^\mu v^\nu)}{\rho + p} + \frac{\sigma E^\nu}{\rho + p}. \tag{2.13}$$

Equations (2.9) and (2.13) now give

$$\dot v_\mu = \frac{\sigma k_{,\mu} + p^{,\alpha}(g_{\alpha\mu} - v_\alpha v_\mu)}{\rho + p + \sigma k} \tag{2.14}$$

and

$$E_\mu = \frac{(\rho + p)k_{,\mu} - kp^{,\alpha}(g_{\alpha\mu} - v_\alpha v_\mu)}{\rho + p + \sigma k}. \tag{2.15}$$

Furthermore, (2.9), (2.12), and (2.13) give

$$2S^\gamma = -\frac{S^\gamma}{2k^2}\left(1 + \frac{\sigma k}{\rho + b}\right) - \frac{2\eta^{\mu\nu\beta\gamma}p^{,\alpha}(g_{\alpha\beta} - v_\alpha v_\beta)\omega_\mu v_\nu}{\rho + p}. \tag{2.16}$$

We now restrict ourselves to the following interesting cases.

Case A: k is constant, $S^\gamma \neq 0$.

From (2.9), (2.10), (2.15), and (2.16) we get

$$k^2 = \tfrac{1}{4}, \quad \text{if} \quad \rho + p \neq 0.$$

It is interesting to note that the same value of k is obtained in the corresponding pressure-free case.[1] The following set of equations now readily follows:

$$\Box A = \pm 2[4\pi\sigma/3 \pm (8\pi/3)(\rho + 3p)], \tag{2.17}$$

$$\Box \psi = -3\psi_{,\mu}A^{,\mu}, \tag{2.18}$$

$$A_{,\mu}A^{,\mu} = p_{,\mu}A^{,\mu}/\rho + p \pm \sigma/2, \tag{2.19}$$

$$[e^{2A}\psi_{,\mu}\psi^{,\mu} - \tfrac{1}{4}A_{,\mu}A^{,\mu}] = (8\pi/3)[\sigma \pm \tfrac{1}{2}(\rho + 3p)], \tag{2.20}$$

with

$$E_\mu = \mp \tfrac{1}{2}A_{,\mu} \tag{2.21}$$

and

$$B_\mu = e^A\psi_{,\mu}. \tag{2.22}$$

In the above set of equations, A is a scalar field, the upper sign refers to $k = \frac{1}{2}$, and the lower sign refers to $k = -\frac{1}{2}$. Equations (2.17)–(2.20) give four equations in the five unknowns ρ,σ,p,A,ψ. Thus one can solve the set of equations by introducing a specific equation of state $p = p(\rho)$.

Case B: k is constant, $S^\gamma = 0$, $E_\mu \neq 0$, $B_\mu \neq 0$.

In this case one readily derives the following set of equations:

$$\Box A + (\chi'/\chi + 2)A_{,\mu}A^{,\mu} = 0, \tag{2.23}$$

$$[k(\chi'/\chi) + \chi^2 + k]A_{,\mu}A^{,\mu} = 4\pi\sigma, \tag{2.24}$$

$$[\chi'/\chi - (1 - (1/2k^2))\chi^2 + k^2 + 2] = 4\pi(\rho + 3p), \tag{2.25}$$

and

$$p_{,\mu} = (\rho + p + \sigma k)A_{,\mu}, \tag{2.26}$$

with

$$E_\mu = -k\dot v_\mu = -kA_{,\mu}, \tag{2.27}$$

$$B_\mu = \chi A_{,\mu}, \tag{2.28}$$

where χ is a function of A and a prime denotes differentiation with respect to A. Equations (2.23)–(2.26) constitute four equations for six unknowns $\rho, p, \sigma, R, \chi, A$. Thus, introducing an equation of state $p = p(\rho)$, one can solve for the unknowns, where the constant nonzero value of k is arbitrarily selected.

Case C: $\dot v_\mu = 0$ (geodesic motion), $S^\gamma = 0$, $E^\mu \neq 0$, $B^\mu \neq 0$.

In this case one readily derives the following set of equations:

$$\Box k + (f'/f - 1/k)k_{,\mu}k^{,\mu} = 0, \tag{2.29}$$

$$[f'/f + (1/k)(f^2 - 1)]k_{,\mu}k^{,\mu} = -4\pi\sigma, \tag{2.30}$$

$$[f^2(1 - 1/2k^2) + 1]k_{,\mu}k^{,\mu} = 4\pi(\rho + 3p), \tag{2.31}$$

and

$$p_{,\mu} = -\sigma k_{,\mu}, \tag{2.32}$$

with

$$E_\mu = k_{,\mu}, \tag{2.33}$$

$$B_\mu = f(k)k_{,\mu}. \tag{2.34}$$

Equations (2.29)–(2.32) constitute four equations with five unknowns ρ, σ, p, f, k. Thus one can introduce an equation of state $p = p(\rho)$ and solve for the unknowns.

Case D: $\dot v^\mu = 0$, $S^\gamma \neq 0$ (geodesic motion).

In this case one readily derives the following set of equations:

$$\Box \psi = (k_{,\mu}/k)\psi^{,\mu}, \tag{2.35}$$

$$k_{,\mu}k^{,\mu} + (1 - 1/2k^2)\psi_{,\mu}\psi^{,\mu} = 4\pi(\rho + 3p), \tag{2.36}$$

$$p_{,\mu} = -\sigma k_{,\mu}, \tag{2.37}$$

$$\rho + p + 2\sigma k = 0, \tag{2.38}$$

with

$$E_\mu = k_{,\mu}, \tag{2.39}$$

$$B_\mu = \psi_{,\mu}. \tag{2.40}$$

Equations (2.35)–(2.38) constitute four equations with five unknowns ρ, σ, p, k, ψ. Thus, introducing an equation of state $p = p(\rho)$ one can solve for the unknowns.

III. CONCLUSION

The main feature that emerges from our investigation is that the considerations of Bonnor[2] and Raychaudhuri in the case of charged dust can be easily generalized to the case of a charged perfect fluid as well. Explicit solutions to the various sets of equations obtained in Sec. II may be obtained relatively easily with the introduction of specific symmetries.

ACKNOWLEDGMENT

The authors are deeply indebted to Professor A. K. Raychaudhuri, Presidency College, Calcutta, India, for suggesting the problem and extending helpful comments.

APPENDIX

From the condition of rigid motion (2.1) we have $\underset{\xi^\alpha}{\mathscr{L}} v^\mu = 0$. (Henceforth $\underset{\xi^\alpha}{\mathscr{L}}$ will be abbreviated to $\mathscr{L}$.) Let us write

$$\Psi_\mu \equiv \mathscr{L}E_\mu = (\mathscr{L}F_{\mu\nu})v^\nu$$

and

$$\theta^{\mu} \equiv \mathscr{L} B^{\mu} = (\mathscr{L} \tfrac{1}{2} \eta^{\alpha\mu\nu\sigma} F_{\nu\sigma}) v_{\alpha}.$$

Thus

$$\Psi_{\mu} v^{\mu} = \theta_{\mu} v^{\mu} = 0.$$

Now

$$G^{\mu}{}_{\nu} = 8\pi[(\rho + p) v^{\mu} v_{\nu} - p\delta^{\mu}{}_{\nu} - (1/4\pi)(\tfrac{1}{2}\delta^{\mu}{}_{\nu} - v^{\mu} v_{\nu}) \times (E^2 + B^2) - (1/4\pi)(E^{\mu} E_{\nu} + B^{\mu} B_{\nu}) - (1/4\pi)(v^{\mu} S_{\nu} + v_{\nu} S^{\mu})]. \tag{A1}$$

Thus, $\mathscr{L} G^{\mu}{}_{\nu} = 0$ gives

$$\mathscr{L}(\rho + p) v^{\mu} v_{\nu} - \mathscr{L} p \delta^{\mu}{}_{\nu} - (1/4\pi)(\tfrac{1}{2}\delta^{\mu}{}_{\nu} - v^{\mu} v_{\nu}) \times \mathscr{L}(E^{\nu} + B^{\nu}) - (1/4\pi)(\Psi^{\mu} E_{\nu} + E^{\mu} \Psi_{\nu} + \theta^{\mu} B_{\nu} + B^{\mu} \theta_{\nu}) - (1/4\pi)(v^{\mu} \phi_{\nu} + v_{\nu} \phi^{\mu}) = 0, \tag{A2}$$

where

$$\phi^{\mu} \equiv \mathscr{L} S^{\mu}$$

and

$$\phi_{\mu} v^{\mu} = 0.$$

Contracting (A2) with v^{ν}, we have

$$\mathscr{L} \rho v^{\mu} + (1/8\pi) v^{\mu} \mathscr{L}(E^2 + B^2) - (1/4\pi)(\phi^{\mu}) = 0.$$

Contracting further with v_{μ}, we have

$$\mathscr{L} \rho + (1/8\pi) \mathscr{L}(E^2 + B^2) = 0 \tag{A3}$$

and

$$\phi^{\mu} = 0. \tag{A4}$$

Again $\mathscr{L} G^{\mu}{}_{;\mu} = 0$ gives

$$\mathscr{L}(\rho - 3p) = 0 \tag{A5}$$

when the Poynting vector is nonzero. Then, contracting Eq. (A2) with S_{ν}, we have

$$\mathscr{L} p + (1/8\pi) \mathscr{L}(E^2 + B^2) = 0. \tag{A6}$$

From (A3), (A5), and (A6) we have

$$\mathscr{L} \rho = \mathscr{L} p = \mathscr{L}(E^2 + B^2) = 0. \tag{A7}$$

We finally have

$$\Psi_{\mu} E_{\nu} + E_{\mu} \Psi_{\nu} + \theta_{\mu} B_{\nu} + B_{\mu} \theta_{\nu} = 0. \tag{A8}$$

Now if none of E^{μ}, B^{μ}, and S^{μ} vanish, they will span the three-space orthogonal to v^{μ}.

Since θ_{μ} and Ψ_{μ} are vectors in this three-space, we can write

$$\begin{aligned} \theta_{\mu} &= \alpha E_{\mu} + \beta B_{\mu} + \gamma S_{\mu}, \\ \Psi_{\mu} &= \alpha' E_{\mu} + \beta' B_{\mu} + \gamma' S_{\mu}. \end{aligned} \tag{A9}$$

The condition $\phi^{\mu} = 0$ gives

$$\eta^{\nu\mu\rho\sigma}(E_{\mu} \theta_{\rho} + \Psi_{\mu} B_{\rho}) v_{\sigma} = 0. \tag{A10}$$

[This is because $S^{\nu} \equiv \eta^{\nu\mu\rho\sigma} E_{\mu} B_{\rho} v_{\sigma}$.] Using (A9) and (A10), we have

$$\eta^{\nu\mu\sigma\rho}[\beta E_{\mu} B_{\sigma} + \gamma E_{\mu} S_{\sigma} + \alpha' E_{\mu} B_{\sigma} + \gamma' S_{\mu} B_{\sigma}] v_{\rho} = 0,$$

or

$$(\alpha' + \beta) S^{\nu} + \eta^{\nu\mu\rho\sigma}(\gamma E_{\mu} S_{\rho} + \gamma' S_{\mu} B_{\rho}) v_{\sigma} = 0.$$

While contraction with E_{ν} gives

$$\gamma' = 0,$$

contraction with B_{μ} gives

$$\gamma = 0$$

and so

$$\alpha' + \beta = 0.$$

Equation (A9) now gives

$$\begin{aligned} \theta_{\mu} &= \alpha E_{\mu} + \beta B_{\mu}, \\ \Psi_{\mu} &= -\beta E_{\mu} + \beta' B_{\mu}. \end{aligned} \tag{A11}$$

Using (A8) and (A11) and contracting with E_{μ} we have

$$[2\beta E^2 + (\alpha + \beta') E_{\mu} B^{\mu}] E_{\nu} + [2\beta E_{\mu} B^{\mu} - (\alpha + \beta') E^2] B_2 = 0.$$

As $S^{\nu} \neq 0$, we have

$$2\beta E^2 + (\alpha + \beta') E_{\mu} B^{\mu} = 0,$$

$$2\beta B_{\mu} E^{\mu} - (\alpha + \beta') E^2 = 0,$$

i.e., $\beta = 0$, or $\alpha' + \beta = 0$. Thus, from (A11),

$$\begin{aligned} \mathscr{L} E_{\mu} &= -\alpha B_{\mu}, \\ \mathscr{L} B_{\mu} &= \alpha E_{\mu}. \end{aligned} \tag{A12}$$

Now from Eqs. (2.13) and (A7), we have

$$\mathscr{L}(\sigma E^{\nu}) = 0 \quad \text{or} \quad \mathscr{L} E^{\nu} = (\mathscr{L}\sigma/\sigma) E^{\nu}. \tag{A13}$$

As $S^{\nu} \neq 0$, so (A12) and (A13) give $\alpha = 0$, or

$$\mathscr{L} E^{\mu} = \mathscr{L} B^{\mu} = 0 \quad \text{and} \quad \mathscr{L}\sigma = 0. \tag{A14}$$

Since we have assumed the relation $A^{\mu} = k v^{\mu}$ to hold in the Lorentz gauge it follows that

$$\mathscr{L} k = \mathscr{L} A_{\mu} = 0.$$

The proof of the above result, for $S^{\mu} = 0$, in the case of charged dust is rather trivial. But, in the case of a perfect fluid with nonvanishing pressure, the above conclusions cannot be proved unless some equation of state is assumed.

[1]A. K. Raychaudhuri, J. Phys. A: Math. Nucl. Gen. **15**, 831 (1982).
[2]W. B. Bonnor, J. Phys. A: Math. Gen. **13**, 3465 (1980).

Plane-fronted gravitational and electromagnetic waves in spaces with cosmological constant

István Ozsváth, Ivor Robinson, and Krzysztof Rózga[a]
Program in Mathematical Sciences, The University of Texas at Dallas, Richardson, Texas 75080

(Received 4 January 1984; accepted for publication 27 July 1984)

Plane-fronted waves in spaces with nonzero cosmological constant are studied. In particular their complete classification, which depends essentially on the sign of the cosmological constant and that of some second-order invariant determined by the congruence of null rays, is provided.

I. INTRODUCTION

Plane waves are such useful objects in many problems of everyday electromagnetism that it is natural to ask whether any similar solutions exist in electromagnetic theory against a cosmological background. Schrödinger[1] considered the problem long ago, and concluded that no such solutions exist for Robertson–Walker universes in general; and more specifically, that there were none in de Sitter space. Without detailed investigation, one might reasonably expect similar results to hold for gravitational waves of the linear approximation, and *a fortiori* for exact solutions representing gravitational plane waves.

In anti-de Sitter space, however, we seem to encounter a very different situation. In the course of investigation of homogeneous spaces, one of us[2] obtained an Einstein space which is null and invariant under a five-parameter group of motions—properties which one would normally regard as characteristic of a plane wave. Apart from the cosomological constant, the solution contains a single parameter, which specifies the amplitude of the wave. When this parameter is zero, the solution reduces to anti-de Sitter space.

The result depends crucially on the sign of the cosmological constant: there is no corresponding family of solutions for de Sitter space. All this is consistent with Schrödinger's observation, but it is in some ways rather puzzling. In order to explore the problem systematically, we turn to the general question of gravitational and electromagnetic waves with plane wave fronts.

In formal terms, our assumption is that the gravitational and electromagnetic fields are subject to the radiation conditions[3]

$$S_{ab[cd}k_{m]} = 0 \quad \text{and} \quad k^m S_{mbcd} = 0, \tag{1.1}$$

where

$$S_{abcd} = R_{abcd} + (R/6) g_{a[c} g_{|b|d]},$$

with $R = -4\Lambda = \text{const}$, and

$$F_{[ab}k_{c]} = 0 \quad \text{and} \quad k^m F_{mc} = 0, \tag{1.2}$$

where the vector field k_a (the propagation vector) is null, geodesic, shear-free, expansion-free, and rotation-free (characteristics of a plane-fronted wave).

In Sec. II we discuss some properties of that vector. In particular an almost-Killing normalization of k_a is defined and an important, second-order invariant is pointed out.

In the next two sections, a geometrical coordinate system is introduced, the radiation conditions are integrated, and, for the purely gravitational case, the general solution of the only vacuum field equation is found. In Sec. V canonical forms of the metric tensor are discussed.

In Sec. VI some specializations are made to obtain homogeneous solutions of Einstein and Einstein–Maxwell equations, first discovered in Refs. 2 and 4, respectively.

Section VII deals with a nonhomogeneous field equation. Its particular solution is represented in terms of double integrals. There are provided also explicit forms of particular solutions corresponding to some simple forms of a profile of an electromagnetic wave.

In the last section we compare our results with those obtained by García Díaz and Plebański.[5]

II. THE PROPAGATION VECTOR

A null vector field k^a is geodesic, shear-free, expansion-free, and twist-free if and only if there exist u_a and v_a such that

$$k_{a;b} = u_a k_b + k_a v_b. \tag{2.1}$$

This equation implies that

$$k_{[a}k_{b,c]} = 0, \quad k_{[a}u_{b,c]} = k_{[a}v_{b,c]}, \tag{2.2}$$

and it is invariant under the transformation

$$u_a \to u_a + \chi k_a, \quad v_a \to v_a - \chi k_a, \tag{2.3}$$

with χ arbitrary: without loss of generality, therefore, we may impose the supplementary condition

$$u_{[a,b]} = v_{[a,b]}. \tag{2.4}$$

It then follows from the Ricci identities that

$$k_p R^p{}_{abc} = 2k_a u_{[bc]} - 2u_{a[b}k_{c]}, \tag{2.5}$$

where $u_{ab} := u_{a;b} - u_a u_b$; and hence

$$k_p R^p{}_{a[bc}k_{d]} = 2k_a k_{[b}u_{c,d]}. \tag{2.6}$$

Considering, moreover, that k^a is null, one sees that

$$k^a k_a = k^a u_a = 0, \quad k^a u_{a;b} = -u^a u_a k_b \tag{2.7}$$

and

$$k^a u^b R_{ab} = -\tfrac{1}{2}(u^a u_a)_{,b} k^b; \tag{2.8}$$

the convention is that $R_{ab} = R^c{}_{abc}$.

We now turn to the radiation conditions. From either of Eqs. (1.1),

[a] Permanent address: Rewolucji Pazdziernikowej 157/159, 25-547 Kielce, Poland.

 0022-2488/85/071755-07$02.50

$$k_p R^p{}_{a[bc}k_{d]} = 0. \tag{2.9}$$

Contracting with $g^{ab}u^c$ and using (2.8), we obtain

$$(u^a u_a)_{,b} k^b = 0. \tag{2.10}$$

From (2.2) and (2.5) it follows then that (2.9) is equivalent to

$$k_{[a} u_{b,c]} + k_{[a} v_{b,c]} = 0, \tag{2.11}$$

which is a necessary and sufficient condition for the existence of scalars L and w such that

$$u_a + v_a = Lk_a - w_{,a}. \tag{2.12}$$

We can remove w from the first equation by rescaling the propagation vector

$$k_a \to e^w k_a. \tag{2.13}$$

By this means, we obtain a sharpened version of (2.1),

$$k_{a;b} = u_a k_b - k_a u_b + Lk_a k_b. \tag{2.14}$$

The normalization adopted here is a special case of affine normalization, since the vector

$$k'_a := k_{a;p} k^p \tag{2.15}$$

vanishes on account of (2.14).

It is conserved under transformations (2.13) where dw is proportional to the propagation vector: then

$$Lk_a \to Lk_a + w_{,a}. \tag{2.16}$$

With respect to an arbitrary vector field k^a, the Lie derivative of a tensor c_{ab} is

$$\mathscr{L}_k c_{ab} = c_{pb} k^p{}_{;a} + c_{ap} k^p{}_{;b} + c_{ab;p} k^p, \tag{2.17}$$

hence

$$\mathscr{L}_k g_{ab} = 2k_{(a;b)} \tag{2.18}$$

and

$$\tfrac{1}{2} \mathscr{L}_k^2 g_{ab} - k'_{(a;b)} = k^p{}_{;a} k_{p;b} + k^p k^q R_{pabq}. \tag{2.19}$$

Now employing (2.1) we obtain

$$\tfrac{1}{2} \mathscr{L}_k^2 g_{ab} - k'_{(a;b)} = (u^p u_p) k_a k_b + k^p k^q R_{pabq}. \tag{2.20}$$

Moreover (1.1) implies

$$k^p k^q R_{pabq} = (\Lambda/3) k_a k_b. \tag{2.21}$$

Hence, finally,

$$\tfrac{1}{2} \mathscr{L}_k^2 g_{ab} - k'_{(a;b)} = L' k_a k_b,$$

where

$$L' = u^a u_a + \Lambda/3. \tag{2.22}$$

The scalar L' is invariant under renormalization of the propagation vector.

If k_a is normalized in such a way that (2.14) holds, then

$$\tfrac{1}{2} \mathscr{L}_k g_{ab} = Lk_a k_b. \tag{2.23}$$

That normalization is further referred to as an almost-Killing normalization.

In the almost-Killing normalization we have

$$L' = \mathscr{L}_k L, \tag{2.24}$$

and because of (2.10) and (2.22),

$$\mathscr{L}_k^2 L = 0. \tag{2.25}$$

III. LINE ELEMENT ADMITTING PLANE-FRONTED WAVE

Let k^a be a propagation vector of a plane-fronted wave in its almost-Killing normalization (2.23).

There exists a coordinate system $\{\rho, \sigma, \zeta, \bar{\zeta}\}$ (bar denotes complex conjugation) subject to the following constraints: (i) the ρ-coordinate is an affine parameter along geodesic lines tangent to k^a, (ii) $k_a\, dx^a \wedge d\sigma = 0$, and (iii) $d\zeta$ and $d\bar{\zeta}$ are null one-forms. Then the metric tensor can be written conveniently in the form

$$ds^2 = -2q^2 p^{-2}\, d\sigma(\bar{Z}\, d\zeta + Z\, d\bar{\zeta} + S\, d\sigma + d\rho) + 2p^{-2}\, d\zeta\, d\bar{\zeta}, \tag{3.1}$$

where S, p, and q are real and Z, $\bar{Z}$ are complex conjugates. It follows also by remarks in Sec. II that

$$k^a \frac{\partial}{\partial x^a} = \Psi(\sigma) \frac{\partial}{\partial \rho}. \tag{3.2}$$

Now, employing (2.23) and (2.25) we infer that

$$Z_\rho = 0 = p_\rho = q_\rho \tag{3.3}$$

and

$$S = -\tfrac{1}{2}\kappa\rho^2 + l\rho + \epsilon, \tag{3.4}$$

where κ, l, and ϵ do not depend on ρ.

Moreover, the invariant L' (2.24) is

$$L' = \kappa p^2/q^2. \tag{3.5}$$

A convenient feature of our coordinates is that we can write down immediately the general expression for an electromagnetic wave having the surfaces of constant σ for its wave fronts [(1.2) and (ii)]:

$$F = \tfrac{1}{2} F_{ab}\, dx^a \wedge dx^b = f(\zeta, \sigma) d\zeta \wedge d\sigma + \bar{f}(\bar{\zeta}, \sigma) d\bar{\zeta} \wedge d\sigma. \tag{3.6}$$

The energy tensor is given by

$$T_{ab} = f\bar{f} p^2 \sigma_{,a} \sigma_{,b}. \tag{3.7}$$

Finally we list coordinate transformations preserving the conditions (i)–(iii) and (3.2):

$$\zeta' = h(\zeta, \sigma), \tag{3.8}$$

$$\rho' = a(\sigma)\rho + b(\zeta, \bar{\zeta}, \sigma), \tag{3.9}$$

$$\sigma' = g(\sigma), \tag{3.10}$$

where $h_\zeta \neq 0$ and $a g_\sigma > 0$. The remaining consequences of radiation conditions are discussed in the next section.

IV. RADIATION CONDITIONS AND THE FIELD EQUATION

We use the null-tetrad formalism.[5,6] The metric tensor is represented in the form

$$ds^2 = 2e^1 e^2 + 2e^3 e^4, \tag{4.1}$$

where

$$e^1 = p^{-2}\, d\zeta, \quad e^2 = p^{-2}\, d\bar{\zeta}, \quad e^3 = -d\sigma,$$
$$e^4 = q^2 p^{-2} (\bar{Z}\, d\zeta + Z\, d\bar{\zeta} + S\, d\sigma + d\rho). \tag{4.2}$$

The radiation conditions are then equivalent to

$$C^{(5)} = C^{(4)} = C^{(3)} = C^{(2)} = 0, \tag{4.3}$$

$$C_{ab} \sim \delta_a^3 \delta_b^3, \tag{4.4}$$

$$R = -4\Lambda, \tag{4.5}$$

where the $C^{(i)}$, $i = 1,2,3,4,5$, are components of the self-dual part of the conformal curvature tensor, C_{ab} is the trace-free Ricci tensor, and R is the Ricci scalar.[6]

Those components can be calculated from the second structural equations.[6] It is convenient to incorporate into them what is already known to be a result of radiation conditions, (3.3). Then we have

$$C^{(5)} = C_{44} = 0 = C^{(4)} = C_{42}, \tag{4.6}$$

$$C^{(3)} - \frac{R}{6} = -2p\left\{\frac{p^2}{q}\left(\frac{q}{p}\right)_{\bar\zeta}\right\}_{\zeta} + 2\frac{p^2}{q}\left(\frac{q}{p}\right)_{\bar\zeta}\left\{p_\zeta - \frac{p^2}{q}\left(\frac{q}{p}\right)_{\zeta}\right\}, \tag{4.7}$$

$$C_{22} = 2p\left\{\frac{p^2}{q}\left(\frac{q}{p}\right)_{\bar\zeta}\right\}_{\bar\zeta} + 2\frac{p^2}{q}\left(\frac{q}{p}\right)_{\bar\zeta}\left\{p_{\bar\zeta} + \frac{p^2}{q}\left(\frac{q}{p}\right)_{\bar\zeta}\right\}, \tag{4.8}$$

$$C^{(3)} + \frac{R}{12} = \frac{1}{2}p^2q^{-2}S_{\rho\rho} - p^2 \times (\ln p)_{\zeta\bar\zeta} + p^4q^{-2}\left(\frac{q}{p}\right)_{\zeta}\left(\frac{q}{p}\right)_{\bar\zeta}, \tag{4.9}$$

$$C^{(2)} = p\left\{-p^{-1}p_\zeta + \frac{p}{q}\left(\frac{q}{p}\right)_{\zeta}\right\}_{\sigma} + p\left\{\frac{1}{2}q^2(Z_\zeta - \bar Z_{\bar\zeta}) + S_\rho - 2\frac{p}{q}\left(\frac{q}{p}\right)_{\sigma}\right\}_{\zeta} - p\bar Z S_{\rho\rho}, \tag{4.10}$$

$$C_{12} = -\frac{1}{2}p^2q^{-2}S_{\rho\rho} - p^2(\ln p)_{\zeta\bar\zeta} - p^4q^{-2}\left(\frac{q}{p}\right)_{\zeta}\left(\frac{q}{p}\right)_{\bar\zeta}, \tag{4.11}$$

$$\begin{aligned} C_{32} = {} & p\left\{p^{-1}p_{\bar\zeta} + \frac{p}{q} - \left(\frac{q}{p}\right)_{\bar\zeta}\right\}_{\sigma} + p\left\{\frac{1}{2}q^2(Z_\zeta - \bar Z_{\bar\zeta}) + S_\rho \right. \\ & \left. - 2\frac{p}{q}\left(\frac{q}{p}\right)_{\sigma}\right\}_{\bar\zeta} - pZS_{\rho\rho} + 2\frac{p^2}{q}\left(\frac{q}{p}\right)_{\bar\zeta} \\ & \times\left\{\frac{1}{2}q^2(Z_\zeta - \bar Z_{\bar\zeta}) - p^{-1}p_\sigma\right\}. \end{aligned} \tag{4.12}$$

Now, assuming (4.3)–(4.5) to be satisfied, we substitute $S_{\rho\rho}$ from (4.11) into (4.9), and we infer that

$$2(pp_{\zeta\bar\zeta} - p_\zeta p_{\bar\zeta}) = \Lambda/3. \tag{4.13}$$

Next (4.7) and (4.8) imply

$$qp_{\zeta\bar\zeta} + pq_{\zeta\bar\zeta} = p_\zeta q_{\bar\zeta} + p_{\bar\zeta}q_\zeta \tag{4.14}$$

and

$$pq_{\zeta\zeta} - qp_{\zeta\zeta} = 0. \tag{4.15}$$

We observe also that by combining the complex conjugate of (4.10) with (4.12) one can derive

$$p^2[(q^3/p)(Z_\zeta - \bar Z_{\bar\zeta})]_{\bar\zeta} + 2qp_{\bar\zeta\sigma} - 2p_\sigma q_{\bar\zeta} = 0. \tag{4.16}$$

Now the form of p can be specialized conveniently. Indeed, a differentiation of both sides of (4.13) with respect to ζ leads to the conclusion that $p_{\zeta\zeta} = G(\zeta,\sigma)p$, where G is some function of ζ and σ only. Moreover, G can be transformed to a zero function by means of a transformation (3.8)–(3.10) in the form of $\zeta = h(\zeta',\sigma')$, $\rho = \rho'$, $\sigma = \sigma'$, with h being a solution of

$$h_{\zeta'\zeta'\zeta'} = \tfrac{3}{2}h_{\zeta'}^{-1}h_{\zeta'\zeta'}^2 + 2G(h(\zeta',\sigma'),\sigma')h_{\zeta'}^3. \tag{4.17}$$

After that, p is a linear function of ζ and $\bar\zeta$. Employing that transformation again [this time h is restricted by (4.17) with $G \equiv 0$] we reduce p to

$$p = 1 + (\Lambda/6)\zeta\bar\zeta. \tag{4.18}$$

In doing this (4.13) is essential. There has been employed also the fact that a sign of p does not affect the form of ds^2 (3.1). It can be shown that transformations preserving that shape of p are given by

$$\zeta = (\zeta' A + \bar B)/[-(\Lambda/6)\zeta' B + \bar A], \tag{4.19}$$

where A and B are functions of σ only, such that

$$A\bar A + (\Lambda/6)B\bar B = 1, \tag{4.20}$$

(3.9), and (3.10). Under these (4.19) transformations the structural functions p and q transform as

$$\begin{aligned} p' &= [-(\Lambda/6)\zeta' B + \bar A]^2 p, \\ q' &= [-(\Lambda/6)\zeta' B + \bar A]^2 q, \end{aligned} \tag{4.21}$$

with

$$\begin{aligned} \alpha' &= (A\bar A - (\Lambda/6)B\bar B)\alpha = \beta\bar A B + \bar\beta A\bar B, \\ \beta' &= (\Lambda/3)\alpha\bar A\bar B + \beta\bar A^2 - (\Lambda/6)\bar\beta\bar B^2. \end{aligned} \tag{4.22}$$

There are two important consequences of that specialization of p. Equations (4.14) and (4.15) can be easily solved for q:

$$q = (1 - (\Lambda/6)\zeta\bar\zeta)\alpha + \zeta\bar\beta + \bar\zeta\beta, \tag{4.23}$$

where α is a real and β a complex function of σ. The equation (4.16) implies that

$$[(q^3/p)(Z_\zeta - \bar Z_{\bar\zeta})]_{\bar\zeta} = 0. \tag{4.24}$$

It is convenient to denote

$$\omega := (q^3/p)(\bar Z_{\bar\zeta} - Z_\zeta). \tag{4.25}$$

Next we find the transformation law of ω under a transformation (4.19) ($\sigma = \sigma'$ and $\rho = \rho'$):

$$\omega' = \omega + \alpha\left(\frac{A_\sigma}{A} - \frac{\bar A_\sigma}{\bar A}\right) + \frac{\bar\beta\bar B A_\sigma}{A\bar A} - \frac{\beta B\bar A_\sigma}{A\bar A} + \frac{\beta B_\sigma}{A} - \frac{\bar\beta\bar B_\sigma}{\bar A} + \frac{\Lambda}{6}\frac{\alpha}{A\bar A}(B\bar B_\sigma - \bar B B_\sigma). \tag{4.26}$$

Therefore by a proper choice of A and B, ω can be reduced to zero, due to (4.24). In this new coordinate system we have

$$Z_\zeta - \bar Z_{\bar\zeta} = 0, \tag{4.27}$$

which means that Z itself can be reduced to zero, by means of a transformation (3.8)–(3.10) with $\zeta' = \zeta$ and $\sigma' = \sigma$.

Now with $Z \equiv 0$ one has three equations on S:

$$S_{\rho\rho} + 2p^2(q/p)_{\bar\zeta}(q/p)_\zeta + 2q^2(\ln p)_{\zeta\bar\zeta} = 0, \tag{4.28}$$

$$S_{\rho\bar\zeta} = (q_\sigma/q)_{\bar\zeta}, \tag{4.29}$$

and its complex conjugate.

Then substituting S from (3.4) into (4.28) one obtains

$$\kappa = \tfrac{1}{3}\Lambda\alpha^2 + 2\beta\bar\beta. \tag{4.30}$$

Equation (4.29) leaves some ambiguity in l, which, however, can be removed by a transformation: $\zeta' = \zeta$, $\sigma' = g(\sigma)$,

$\rho' = \rho$ (or $\rho' = -\rho$), so that finally

$$l = \frac{\partial}{\partial\sigma}(\ln|q|). \tag{4.31}$$

Further it is convenient to write

$$H(\sigma, \zeta, \bar{\zeta}) := \tfrac{1}{2}(q/p)\epsilon. \tag{4.32}$$

The remaining components of the curvature tensor C_{33} and $C^{(1)}$ can be calculated now. One obtains

$$C_{33} = -pq\,(H_{\zeta\bar{\zeta}} + (\Lambda/3)p^{-2}H) \tag{4.33}$$

and

$$C^{(1)} = [q^2((p/q)H)_{\zeta}]_{\zeta}.$$

For a purely gravitational wave, therefore, the empty space equation is

$$H_{\zeta\bar{\zeta}} + (\Lambda/3)p^{-2}H = 0, \tag{4.34}$$

and for a combined gravitational and electromagnetic wave,

$$H_{\zeta\bar{\zeta}} + (\Lambda/3)p^{-2}H = f\bar{f}p/q. \tag{4.35}$$

The final form of the metric structure is

$$ds^2 = ds_1{}^2 - p^{-1}q\,H\,d\sigma^2, \tag{4.36}$$

where

$$ds_1{}^2 = -2q^2p^{-2}\,d\sigma[dp + (-\tfrac{1}{2}\kappa p + l)\,p\,d\sigma] + 2p^{-2}\,d\zeta\,d\bar{\zeta},$$

with p,q,κ, and l given by (4.18), (4.23), (4.30), and (4.31), respectively.

We notice two simple facts concerning that metric structure. It is conformally flat ($C^{(1)} = 0$) iff

$$H = p^{-1}(u + \bar{v}\zeta + v\bar{\zeta} + w\zeta\bar{\zeta}), \tag{4.37}$$

where u, w, and v are arbitrary, u and w real, functions of σ. In particular, the metric is of a constant curvature iff $w = -(\Lambda/6)\,u$, i.e.,

$$H = p^{-1}[u(1 - (\Lambda/6)\zeta\bar{\zeta}) + \bar{v}\zeta + v\bar{\zeta}]. \tag{4.38}$$

Now, let $\Phi = \Phi(\zeta, \sigma)$ be an arbitrary function of ζ and σ (holomorphic in ζ). Then one verifies that the combination $\Phi_{\zeta} - (\Lambda/3)(\bar{\zeta}/p)\Phi$ is a complex solution of (4.34). Hence,

$$H = \Phi_{\zeta} - \frac{\Lambda}{3}\frac{\bar{\zeta}}{p}\Phi + \bar{\Phi}_{\bar{\zeta}} - \frac{\Lambda}{3}\frac{\zeta}{p}\bar{\Phi} \tag{4.39}$$

is its real solution. Moreover, one can prove that this is the general form of real solutions of (4.34), as one could expect, since it depends on one complex, arbitrary function Φ.

Solutions of (4.35) are discussed in Sec. VII.

V. CANONICAL FORMS OF THE METRIC STRUCTURE

By a proper choice of the coordinate system [transformations (3.8)–(3.10)] the whole ambiguity in the metric structure (4.36) can be incorporated into the H-term. (The argument does not depend on the field equation at all.) There are four cases to be discussed independently. A distinction between them is provided by the sign of the invariant L' (3.5) and that of the cosmological constant Λ.

As a consequence of (3.5) and (4.30) we have to consider

I : $\Lambda > 0,\quad \kappa > 0,$
II : $\Lambda < 0,\quad \kappa > 0,$
III : $\Lambda < 0,\quad \kappa < 0,$
IV : $\Lambda < 0,\quad \kappa = 0.$

We remark that metrics with $\Lambda = 0$ are divided into two classes: R-waves ($\kappa = 0$) and Kundt's waves ($\kappa > 0$) (See Refs. 7–10). If $\Lambda \neq 0$, one can introduce a similar distinction, which, however, turns out to be essential for $\Lambda < 0$ only. To make that point clear, let $R(\Lambda,\alpha,\beta)$ denote the class of metrics of the form (4.36) with Λ, α, and β being fixed. One can prove that for $\Lambda \neq 0$ and $\kappa \neq 0$ it suffices to consider two classes only: generalized R-waves, $R(\Lambda) := R(\Lambda, 1, 0)$; and generalized Kundt's waves, $K(\Lambda) := R(\Lambda, 0, 1)$. Indeed, one can show that for $\Lambda > 0$ (case I),

$$R(\Lambda^+, \alpha, \beta) = R(\Lambda^+) = K(\Lambda^+),$$

for $\Lambda < 0$, however, it turns out that

$$R(\Lambda^-, \alpha, \beta) = K(\Lambda^-), \quad \text{when } \kappa > 0 \text{ (case II)}$$

and

$$R(\Lambda^-, \alpha, \beta) = R(\Lambda^-), \quad \text{when } \kappa < 0 \text{ (case III)}.$$

Obviously, the generalized waves $R(\Lambda^-)$ and $K(\Lambda^-)$ are distinct since the sign of κ is an invariant.

Within the case IV, there are two families of metrics to be considered independently. The first one $(\mathrm{IV})_0$ consists of metrics admitting $\partial/\partial p$ as their Killing vector. The remaining metrics form the second family $(\mathrm{IV})_1$.

One can show the existence of a coordinate system, in which the metrics from the family $(\mathrm{IV})_0$ are of the form (4.36) with

$$q = (1 + \lambda\zeta)(1 + \lambda\bar{\zeta}), \quad \lambda = \sqrt{-\Lambda/6}. \tag{5.1}$$

For the metrics from the family $(\mathrm{IV})_1$, the function q can be reduced to

$$q = (1 + \lambda\zeta e^{i\lambda\sigma})(1 + \lambda\bar{\zeta}e^{-i\lambda\sigma}). \tag{5.2}$$

VI. SPECIAL, COMBINED ELECTROMAGNETIC, AND GRAVIATIONAL PLANE WAVES WITH SYMMETRIES

We consider metrics from the family $(\mathrm{IV})_0$, which admit a three-parameter group of motions being at the same time a symmetry group of the electromagnetic field, which acts along wave fronts: hypersurfaces $\sigma = \text{const}$. We assume also, that it is a subgroup of the ten-parameter anti-de Sitter group, i.e., the group of motions of the metric

$$ds_\Lambda^2 = -2q^2p^{-2}dp\,d\sigma + 2p^{-2}\,d\zeta\,d\bar{\zeta}, \tag{6.1}$$

where q is given by (5.1).

To specify that subgroup in terms of its generators, we find at first, a maximal, four-parameter subgroup of motions of (6.1), acting along hypersurfaces $\sigma = \text{const}$. Its generators are

$$X_1 = \frac{\partial}{\partial p}, \tag{6.2}$$

$$X_2 = i(1 + \lambda\zeta)^2\frac{\partial}{\partial\zeta} - i(1 + \lambda\bar{\zeta})^2\frac{\partial}{\partial\bar{\zeta}}, \tag{6.3}$$

$$X_3 = \frac{i(\bar{\zeta} - \zeta)}{(1 + \lambda\zeta)(1 + \lambda\bar{\zeta})}X_1 + \sigma X_2, \tag{6.4}$$

and

$$Y = -4\lambda\rho\frac{\partial}{\partial\rho} + (1-\lambda^2\zeta^2)\frac{\partial}{\partial\zeta} + (1-\lambda^2\bar{\zeta}^2)\frac{\partial}{\partial\bar{\zeta}}. \tag{6.5}$$

We find also the commutation relations

$$[X_1, X_2] = 0, \quad [X_1, X_3] = 0, \quad [X_2, X_3] = 2X_1,$$
$$[X_1, Y] = -4\lambda X_1, \quad [X_2, Y] = -2\lambda X_2,$$
$$[X_3, Y] = -2\lambda X_3. \tag{6.6}$$

A subgroup to be considered further is that generated by X_1, X_2, and X_3. Thus we require that

$$\underset{X_\alpha}{\mathscr{L}} ds^2 = 0, \tag{6.7}$$

and

$$\underset{X_\alpha}{\mathscr{L}} F = 0, \quad \text{for} \quad \alpha = 1,2,3. \tag{6.8}$$

Then we find that (6.7) is equivalent to

$$(1+\lambda\zeta)^2 H_\zeta - (1+\lambda\bar{\zeta})^2 H_{\bar{\zeta}} = 0, \tag{6.9}$$

while (6.8) implies that

$$f(\zeta,\sigma) = \phi(\sigma)/(1+\lambda\zeta)^2. \tag{6.10}$$

Equation (6.9) can be solved immediatley, and the result is that H depends on ζ and $\bar{\zeta}$ through a function s,

$$s := p/q. \tag{6.11}$$

Then the field equation (4.33) takes the form of

$$s^2 H_{ss} - 2H = (\phi\bar{\phi}/\lambda^2)s^3. \tag{6.12}$$

Its general solution is

$$H(\sigma, s) = U(\sigma)s^2 + V(\sigma)s^{-1} + (\phi\bar{\phi}/4\lambda^2)s^3, \tag{6.13}$$

and therefore

$$H(\sigma,\zeta,\bar{\zeta}) = \frac{U(\sigma)p^2}{q^2} + \frac{V(\sigma)q}{p} + \frac{\phi\bar{\phi}}{4\lambda^2}\frac{p^3}{q^3}. \tag{6.14}$$

A function $V(\sigma)$ can be eliminated by the coordinate transformation $\rho = \rho' - b(\sigma)$, $\sigma = \sigma'$, $\zeta = \zeta'$, where $b = V/2$. Thus without lost generality we may put $V \equiv 0$ and

$$H = \frac{Up^2}{q^2} + \frac{\phi\bar{\phi}}{4\lambda^2}\frac{p^3}{q^3}. \tag{6.15}$$

The functions H and f given by (6.15) and (6.10) determine gravitational and electromagnetic waves with a three-parameter group of symmetries built into them. It is interesting to observe the correspondence between this result and those obtained by one of us in Refs. 2 and 4. For this purpose we make the coordinate transformation $\zeta \to (2/\lambda) \times (\sqrt{2\lambda\zeta} + 1) - (1/\lambda)$, and next we define $x := \frac{1}{2}(\zeta + \bar{\zeta})$ and $y := (1/2i)(\zeta - \bar{\zeta})$.

Then

$$ds^2 = -\frac{1}{\lambda^2 x^2}d\rho\, d\sigma + \frac{1}{2\lambda^2 x^2}(dx^2 + dy^2)$$
$$-\sqrt{2}\lambda Ux\, d\sigma^2 - \frac{\phi\bar{\phi}}{2}x^2\, d\sigma^2. \tag{6.16}$$

Now for $U = \text{const}$ and $\Phi = 0$ we perform the coordinate transformation $x \to (1/\lambda\sqrt{2})e^{-\lambda\sqrt{2}x}$, which results in

$$ds^2 = -2e^{2\sqrt{2}\lambda x}d\rho\, d\sigma + e^{2\sqrt{2}\lambda x}dy^2$$
$$+ dx^2 - Ue^{-\sqrt{2}\lambda x}d\sigma^2. \tag{6.17}$$

This metric structure is easily recognizable as that of a homogeneous solution of Einstein equations, of Petrov's type N, with $\Lambda \neq 0$ (see Ref. 2).

Now, let $U \equiv 0$ and ϕ be such that $\phi\bar{\phi} = \text{const}$. Then the transformation $x \to e^x$ leads to

$$ds^2 = -\frac{e^{-2x}}{\lambda^2}d\rho\, d\sigma$$
$$+ \frac{1}{2\lambda^2}(dx^2 + e^{-2x}dy^2) - \frac{\phi\bar{\phi}}{2}e^{2x}d\sigma^2. \tag{6.18}$$

We identify this metric as a homogeneous solution of Einstein–Maxwell's equations with a null electromagnetic tensor and $\Lambda \neq 0$ (see Ref. 4).

We remark that if the parameter U in (6.17) and $\phi\bar{\phi}$ in (6.18) is nonzero, it can be reduced to a value prescribed in advance by a coordinate transformation $\sigma \to c^{-1}\sigma$, $\rho \to c\rho$, where $c = \text{const} \neq 0$. And so, for instance, one can impose on $\phi\bar{\phi}$ the condition $\phi\bar{\phi} = 1/\lambda^2$ (see Ref. 4).

In both cases, (6.17) and (6.18), five-parameter groups of motions are admitted. In addition to symmetries generated by X_1, X_2, X_3 [(6.2)–(6.4)], there are new ones, generated by $X_4 = \partial/\partial\sigma$ and

$$X_5 = \lambda\sigma\frac{\partial}{\partial\sigma} - 5\lambda\rho\frac{\partial}{\partial\rho} + (1-\lambda^2\zeta^2)\frac{\partial}{\partial\zeta}$$
$$+ (1-\lambda^2\bar{\zeta}^2)\frac{\partial}{\partial\bar{\zeta}},$$
$$\text{for (6.17)}, \tag{6.19}$$

and

$$X_5' = 2\lambda\sigma\frac{\partial}{\partial\sigma} - 6\lambda\rho\frac{\partial}{\partial\rho} + (1-\lambda^2\zeta^2)\frac{\partial}{\partial\zeta}$$
$$+ (1-\lambda^2\bar{\zeta}^2)\frac{\partial}{\partial\bar{\zeta}}, \tag{6.20}$$
$$\text{for (6.18)}.$$

(The expressions for X_5 and X_5' are given in the original coordinates $\sigma, \rho, \zeta, \bar{\zeta}$.) It is not difficult to show also that X_5 and X_5' generate symmetries of the anti-de Sitter space (6.1) as well.

In the conclusion of this section we remark the possibility of a limit transition $\lambda \to 0$. It cannot be performed on H in the form of (6.14) with U and V being λ-independent. However, the substitution $U \to U - \phi\bar{\phi}/3\lambda^2$ and $V \to V + \phi\bar{\phi}/12\lambda^2$ makes that process meaningful.

VII. THE NONHOMOGENEOUS FIELD EQUATION

In this section we construct a particular solution of the field equation (4.34), and therefore, because of the results of Sec. IV, its general solution as well.

We look for H_{part} in the form of

$$H_{\text{part}} = \mu_\zeta - \frac{\Lambda}{3}\frac{\bar{\zeta}}{p}\mu + \bar{\mu}_{\bar{\zeta}} - \frac{\Lambda}{3}\frac{\zeta\bar{\mu}}{p} \tag{7.1}$$

[compare (4.37)], where μ depends on both ζ and $\bar{\zeta}$, such that the function

$$H_1 := \mu_\zeta - \frac{\Lambda}{3}\frac{\bar{\zeta}}{p}\mu \tag{7.2}$$

satisfies the equation

$$H_{1,\zeta\bar\zeta} + \frac{\Lambda}{3}\frac{H_1}{p^2} = \frac{1}{2} f\bar f \frac{p}{q}. \tag{7.3}$$

Then it follows that μ itself is subject to

$$(\mu_{\bar\zeta})_{\zeta\zeta} - \frac{\Lambda}{3}\Big(\frac{\bar\zeta}{p}\mu_{\bar\zeta}\Big)_\zeta = \frac{1}{2} f\bar f\frac{p}{q}. \tag{7.4}$$

A particular solution of (7.4) can be found easily. Indeed, an integration of both sides of (7.4) reduces that problem to a second-order one

$$(\mu_{\bar\zeta})_\zeta - \frac{\Lambda}{3}\frac{\bar\zeta}{p}\mu_{\bar\zeta} = \frac{1}{2}\bar f\int^\zeta \frac{fp}{q}\,d\zeta, \tag{7.5}$$

which can be solved for $\mu_{\bar\zeta}$ by a standard technique. Then another integration with respect to $\bar\zeta$ has to be performed to obtain

$$\mu = \frac{1}{2}\int^{\bar\zeta} d\bar\zeta\, p^2 \int^{\zeta} d\zeta'\, p^{-2}\int^{\zeta'}\frac{f\bar f p}{q}\,d\zeta''. \tag{7.6}$$

Substituting μ back into (7.1) one arrives at

$$\begin{aligned} H_{\text{part}} = {} & \int^{\bar\zeta}\int^{\zeta} f\bar f\frac{p}{q}\,d\zeta\, d\bar\zeta \\ & + \frac{\Lambda}{6}\int^{\bar\zeta} d\bar\zeta\, p\bar\zeta\int^{\zeta} d\zeta'\, p^{-2}\int^{\zeta'}\frac{f\bar f p}{q}\,d\zeta'' \\ & + \frac{\Lambda}{6}\int^{\zeta} d\zeta\, p\zeta\int^{\bar\zeta} d\bar\zeta'\, p^{-2}\int^{\bar\zeta'}\frac{f\bar f p}{q}\,d\bar\zeta'' \\ & - \frac{\Lambda}{6}\frac{\bar\zeta}{p}\int^{\bar\zeta} d\bar\zeta\, p^2\int^{\zeta} d\zeta'\, p^{-2}\int^{\zeta'}\frac{f\bar f p}{q}\,d\zeta'' \\ & - \frac{\Lambda}{6}\frac{\zeta}{p}\int^{\zeta} d\zeta\, p^2\int^{\bar\zeta} d\bar\zeta'\, p^{-2}\int^{\bar\zeta'}\frac{f\bar f p}{q}\,d\bar\zeta''. \end{aligned} \tag{7.7}$$

It is to be remarked that all integrals in this formula are contour integrals in the corresponding complex plane. The variables ζ and $\bar\zeta$ are treated as independent complex variables until all integrations have been performed. Then the relation between them is recalled and incorporated into the final formula for H_{part}.

Further, some special cases of combined electromagnetic and gravitational waves, within the class $R(\Lambda)$ are discussed.

Let ds^2 be in its canonical form (4.36) ($\alpha \equiv 1, \beta \equiv 0$), and let $f(\zeta, \sigma)$ be a polynomial function of ζ and ζ^{-1}. Then instead of the formula (7.7) it is perhaps advantageous to use a slightly different approach in which a new variable $t := (\Lambda/6)\zeta\bar\zeta$ is employed.

Indeed, it is easily seen, that the problem is reduced to that of a particular solution of Eq. (4.35) with its right-hand side equal to $(p/q)\,\bar\zeta^n\,\zeta^{n+k}$, where n and k are some integer numbers. Then the substitution of

$$H = \zeta^k Y(t)/(\Lambda/6)^{n+1} \tag{7.8}$$

into (4.33) leads to an equation on Y:

$$tY'' + (k+1)Y' + \frac{2Y}{(1+t)^2} = \frac{1+t}{1-t}\,t^n. \tag{7.9}$$

A particular solution of (7.9) can be found by the method of variation of parameters, given two independent solutions of the homogeneous problem. [Those, in fact are given, because the general solution of (4.34) is known (4.39) (see the Appendix).]

If, for simplicity,

$$f(\zeta,\sigma) = \zeta^n, \quad n = 0, \pm 1, \pm 2,\ldots, \tag{7.10}$$

then one obtains the following expression for the H_{part}:

(i) $n = -m < -1$,

$$\begin{aligned} H_{\text{part}} = {} & \frac{1-\Lambda/6\zeta\bar\zeta}{1+\Lambda/6\zeta\bar\zeta}\Big\{\frac{(\zeta\bar\zeta)^{1-m}}{(1-m)^2} \\ & + 4\Big(\frac{\Lambda}{6}\Big)^{m-1}\ln\Big|1-\frac{\Lambda}{6\zeta\bar\zeta}\Big| - 4\Big(\frac{\Lambda}{6}\Big)^{m-1} \\ & \times \ln\Big|\frac{\Lambda}{6\zeta\bar\zeta}\Big| + 4\sum_{r=1}^{m-1}\frac{1}{r}\Big(\frac{\Lambda}{6}\Big)^{m-1-r}(\zeta\bar\zeta)^{-r}\Big\} \\ & + \frac{4}{(1-m)(1+\Lambda/6\zeta\bar\zeta)}(\zeta\bar\zeta)^{1-m}, \end{aligned} \tag{7.11}$$

(ii) $n = -1$,

$$\begin{aligned} H_{\text{part}} = {} & \frac{1}{1+\Lambda/6\zeta\bar\zeta}\Big\{4\Big(1-\frac{\Lambda}{6\zeta\bar\zeta}\Big)\ln\Big|1-\frac{\Lambda}{6\zeta\bar\zeta}\Big| \\ & + \frac{2}{3}\Lambda\zeta\bar\zeta\ln\zeta\bar\zeta + \frac{1}{2}\Big(1-\frac{\Lambda}{6\zeta\bar\zeta}\Big)\ln^2(\zeta\bar\zeta)\Big\}, \end{aligned} \tag{7.12}$$

(iii) $n > -1$,

$$\begin{aligned} H_{\text{part}} = {} & 4\,\frac{1-\Lambda/6\zeta\bar\zeta}{1+\Lambda/6\zeta\bar\zeta}\Big(\frac{\Lambda}{6}\Big)^{-n-1}\Big\{\ln\Big|1-\frac{\Lambda}{6\zeta\bar\zeta}\Big| \\ & + \sum_{r=1}^{n}\frac{1}{r}\binom{n}{r}\Big(\frac{\Lambda}{6\zeta\bar\zeta}-1\Big)^r - \sum_{r=1}^{n}\frac{1}{r}\binom{n}{r}(-1)^r\Big\} \\ & + \frac{(\zeta\bar\zeta)^{n+1}}{(n+1)^2(1+\Lambda/6\zeta\bar\zeta)}\Big(5+4n-\frac{\Lambda}{6\zeta\bar\zeta}\Big), \end{aligned}$$

for $n = 0$, $\Sigma_{r=1}^{0} := 0$.

The form of these solutions has been chosen in such a way that the limit transition $\Lambda \to 0$ could be performed.

VIII. DISCUSSION

Plane-fronted, purely gravitational waves with the cosmological constant have been investigated by García Díaz and Plebański in Ref. 5. In this paper we reestablish their main result—generalized Kundt waves $K(\Lambda)$. Moreover, it turns out that for $\Lambda < 0$ there is another family of plane-fronted waves $R(\Lambda)$ essentially distinct from $K(\Lambda)$. The negative sign of Λ is a decisive factor in that distinction. For $\Lambda > 0$ these families coincide.

The presence of an electromagnetic plane-fronted wave does not change the metric structure significantly. It is of the form (4.36), with H given by

$$H = \Phi_\zeta - \frac{\Lambda}{3}\frac{\bar\zeta}{p}\Phi + \bar\Phi_{\bar\zeta} - \frac{\Lambda}{3}\frac{\zeta}{p}\bar\Phi + H_{\text{part}}, \tag{8.1}$$

where Φ is a holomorphic function of ζ and the H_{part} is determined by (7.7).

ACKNOWLEDGMENT

This work was supported by the National Science Foundation under Grant No. PHY-8306104.

APPENDIX

(A1) Two independent solutions of the homogeneous problem related to (7.9) are

$$Y_1 = \frac{(k-1)t + k + 1}{t+1} \quad \text{and} \quad Y_2 = Y_1 \int \frac{e^{-\int[(k+1)/t]dt}}{Y_1^2} dt.$$

In particular for $k = 0$, we have

$$Y_1 = \frac{1-t}{1+t}, \quad Y_2 = \frac{4}{1+t} + \frac{1-t}{1+t} \ln|t|.$$

(A2) Under the transformations (4.19), with the rules (4.22), by choosing $A = -1/\sqrt{2}, B = \sqrt{3/\Lambda}$ $(\Lambda > 0)$ and $\rho' = \rho, \sigma' = (\Lambda/6)\sigma$, it follows that

$$R(\Lambda^+) \equiv K(\Lambda^+).$$

The explicit transformations to $K(\Lambda)$ as given in the García Díaz–Plebański notation from our form of $R(\Lambda^+)$, $K(\Lambda^+)$, and $K(\Lambda^-)$ are

(i)
$$\zeta = \frac{1}{\sqrt{\lambda}} e^{-2\sqrt{\lambda}\xi}, \qquad \lambda = \frac{\Lambda}{6} > 0,$$
$$\rho = \frac{r}{\sinh\sqrt{\lambda}(\xi + \bar{\xi})}, \quad \sigma = -\frac{t}{\sqrt{\lambda}}, \quad 2H = -\sqrt{\lambda}\,\tilde{H},$$

(ii)
$$\zeta = \frac{1}{\sqrt{\lambda}} \tanh\sqrt{\lambda}\,\xi, \quad \lambda = \frac{\Lambda}{6} > 0,$$
$$\rho = \frac{r}{\sinh\sqrt{\lambda}(\xi + \bar{\xi})}, \quad \sigma = -\sqrt{\lambda}\,t, \quad 2H = -\frac{\tilde{H}}{\lambda},$$

and

(iii)
$$\zeta = \frac{1}{\sqrt{\lambda}} \tan\sqrt{\lambda}\,\xi, \qquad \lambda = -\frac{\Lambda}{6} > 0,$$
$$\rho = \frac{r}{\sin\sqrt{\lambda}(\xi + \bar{\xi})}, \quad \sigma = -\sqrt{\lambda}\,t, \qquad 2H = -\frac{\tilde{H}}{\lambda}.$$

[1]E. Schrödinger, *Expanding Universes* (Cambridge U.P., Cambridge, 1956).
[2]I. Ozsvath, UTD Technical Report No. 45, 1978.
[3]A. Lichnerowicz, "Radiation en Relativité General," in *Les Theoris Relativistes de la Gravitation* (CNRS, Paris, 1962).
[4]I. Ozsvath, J. Math. Phys. **6**, 1255 (1965).
[5]A. García Díaz and J. F. Plebański, J. Math. Phys. **22**, 2655 (1981).
[6]J. F. Plebański, "Spinors, Tetrads and Forms," a monograph of Centro de Investigacion y Estudios Avanzados, IPN.
[7]I. Robinson, Lecture at King's College, London, 1956; Report to the Rayaumont Conference, 1959.
[8]H. W. Brinkmann, Proc. Nat. Acad. Sci. (USA) **9**, 1 (1923).
[9]J. Hely, C. R. Acad. Sci. Paris **249**, 1867 (1959).
[10]W. Kundt, Z. Phys. **163**, 77 (1961).

An approximate solution of the monomer–dimer problem on a square lattice. II

Alain J. Phares, Donald E. Shaw, and Francis J. Wunderlich
Department of Physics, Villanova University, Villanova, Pennsylvania 19085

(Received 1 October 1984; accepted for publication 1 March 1985)

The mathematical method developed in paper I is applied to obtain the partition function and thermodynamical properties of the monomer–dimer problem for a square lattice in terms of the absolute activity x. We also obtain by extrapolation an approximate expression of the partition function which is accurate to better than 0.1% in the range $0\leqslant x\leqslant 10$. The expectation of the statistics $\langle\theta(x)\rangle$ is calculated in two different ways and numerical results agree to better than 3%, thus showing the consistency of the underlying mathematical method. Consistent with earlier studies, there is no phase transition. Approximation methods used in earlier work are also found to be in good agreement with our analytic study.

I. INTRODUCTION

This article is a continuation of paper I entitled "The occupation statistics for indistinguishable dumbbells on a rectangular lattice space."[1] In paper I, the monomer–dimer problem was formulated by considering dimers (or dumbbells) distributed on an $L\times M\times N$ rectangular lattice, where L and M are fixed and N is allowed to become very large. $A(q,N)$ represented the number of ways of arranging q indistinguishable dimers on the $L\times M\times N$ lattice. Following McQuistan and his collaborators,[2] the bivariant generating function $G(x,y)$ was introduced

$$G(x,y)=\sum_{N=0}^{\infty}\sum_{q=0}^{[LMN/2]}A(q,N)x^q y^N. \tag{1.1}$$

The notation $[LMN/2]$ means the largest integer contained in $(LMN/2)$. This is the largest possible number $q_{\max}$ of dimers that can be arranged on the lattice. The variable x is identified as the absolute activity of the dimer. Parameter y is chosen in such a way as to secure the convergence of this infinite series. The results obtained are, of course, independent of the particular choice of y securing the convergence of the series. It then follows that the bivariant generating function is obtained in closed form as a ratio of two polynomials

$$G(x,y)=H(x,y)/D(x,y). \tag{1.2}$$

The configurational grand-canonical partition function for a given absolute activity x is

$$\Delta_{LMN}(x)=\sum_{q=0}^{q_{\max}}A(q,N)x^q. \tag{1.3}$$

The expectation of the coverage by dimers as a function of the absolute activity is

$$\langle\theta(x)\rangle_{LMN}=\frac{1}{q_{\max}\Delta_{LMN}(x)}\sum_{q=0}^{q_{\max}}qA(q,N)x^q. \tag{1.4}$$

The dispersion in θ is

$$\langle\theta^2(x)\rangle_{LMN}=\frac{1}{q_{\max}\Delta_{LMN}(x)}\sum_{q=0}^{q_{\max}}q^2A(q,N)x^q. \tag{1.5}$$

Keeping L and M fixed, the partition function in the thermodynamic limit is

$$\Xi_{LM}(x)=\lim_{N\to\infty}[\Delta_{LMN}(x)]^{1/LMN}. \tag{1.6}$$

In paper I, it was shown that the knowledge of $D(x,y)$ was sufficient to obtain $\langle\theta(x)\rangle_{LM\infty}$, $\langle\theta^2(x)\rangle_{LM\infty}$, and $\Xi_{LM}(x)$. Let $R_1(x;L,M)$ be the largest z root of $(z=1/y)$

$$D(x,1/z)=0. \tag{1.7}$$

It then followed that

$$\langle\theta(x)\rangle_{LM\infty}=(\langle\theta^2(x)\rangle_{LM\infty})^{1/2}=\frac{2}{LM}\left(\frac{x}{y}\frac{\partial D/\partial x}{\partial D/\partial y}\right)_{y=1/R_1(x)} \tag{1.8}$$

and

$$\Xi_{LM}(x)=[R_1(x;L,M)]^{1/LM}. \tag{1.9}$$

The function $D(x,y)$ was explicitly calculated for planar lattices ($L=1$) and up to and including four rows, i.e., for $M=1,2,3,$ and 4. We list in Table I the expressions of $D(x,y)$ derived in paper I. The analytic expressions for the largest roots $R_1(x;1,M)$ were easily calculated for $M=1$ and $M=2$ simply because $D(x,1/z)$ turned out to be a quadratic and cubic polynomial in y, respectively. The roots were calculated to be

$$R_1(x;1,1)=\tfrac{1}{2}[1+\sqrt{1+4x}], \tag{1.10}$$

$$R_1(x;1,2)=(1+2x)/3+\tfrac{2}{3}[4x^2+7x+1]^{1/2}\times\cos(\phi(x)/3), \tag{1.11a}$$

TABLE I. The polynomial $D(x,y)$ for lattices with $M=1$, 2, 3, and 4.

M	$D(x,y)$
1	$1-y-xy^2$
2	$1-(1+2x)y-xy^2+x^3y^3$
3	$1-(1+3x)y-(2+7x+5x^2)xy^2-(1+x-2x^2)x^2y^3$ $+(2+3x+5x^2)x^4y^4-(1-x)x^6y^5-x^9y^6$
4	$1-(1+6x+2x^2)y-(1+9x+28x^2+10x^3)xy^2$ $+(1+12x+39x^2+31x^3+14x^4)x^2y^3$ $+(1+16x+84x^2-176x^3+143x^4+41x^5)x^3y^4$ $+(2+14x+21x^2-36x^3-57x^4-34x^5)x^5y^5$ $-(7+60x+227x^2+429x^3+297x^4+82x^5)x^7y^6$ $-(16+70x+124x^2-14x^3-38x^4)x^{10}y^7$ $+(7+56x+175x^2+285x^3+273x^4+86x^5)x^{11}y^8$ $-(2-2x-43x^2-94x^3-34x^4+20x^5)x^{13}y^9$ $-(1+12x+32x^2+64x^3+103x^4+47x^5)x^{15}y^{10}$ $-(1+6x+9x^2+17x^3-4x^4)x^{18}y^{11}$ $+(1+5x+12x^2+12x^3)x^{21}y^{12}+x^{24}y^{13}-x^{28}y^{14}$

 0022-2488/85/071762-07$02.50

where

$$\cos\phi(x) = -\frac{(11x^3 - 42x^2 - 21x - 2)}{2(4x^2 + 7x + 1)^{3/2}}. \quad (1.11b)$$

Also in paper I, it was shown that, in the large-M limit, the first-order correction to $\langle\theta(x)\rangle_{LM\infty}$ is of order $(1/M)$. Setting the absolute activity to be unity, $x = 1$, a plot of $\langle\theta(1)\rangle_{1M\infty}$ vs $(1/M)$ showed a rapid convergence toward a linear curve beyond $M = 1$. A least-square fit of the three points corresponding to $M = 2, 3, 4$ predicted an expectation of dimers on a planar lattice of infinite extent to be

$$\langle\theta(x=1)\rangle_{1\infty\infty} = 63.4\%. \quad (1.12)$$

Again, setting $x = 1$, paper I exhibited the exponential behavior of the largest root $R_1(1;1,M)$ as a function of M.

In this article, we plan to extend these results to other values of the activity. Explicit numerical results are obtained in the range $0 \leqslant x \leqslant 10$. A closed form analytic expression of the partition function $\Xi_{1\infty\infty}(x)$ for the infinte two-dimensional lattice is obtained by extrapolation and from it other thermodynamic quantities are derived in the usual way.

II. THE PARTITION FUNCTION $\Xi(x)$

Table II gives the values of the largest root $R_1(x;1,M)$ for $M = 1$, 2, 3, and 4 and various values of the absolute activity x in the range $0 \leqslant x \leqslant 10$. These values are computed from the analytical expressions (1.10) and (1.11) for $M = 1$ and 2, and by a direct numerical search for the largest z root of the polynomial $D(x,1/z)$, listed in Table I, for $M = 3$ and 4. A numerical study of the z roots as a function of x showed the leading root to remain the leading root for the values of x in the range $0 \leqslant x \leqslant 10$. Although we could not come up with any rigorous mathematical proof, we find it is reasonable to conjecture that this is the case for all values of x, not only for $M = 1$ and 2 but also for all other values of M. Figure 1 gives the plot of the z roots against the absolute activity x for $M = 3$.

The next step is to exhibit for any given value of x the exponential behavior of $R_1(x;1,M)$ as a function of M. These

TABLE II. The largest x root $R_1(x;L=1,M)$ of polynomial $D(x,1/z)$ for various lattices ($M = 1$, 2, 3, and 4) and x in the range $0 \leqslant x \leqslant 10$.

X	$M=1$	$M=2$	$M=3$	$M=4$
0.1	1.091 61	1.277 66	1.489 43	1.736 71
0.2	1.170 82	1.527 50	1.975 39	2.556 47
0.3	1.241 62	1.761 60	2.467 87	3.461 70
0.4	1.306 23	1.985 25	2.969 94	4.451 01
0.5	1.366 03	2.201 34	3.482 71	5.522 57
0.6	1.421 95	2.411 65	4.006 57	6.674 61
0.7	1.474 68	2.617 37	4.541 56	7.905 58
0.8	1.524 69	2.819 34	5.087 59	9.214 10
0.9	1.572 38	3.018 17	5.644 49	10.599 0
1.0	1.618 03	3.214 32	6.212 07	12.059 1
2.0	2.000 00	5.083 87	12.429 3	30.630 6
3.0	2.302 78	6.863 99	19.511 2	56.075 9
4.0	2.561 55	8.599 75	27.335 4	88.097 5
5.0	2.791 29	10.308 8	35.820 3	126.519
6.0	3.000 00	12.000 0	44.905 1	171.220
7.0	3.192 59	13.678 5	54.544 5	222.115
8.0	3.372 28	15.347 6	64.701 2	279.138
9.0	3.541 38	17.009 4	75.344 3	342.235
10.0	3.701 56	18.665 5	86.449 1	411.366

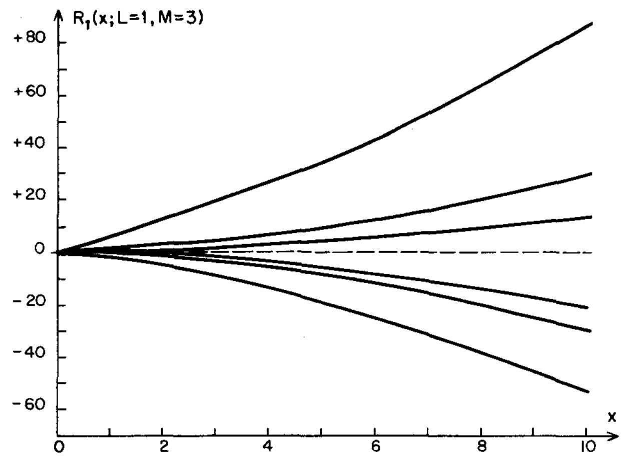


FIG. 1. The z roots of $D(x,1/z)$ as a function of the absolute activity x for $M = 3$.

roots cannot be calculated analytically for $M = 3$ and $M = 4$, as already mentioned earlier; they were obtained numerically for various values of x. For $M = 1$ and $M = 2$, the analytic expressions (1.10) and (1.11) were used to obtain the numerical values of the roots. We plotted on a semilog graph paper these roots versus M. As exhibited in Fig. 2, we found (for any given value of x) the four points plotted ($M = 1$, 2, 3, and 4) to fall along a straight line. A least-square fit of these points gives a correlation coefficient of 0.99. Based on this approximation, we express $R_1(x;1,M)$ as

$$R_1(x;1,M) = A(x)\exp[B(x)M], \quad (2.1)$$

or

$$\ln[R_1(x;1,M)] = \ln[A(x)] + B(x)M. \quad (2.2)$$

Following paper I, we obtain the values of $B(x)$ by a least-square fit of $\ln R_1$ vs M. Obviously, $B(x)$ is the slope for any given value of x, with only four points corresponding to

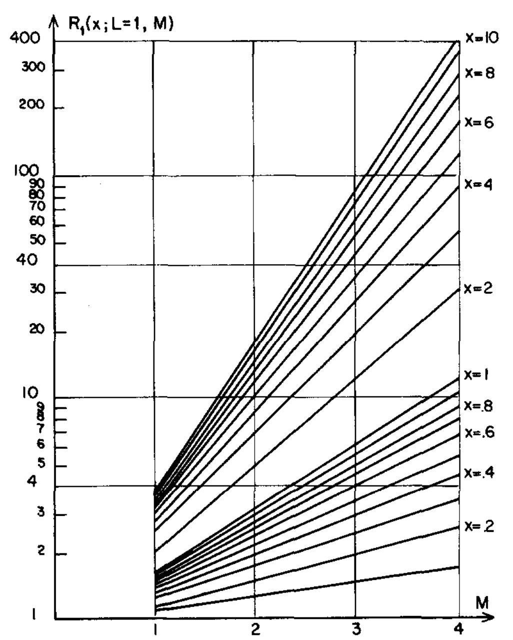


FIG. 2. Semilogarithmic plot of $R_1(x;1,M)$ vs M for various values of the absolute activity x.

$M = 1, 2, 3$, and 4. Since the partition function of the infinite planar lattice is[1]

$$\Xi(x) = \lim_{M\to\infty} [R_1(x;1,M)]^{1/M}, \tag{2.3a}$$

it then follows that

$$\Xi(x) = \lim_{M\to\infty} \{[A(x)]^{1/M}\exp[B(x)]\} = \exp[B(x)]. \tag{2.3b}$$

Table III summarizes all the numerical results obtained for values of x in the range $0 \leqslant x \leqslant 10$.

We expect the approximate exponential behavior to become more accurate with increasing values of M. Thus, the slope $B(x)$, based on the first four points, is not expected to be the same when computed from a linear fit of the next four points. Nevertheless, assuming that (2.1) is perfectly accurate for all values of M including $M = 1$ and $M = 2$, then it is obvious that by taking the ratio $R_1(x;1,M = 2)/R_1(x;1,M = 1)$ one obtains $\exp[B(x)]$ which is the partition function for the infinite planar lattice system. The benefit of this is having an approximate expression for the partition function of a system of dimers distributed on a planar lattice of infinite extent, namely,

$$\Xi^A(x) = \tfrac{1}{2}[1 + \sqrt{1+4x}]\{(1+2x)/3 + \tfrac{2}{3}(4x^2 + 7x + 1)^{1/2}\times\cos[\phi(w)/3]\}^{-1}, \tag{2.4}$$

where $\phi(x)$ is given by Eq. (1.11b). The values of $\Xi^A(x)$ are listed, for comparison, in Table III. In support of our claim, we have also computed the ratio $R_1(x;1,4)/R_1(x;1,3)$. As anticipated, the numbers obtained are closer to the exact values of $\Xi(x)$ to better than 1% in the range $0 \leqslant x \leqslant 10$. Indeed, in general we would have

$$\Xi(x) = R_1(x;1,M+1)/R_1(x;1,M). \tag{2.5}$$

Such a relation becomes more accurate as M becomes increasingly large. Since no analytic expression of $R_1(x;1,M)$ can be obtained for M greater than 2, we computed a correction term to the approximate partition function $\Xi^A(x)$ by obtaining an exponential fit to the difference $[\Xi^A(x) - \Xi(x)]$ for values of x in the range $0 \leqslant x \leqslant 10$; one finds

$$\Xi^C(x) = \Xi^A(x) - 0.034\,63x^{0.858\,656}. \tag{2.6}$$

To have an idea of the limitations of such a closed-form analytic formula, let us consider the large x behavior of $\Xi^A(x)$. Equation (1.11b) giving the explicit expression of $\cos\phi(x)$ shows that

$$\lim_{x\to\infty} \cos\phi(x) = -\tfrac{11}{16}, \quad \phi(\infty)/3 = 44.48°. \tag{2.7}$$

Thus the leading term of $\Xi^A(x)$ as x becomes very large is

$$\Xi_A(x) \sim 1.618\sqrt{x}. \tag{2.8}$$

This shows that the correction term $0.034\,63x^{0.859}$ becomes a leading term in the expression (2.6) of the partition function. Since the partition function cannot be negative, it is evident that its closed form (2.6) will fail to give the proper value. Formally, Eq. (2.6) becomes negative at values of x larger than

$$(1.618/0.034\,63)^{1/0.358\,656} = 45\,180. \tag{2.9}$$

Therefore, we can safely argue that the validity of the closed-form analytic expression (2.6) of the partition function may be extended over a range which is several times larger than $0 \leqslant x \leqslant 10$. Figure 3 is a plot of Ξ^A and Ξ^C versus the absolute activity x in the range $0 \leqslant x \leqslant 10$, as well as Ξ for $M = 1$ and $M = 2$ calculated from Eqs. (1.9), (1.10), and (1.11).

III. THE EXPECTATION OF THE STATISTICS $\langle\theta(x)\rangle$ AND THE NUMBER DENSITY $\rho(x)$

We follow the same procedure as the one presented in paper I for $x = 1$. We use the analytic expression (1.8) of the expectation on a planar lattice ($L = 1$) using for $D(x,y)$ the polynomials given in Table I. For all values of x, Fig. 4 shows

TABLE III. Numerical values of the partition function calculated in several different ways are listed in this table for comparison.

X	$\Xi^A(x)$	$\dfrac{R_1(M=4)}{R_1(M=3)}$	$B(x)$	$\Xi(x) = e^{B(x)}$	$\Xi^C(x)$
0.1	1.17	1.17	0.155	1.17	1.17
0.2	1.30	1.29	0.260	1.30	1.30
0.3	1.42	1.40	0.341	1.41	1.41
0.4	1.52	1.50	0.408	1.50	1.50
0.5	1.61	1.59	0.465	1.59	1.59
0.6	1.70	1.67	0.515	1.67	1.67
0.7	1.77	1.74	0.559	1.75	1.75
0.8	1.85	1.81	0.599	1.82	1.82
0.9	1.92	1.88	0.635	1.89	1.89
1.0	1.99	1.94	0.668	1.95	1.96
2.0	2.54	2.46	0.908	2.48	2.48
3.0	2.98	2.87	1.06	2.89	2.90
4.0	3.36	3.22	1.18	3.24	3.25
5.0	3.69	3.53	1.27	3.56	3.56
6.0	4.00	3.81	1.35	3.84	3.84
7.0	4.28	4.07	1.41	4.10	4.10
8.0	4.55	4.31	1.47	4.34	4.35
9.0	4.80	4.54	1.52	4.57	4.58
10.0	5.04	4.76	1.57	4.79	4.79

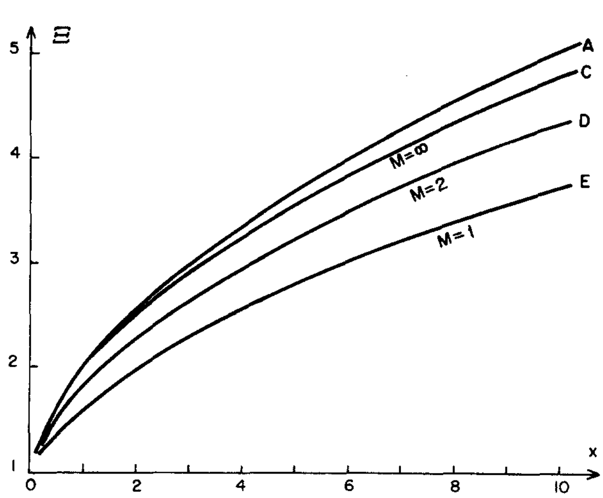


FIG. 3. Plot of the partition function versus the absolute activity. Curve (A) is the approximate partition function $\Xi^A(x)$ as computed analytically. Curve (C) is $\Xi^A(x)$ with the correction added to it; this curve coincides with the exact partition function calculated numerically. Curves (D) and (E) represent the partition function for lattices with $M = 2$ and $M = 3$, respectively, and as computed analytically from Eqs. (1.10) and (1.11a) and (1.11b).

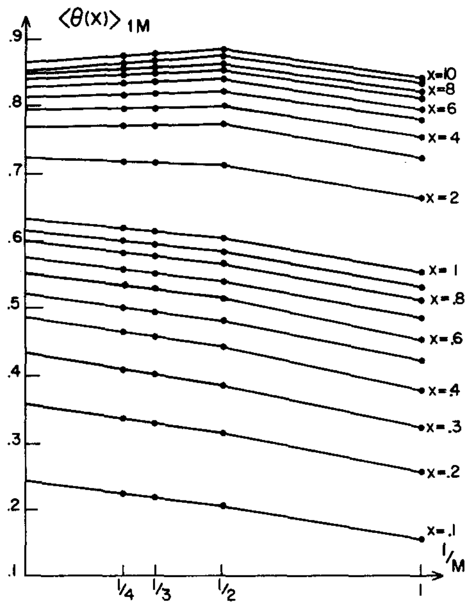


FIG. 4. The expectation of the statistics $\langle\theta(x)\rangle_{1M}$ vs $1/M$ for various values of the activity x in the range $0<x<10$.

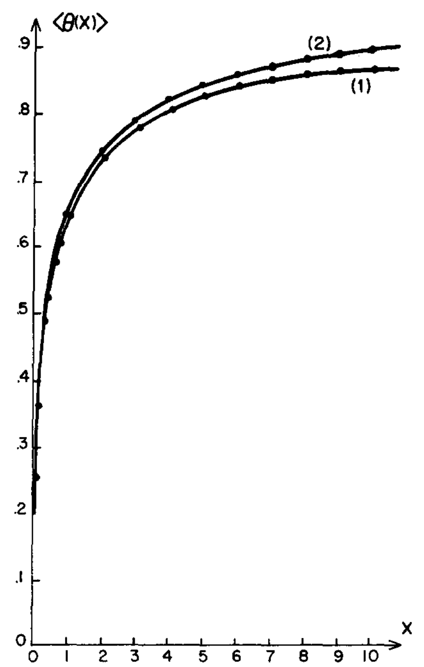


FIG. 5. Plot of the expectation of the statistics, $\langle\theta(x)\rangle$, on a planar lattice of infinite extent versus the absolute activity x. Curve (1) is $\langle\theta(x)\rangle$ computed from the linear extrapolation shown in Fig. 4 and whose values are listed in the last column of Table IV. Curve (2) is $\rho(x)/\rho_0 \equiv \langle\theta(x)\rangle$ computed from the knowledge of the partition function, and whose values are listed in Table V.

that linearity is almost achieved beyond $M=1$ when plotting $\langle\theta(x)\rangle_{1M_\infty}$ vs $(1/M)$. A least-square fit of the points for $M=2$, 3, and 4 gives an intercept $\langle\theta(x)\rangle$, which is precisely the expectation of the statistics on the planar lattice of infinite extent, i.e.,

$$\langle\theta(x)\rangle = \lim_{M\to\infty} \langle\theta(x)\rangle_{1M_\infty}. \tag{3.1}$$

Table IV gives the values of the expectation $\langle\theta(x)\rangle_{1M_\infty}$ computed analytically for different values of x in the range $0<x\leqslant 10$ and for $M=1, 2, 3$, and 4. The last column of Table IV gives the values of the expectation $\langle\theta(x)\rangle$ obtained from the linear extrapolation as exhibited in Fig. 4. Finally, curve (1) in Fig. 5 is the plot of $\langle\theta(x)\rangle$ vs x, obtained from the values listed in the last column of Table IV.

The expectation of the statistics could be calculated using a different approach based on the knowledge of the partition function. This is accomplished by combining first Eqs. (1.3) and (1.4)

$$\langle\theta(x)\rangle_{LMN} = (x/[LMN/2])(\Delta'_{LMN}(x)/\Delta_{LMN}(x)). \tag{3.2}$$

Setting $\Xi_{LMN}(x) = [\Delta_{LMN}(x)]^{1/LMN}$, then the partition function, $\Xi_{LM}(x)$ is the limit as N approaches infinity of $\Xi_{LMN}(x)$. It then follows that

TABLE IV. Values of $\langle\theta(x)\rangle_{1M_\infty}$ are listed for x in the range $0<x\leqslant 10$ and for $M=1, 2, 3$, and 4. The last column is the computed extrapolated values for $M=\infty$.

X	$M=1$	$M=2$	$M=3$	$M=4$	···	$M=\infty$
0.1	0.154 845	0.204 053	0.217 019	0.223 719	···	0.243 291
0.2	0.254 640	0.315 242	0.329 672	0.337 426		0.359 377
0.3	0.325 800	0.388 717	0.402 770	0.410 595		0.432 128
0.4	0.379 824	0.442 273	0.455 514	0.463 148		0.483 587
0.5	0.422 645	0.483 721	0.495 995	0.503 424		0.522 571
0.6	0.457 674	0.517 127	0.528 561	0.535 650		0.562 055
0.7	0.487 053	0.544 852	0.555 424	0.562.244		0.578 977
0.8	0.512 048	0.568 380	0.578 139	0.584 704		0.600 304
0.9	0.533 748	0.588 698	0.597 691	0.604 020		0.618 603
1.0	0.552 787	0.606 484	0.614 764	0.620 875		0.634 420
2.0	0.666 666	0.713 190	0.715 996	0.720 648		0.726 712
3.0	0.722 641	0.766 518	0.765 657	0.769 569		0.770 758
4.0	0.757 464	0.800 216	0.796 625	0.800 120		0.797 756
5.0	0.781 782	0.824 081	0.818 295	0.821 535		0.816 360
6.0	0.799 998	0.842 105	0.834 545	0.837 625		0.830 204
7.0	0.814 304	0.856 356	0.847 306	0.850 281		0.840 991
8.0	0.825 922	0.867 970	0.857 663	0.860 565		0.849 707
9.0	0.835 601	0.877 658	0.866 286	0.869 129		0.856 945
10.0	0.834 826	0.885 904	0.873 600	0.876 398		0.863 056

TABLE V. Computed thermodynamic quantites for different values of the activity x.

x	$\langle\theta(x)\rangle_{1\infty\infty}$	$\frac{\rho(x)}{\rho_0}$	$\Gamma(x)$	$\frac{S_v(x)}{k_B}$	$k_B T\rho^2 K_T(x)$	$\frac{C_v(x)}{k_B}$
0.1	0.243	0.247	0.0766	0.219	0.0412	0.0633
0.2	0.359	0.367	0.130	0.277	0.0460	0.0919
0.3	0.432	0.441	0.171	0.303	0.0457	0.109
0.4	0.484	0.492	0.204	0.317	0.0443	0.122
0.5	0.523	0.531	0.234	0.325	0.0428	0.132
0.6	0.562	0.561	0.258	0.329	0.0414	0.139
0.7	0.579	0.586	0.280	0.332	0.0402	0.146
0.8	0.600	0.608	0.300	0.333	0.0390	0.152
0.9	0.619	0.626	0.318	0.334	0.0380	0.157
1.0	0.634	0.641	0.334	0.334	0.0370	0.161
2.0	0.727	0.732	0.454	0.327	0.0308	0.188
3.0	0.771	0.780	0.531	0.316	0.0272	0.200
4.0	0.798	0.812	0.588	0.307	0.0246	0.207
5.0	0.816	0.833	0.634	0.299	0.0226	0.209
6.0	0.830	0.849	0.673	0.292	0.0210	0.211
7.0	0.841	0.862	0.706	0.286	0.0191	0.205
8.0	0.850	0.871	0.734	0.281	0.0181	0.205
9.0	0.857	0.880	0.760	0.227	0.0172	0.205
10.0	0.863	0.887	0.784	0.273	0.0163	0.204

$$\frac{\Delta'_{LMN}(x)}{\Delta_{LMN}(x)} = \frac{d}{dx}\left(\ln\left[\Xi_{LMN}(x)\right]^{LMN}\right)$$
$$= (LMN)\left[\Xi'_{LMN}(x)/\Xi_{LMN}(x)\right]. \quad (3.3)$$

Combining Eqs. (3.2) and (3.3) and then taking the limit as N approaches infinity, one obtains

$$\langle\theta(x)\rangle_{LM\infty} = 2x\left[\Xi'_{LM}(x)/\Xi_{LM}(x)\right]. \quad (3.4)$$

For the case of the infinite planar lattice, $L = 1$ and $M = \infty$, one finds

$$\langle\theta(x)\rangle = \lim_{M\to\infty}\langle\theta(x)\rangle_{1M\infty} = 2x\left[\Xi'(x)/\Xi(x)\right], \quad (3.5)$$

where we have used the notation introduced earlier in (2.3a), namely,

$$\Xi(x) = \lim_{M\to\infty}\Xi_{1M}(x). \quad (3.6)$$

Following Gaunt,[3] the number density for the square lattice is given by

$$\rho(x) = x\frac{d\Gamma(x)}{dx}, \quad (3.7)$$

where $\Gamma(x)$ is the grand potential function for this square lattice, namely,

$$\Gamma(x) = \tfrac{1}{2}\ln[\Xi(x)]. \quad (3.8)$$

Recalling that the density number for close packing of dimers, ρ_0, is $\frac{1}{4}$ on the infinite square lattice, then the normalized number density is

$$\frac{\rho(x)}{\rho_0} = (4x)\frac{1}{2}\frac{d}{dx}(\ln[\Xi(x)]). \quad (3.9)$$

The above expression is identically the same as the expectation of the statistics $\langle\theta(x)\rangle$, Eq. (3.5)

$$\langle\theta(x)\rangle = \rho(x)/\rho_0 = 2x\left[\Xi'(x)/\Xi(x)\right]. \quad (3.10)$$

Curve (2) in Fig. 5 is the plot of $\langle\theta(x)\rangle$ or $\rho(x)/\rho_0$ as calculated from Eq. (3.10) using for the partition function the analytic expression, Eqs. (2.4) and (2.6), derived earlier. The second and third columns of Table V list the values of $\langle\theta(x)\rangle_{1\infty\infty}$ and $\rho(x)/\rho_0$, respectively, up to three significant figures. According to Eq. (3.10), there should be an identical matching between these results. Since the methods of obtaining these numerical results are different, the largest deviation being better than 3% shows the consistency of the underlying mathematical theory. The method of calculating $\rho(x)/\rho_0$, using Eq. (3.10) and the analytic expression of the partition function, is certainly more accurate than the other extrapolation method, Eq. (3.1).

IV. OTHER THERMODYNAMICAL PROPERTIES

From the knowledge of the analytic partition function $\Xi^c(x)$ we calculate the grand potential $\Gamma(x)$ using Eq. (3.8) (see Fig. 6). Other thermodynamical functions are calculated in the usual way, namely, as follows.

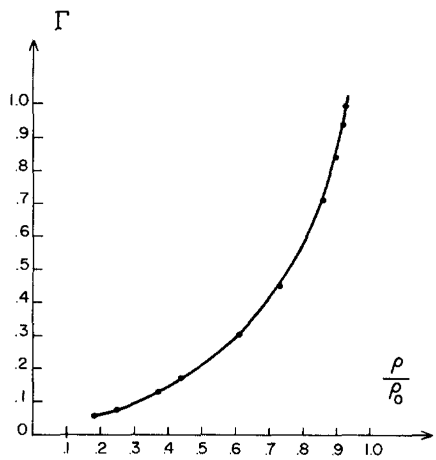


FIG. 6. The grand potential function Γ, plotted against the normalized number density (ρ/ρ_0).

TABLE VI. The results by Gaunt for the maximum values of the thermodynamical quantities as compared to the results obtained in this paper (PSW).

Thermodynamical quantity	(ρ/ρ_0)		Maximum	
	Gaunt	PSW	Gaunt	PSW
S_v/k_B	0.636	0.641	0.331	0.334
$k_B T\rho^2 K_T$	0.42	0.391	0.0454	0.0461
C_V/k_B	0.85	0.843	0.20	0.211

(a) The entropy per unit volume S_v (k_B = Boltzmann's constant[3] is

$$S_v(x)/k_B = -\rho(x)\ln x + \Gamma(x). \tag{4.1}$$

(b) The isothermal compressibility K_T (T = absolute temperature)[3] is

$$k_B T\rho^2(x)K_T(x) = x\,\frac{d\rho}{dx}. \tag{4.2}$$

(c) The constant-pressure specific heat per unit C_v (Ref. 3) is

$$\frac{C_v(x)}{k_B} = x\left(\frac{d\rho}{dx}\right)\left(\frac{\Gamma(x)}{\rho(x)}\right)^2. \tag{4.3}$$

In Table V we list the values of these thermodynamical functions for absolute activities in the range $0<x<10$.

As predicted by the approximate calculations made by Gaunt,[3] the three thermodynamical quantities (4.1), (4.2), and (4.3) have a maximum. The maximum values and the corresponding values of the normalized number density (ρ/ρ_0) obtained by Gaunt[3] are listed in Table VI for comparison with our results. Agreement between our analytical approach and Gaunt's approximate method is quite good. The dependence of these thermodynamical quantities as a function of (ρ/ρ_0) is shown in Figs. 7, 8, and 9.

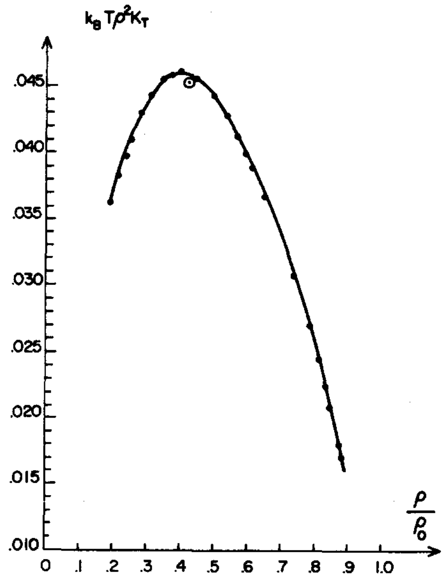


FIG. 8. $k_B T\rho^2 K_T$ plotted against the normalized number density (ρ/ρ_0). The circled data point is the maximum obtained by Gaunt.[3]

V. CONCLUSION

We obtained a closed-form analytic expression $\Xi^C(x)$ of the partition function of a system of dimers on a planar square lattice. This expression was arrived at by conjecturing that the largest root $R_1(x;L=1,M)$ increases exponentially with M. Such a conjecture is justified by the exponential behavior of $\Delta_N(x)$ with increasing values of N already made explicit in paper I.

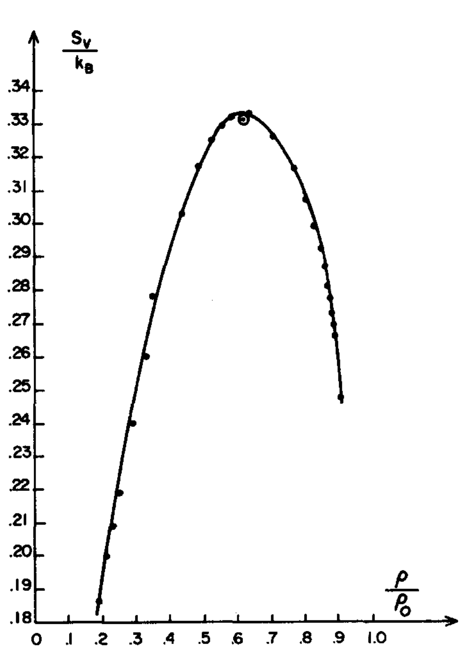


FIG. 7. S_v/k_B plotted against the normalized number density (ρ/ρ_0). The circled data point is the maximum obtained by Gaunt.[3]

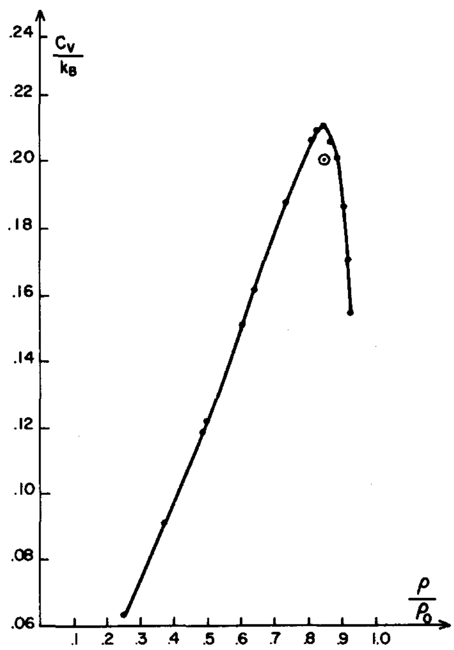


FIG. 9. C_v/k_B plotted against the normalized number density (ρ/ρ_0). The circled data point is the maximum obtained by Gaunt.[3]

The internal consistency of the underlying mathematical method was verified by calculating the expectation of the statistics $\langle\theta(x)\rangle$ in two different ways with an agreement of better than 3%.

Not available in the literature, we give the values of several thermodynamical quantities as a function of the absolute activity x and also as a function of the normalized number density ρ/ρ_0.

As expected, no phase transition is observed. The maxima reported by Gaunt[3] for the square lattice using the series expansion method are found to be in good agreement with our closed-form analytic approach.

[1] A. J. Phares, J. Math. Phys. **25**, 1756 (1984).

[2] D. Lichtman and R. B. McQuistan, J. Math. Phys. **8**, 2441 (1967); R. B. McQuistan, Nuovo Cimento B **58**, 86 (1968); R. B. McQuistan and S. J. Lichtman, J. Math. Phys. **11**, 3095 (1970); R. B. McQuistan and J. L. Hock, Fibonacci Quart. **21**, 196 (August 1983); J. L. Hock and R. B. McQuistan, J. Math. Phys. **24**, 1859 (1983).

[3] D. S. Gaunt, Phys. Rev. **179**, 174 (1969).

A multitype random sequential process. II. Distribution of particle size and vacant space length in the saturation limit

B. Mellein
Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA), División Química Teórica, Sucursal 4, Casilla de Correo 16, 1900 La Plata, Argentina

(Received 26 September 1984; accepted for publication 28 December 1984)

A one-dimensional lattice space of n equivalent compartments is filled sequentially at random with nonoverlapping particles of integral length β (β-bell particles), the latter assumed to be a random variable with probability distribution $\{p_q,...,p_r\}$ on $\{q,q+1,...,r-1,r\}$, $q\geqslant 1$. Due to configurational degeneracies the relative probability $p^*_{k,n}$ of ultimately finding a k-bell particle on the saturated lattice space will generally not coincide with p_k, the "input" probability. In the present paper we shall determine $p^*_{k,n}$, $k=q,...,r$, and its limit as n tends to infinity. Some more insight into the occupation configuration of the lattice space in the jammed state is gotten by means of the length distribution of stretches of unoccupied compartments (gaps).

I. INTRODUCTION

One-type random sequential space filling procedures in one dimension have been formulated to describe statistical aspects of physical–chemical phenomena such as adsorption,[1–3] crystallization,[4,5] and intramolecular reactions of polymers.[5–11] As pointed out recently,[12] the more general concept of a multitype random sequential process constitutes an improvement upon former approaches in the sense that it covers situations (additionally to those previously cited we mention the case of a discrete finite cascade process[13] and a concrete example referred to in Boucher and Nisbet[14]) which allow for sequential addition of various kinds of objects. The classical notion of a one-type random sequential process, however, does not extend uniquely to a multitype formulation, and two distinct multitype random sequential processes, designated model I and model II, have been introduced.[12] This research represents a generalization of previously published results[12] concerning model II.

In model II particles of random lengths are placed on a $1\times n$ lattice space, one at a time and randomly and subject to the condition that no two particles overlap. The formal description of the filling procedure is as follows. From the probability distribution $P=\{p_q,...,p_r\}$ on $\{q,q+1,...,r-1,r\}$, with $q\geqslant 1$ and $p_q>0$, we observe an integer β_1 and proceed to place a β_1-bell particle on the $1\times n$ array at random, i.e., the particle's left-hand end point occupies any of the sites $1,2,..., n-\beta_1+1$ with equal probabilities $1/(n-\beta_1+1)$. The thus-arising random $1\times n_1$ and $1\times n_1'$ subarrays $[n_1+n_1'+\beta_1=n]$ will be filled independently and (statistically) identically in the following manner: If the $1\times n_1$ lattice space is still accessible, i.e., if $n_1\geqslant q$, we sample from the probability distribution P until observing a first value β_2 not exceeding n_1 and subsequently insert a β_2-bell particle at random in the $1\times n_1$ array. Next we turn to a first further, similarly effected occupation of the $1\times n_1'$ lattice space and thus continue filling the arising random subarrays. Ultimately, in the so-called terminal, jammed, or saturated state, no further particle fits and the placement process has come to an end.

Several random variables evolve in the analysis of the occupation configuration of a $1\times n$ array in the jammed state. One, particularly important, is A_n, the total number of unoccupied compartments, and this has been studied in some detail.[12] In the present paper we shall examine two sets of random variables, each one determining A_n and thus providing a more complete picture of how the array's saturation occupation is made up: B^k_n, $k=q,...,r$, the number of k-bell particles accommodated on the $1\times n$ lattice space and C^m_n, $m=0,1,...,q-1$, the number of m-gaps (i.e., stretches of exactly m contiguous unoccupied compartments) are ultimately present. Then, as asserted, there are the following relations:

$$A_n=n-\sum_{k=q}^{r}kB^k_n=\sum_{m=1}^{q-1}mC^m_n. \tag{1}$$

It is the asymptotic behavior $(n\to\infty)$ of b^k_n and c^m_n, the expectations of B^k_n and C^m_n, respectively, we are interested in. On the basis of recursion relationships satisfied by b^k_n and c^m_n the appropriate generating functions will be deduced and utilized to obtain, for all $0<\epsilon<\rho=1/(r-1)$, as $n\to\infty$,

$$b^k_n=\lambda_k n+\lambda'_k+O(n^{-n(\rho-\epsilon)}),\quad k=q,...,r \tag{2}$$

and

$$c^m_n=\rho_m n+\rho'_m+O(n^{-n(\rho-\epsilon)}),\quad m=0,1,...,q-1, \tag{3}$$

where λ_k, λ'_k, ρ_m, and ρ'_m are constants (clearly depending on $p_q,...,p_r$, but independent of n) stated explicitly in the text [see Eqs. (28) and (29) below].

From the results given in Eq. (2) the relative probabilities

$$p^*_k=\frac{\lambda_k}{\Sigma^r_{i=q}\lambda_i},\quad k=q,...,r, \tag{4}$$

of finding a k-bell particle among the particles filling an infinite lattice space can be computed. In Sec. VII we will present some numerical calculations which show that the difference between p^*_k and the "input" probability p_k may be considerable. This is particularly true for values of k either close to q or close to r, p^*_k being larger than p_k in the former case and smaller in the latter one. However, these differences are even more pronounced[15] in model I in which the larger particles are put at a disadvantage[12] from the very beginning

 0022-2488/85/071769-06$02.50

of the filling process.

Similarly, Eq. (3) yields

$$q_m = \frac{\rho_m}{\Sigma_{i=0}^{q-1}\rho_i}, \quad m = 0,\dots,q-1, \tag{5}$$

the relative probability of a stretch of m adjacent vacant sites (m-gap) on a saturated infinite array, which, in the one-type case $q = r$, has been first determined by Mackenzie.[2]

II. RECURSION RELATIONSHIPS

To establish a recursion relation for b_n^i, $i = q,\dots,r$, fix $k \in \{q,\dots,r\}$, suppose that $n \geqslant r$, and recall from our model assumptions that (1) the first particle placed on the $1\times n$ array will be an i-bell particle with probability p_i, (2) being the first particle to be inserted an i-bell particle, there are $n - i + 1$ equally probable choices for placing the particle, and (3) if the first particle getting stuck is an i-bell particle occupying compartments $j, j+1,\dots,j+i-1$, further occupation will be directed, in a statistically identical manner, to two subsequently independent subarrays consisting of $j-1$ and $n-i-j+1$ sites. Since we only count k-bell particles it thus follows that

$$\begin{aligned} b_n^k &= \sum_{i=q}^{r}\sum_{j=1}^{n-i+1}\frac{p_i}{n-i+1}(\delta_{ik} + b_{j-1}^k + b_{n-i-j+1}^k) \\ &= p_k + \sum_{i=q}^{r}\frac{2p_i}{n-i+1}\sum_{j=1}^{n-i} b_j^k, \end{aligned} \tag{6}$$

where δ_{ik} is the Kronecker delta. On the other hand, if $n < r$, particles of lengths $n+1,\dots,r$ do not participate in the occupation process of a $1\times n$ array and p_i in Eq. (6) must be replaced by

$$p_{i,n} = \frac{p_i}{\Sigma_{m=q}^{n} p_m}, \quad i = q,\dots,n, \quad n = q,\dots,r-1, \tag{7}$$

the (conditional) probability of choosing an i-bell particle from among particles of lengths $q,\dots,n$ and relative frequencies $p_q,\dots,p_n$. Giving an empty sum (here and in the sequel) the value zero, we may therefore write, for any $k = q,\dots,r$,

$$b_n^k = \begin{cases} 0, & \text{if } n = 1,\dots,k-1, \\ p_{k,n} + \displaystyle\sum_{i=q}^{n-k}\frac{2p_{i,n}}{n-i+1}\sum_{j=k}^{n-i} b_j^k, & \text{if } n = k,\dots,r-1, \\ p_k + \displaystyle\sum_{i=q}^{r}\frac{2p_i}{n-i+1}\sum_{j=k}^{n-i} b_j^k, & \text{if } n = r, r+1,\dots . \end{cases} \tag{8}$$

Rather similar considerations lead to the following recursion scheme giving c_n^m, the mean number of m-gaps in a saturated $1\times n$ array, for any $m = 0,1,\dots,q-1$,

$$c_n^m = \begin{cases} \delta_{nm}, & \text{if } n = 0,1,\dots,m+q-1, \\ \displaystyle\sum_{i=q}^{n-m}\frac{2p_{i,n}}{n-i+1}\sum_{j=m}^{n-i} c_j^m, & \text{if } n = m+q,\dots,r-1, \\ \displaystyle\sum_{i=q}^{r}\frac{2p_i}{n-i+1}\sum_{j=m}^{n-i} c_j^m, & \text{if } n = r, r+1,\dots, \end{cases} \tag{9}$$

where $p_{i,n}$ is the conditional probability defined in Eq. (7).

Due to Eq. (1), $\Sigma_{k=1}^{q-1} kc_n^k$ (as well as $n - \Sigma_{k=q}^{r} kb_n^k$) coincides with a_n, the mean total number of compartments ultimately remaining vacant in a $1\times n$ lattice space, and it is this quantity which has been investigated formerly.[12]

III. GENERATING FUNCTIONS

We introduce the generating functions

$$F_k(s) = \sum_{n=1}^{\infty} b_n^k s^n, \quad k = q,\dots,r \tag{10}$$

and

$$G_m(s) = \sum_{n=0}^{\infty} c_n^m s^n, \quad m = 0,1,\dots,q-1 \tag{11}$$

and set

$$\begin{aligned} \phi_k(s) = \sum_{n=k}^{r-1} b_n^k s^n - 2\sum_{i=q+k}^{r-1} s^i \sum_{j=k}^{i-q}\frac{p_{i-j}}{j+1}\sum_{n=k}^{j} b_n^k \\ + p_k s^r/(1-s), \end{aligned} \tag{12}$$

$$\psi_m(s) = \sum_{n=m}^{r-1} c_n^m s^n - 2\sum_{i=q+m}^{r-1} s^i \sum_{j=m}^{i-q}\frac{p_{i-j}}{j+1}\sum_{n=m}^{j} c_n^m, \tag{13}$$

$$g(s) = \sum_{i=q}^{r} p_i s^{i-1}, \tag{14}$$

and

$$\xi(s) = \sum_{i=1}^{q-1}\frac{s^i}{i} + \sum_{m=q}^{r-1}\frac{1}{m}s^m\sum_{j=m+1}^{r} p_j. \tag{15}$$

First starting from Eq. (9) and following the lines of thought applied previously,[12] we obtain

$$G_m(s) = 2g(s)\int_0^s \frac{G_m(x)}{1-x}\,dx + \psi_m(s)$$

or

$$G_m'(s) - G_m(s)\left[\frac{g'(s)}{g(s)} + \frac{2g(s)}{1-s}\right] = g(s)\left(\frac{\psi_m}{g}\right)'(s), \tag{16}$$

which has to be solved with initial condition $G_m(0) = \delta_{0m}$. Rather than doing this directly we substitute

$$G_m(s) = \psi_m(s) + g(s)e^{-2\xi(s)}H_m(s)(1-s)^{-2} \tag{17}$$

into Eq. (16) which yields the most simple equation

$$H_m'(s) = 2(1-s)\psi_m(s)e^{2\xi(s)}. \tag{18}$$

Since, due to Eqs. (13), (14), and (17), $\lim_{s\to 0} H_m(s) = 0$, Eq. (18) is subject to the initial condition $H_m(0) = 0$. It thus follows readily from Eq. (18) that

$$H_m(s) = 2\int_0^s (1-x)\psi_m(x)e^{2\xi(x)}\,dx$$

or

$$G_m(s) = \psi_m(s) + M_m(s)/(1-s)^2, \quad m = 0,1,\dots,q-1, \tag{19}$$

where we put

$$M_m(s) = 2g(s)e^{-2\xi(s)}\int_0^s (1-x)\psi_m(x)e^{2\xi(x)}\,dx. \tag{20}$$

Proceeding from Eq. (8) and utilizing Eqs. (10), (12), (14), and (15) we obtain quite similarly

$$F_k(s) = \phi_k(s) + L_k(s)/(1-s)^2, \quad k = q,\dots,r, \tag{21}$$

with

$$L_k(s) = 2g(s)e^{-2\xi(s)}\int_0^s (1-x)\phi_k(x)e^{2\xi(x)}\,dx. \tag{22}$$

IV. THE ASYMPTOTIC FORM OF b_n^k AND c_n^m

We put $\theta = \Sigma_{i=q}^r ip_i$ (which represents the average size of β-bell particles taking part in the occupation process) and note that [see Eqs. (14) and (15)]

$$g'(1) = \xi'(1) = \theta - 1, \quad g(1) = 1. \tag{23}$$

Upon differentiating the functions defined by Eqs. (20) and (22) and recalling Eqs. (12) and (13) we thus obtain

$$M'_m(1) = M_m(1)[1 - \theta], \quad m = 0,1,\ldots,q-1 \tag{24}$$

and

$$L'_k(1) = L_k(1)[1-\theta] + 2p_k, \quad k = q,\ldots,r. \tag{25}$$

But inspection of Eqs. (12), (13), and (19)–(22) reveals that the generating functions F_k and G_m have a pole of order 2 at 1 and we henceforth conclude that (note that ϕ_k too has a singularity at 1), as $n \to \infty$,

$$b_n^k \sim L_k(1)n + L_k(1) - L'_k(1) + p_k$$

and

$$c_n^m \sim M_m(1)n + M_m(1) - M'_m(1),$$

or, making use of Eqs. (24) and (25), as $n \to \infty$,

$$b_n^k \sim (n+\theta)L_k(1) - p_k, \quad k = q,\ldots,r \tag{26}$$

and

$$c_n^m \sim (n+\theta)M_m(1), \quad m = 0,\ldots,q-1. \tag{27}$$

Finally, upon observing that L_k and M_m are integer functions[16] of order $r-1$ [reconsider Eq. (15) to see that this affirmation is only true if $p_r > 0$; but also note that lower orders provide even smaller error terms in Eqs. (2) and (3)], the relations (2) and (3) are a consequence[16] of Eqs. (26) and (27), respectively, with

$$\lambda_k = L_k(1), \quad \lambda'_k = \theta L_k(1) - p_k, \quad k = q,\ldots,r \tag{28}$$

and

$$\rho_m = M_m(1), \quad \rho'_m = \theta M_m(1), \quad m = 0,1,\ldots,q-1. \tag{29}$$

Clearly, the just-adopted argument may be used equally well to strengthen Eq. (6) in Ref. 12 to

$$a_n = (n+\theta)L(1) + O(n^{-n(\rho-\epsilon)}), \quad \text{as } n \to \infty, \tag{30}$$

for any $0 < \epsilon < \rho = 1/(r-1)$.

V. DISPERSION OF B_n^k AND C_n^m

The variance of A_n has been found[12] to obey an asymptotic law similar in form to that of a_n, the mean of A_n [see Eq. (30)]. However, calculations proved to be rather cumbersome and the constants involved in that asymptotic formula gave somewhat unwieldy results. Concerning the dispersions of B_n^k and C_n^m the situation is not really different. We therefore desist from deriving the exact limit laws and confine ourselves to state the following: There are constants $C_1 \geqslant 0$, $C_2 \geqslant 0$, D_1, and D_2 independent of n, such that for any $0 < \epsilon < \rho = 1/(r-1)$, as $n \to \infty$,

$$\langle (B_n^k)^2\rangle - \langle B_n^k\rangle^2 = C_1 n + D_1 + O(n^{-n(\rho-\epsilon)}),$$
$$k = q,\ldots,r \tag{31}$$

and

$$\langle (C_n^m)^2\rangle - \langle C_n^m\rangle^2 = C_2 n + D_2 + O(n^{-n(\rho-\epsilon)}),$$
$$m = 0,\ldots,q-1. \tag{32}$$

Clearly, as a consequence of Eqs. (31) and (32), there is stochastic convergence ($n \to \infty$) of B_n^k/n and C_n^m/n to λ_k and ρ_m, respectively [see Eqs. (28) and (29)].

VI. PARTICLE SIZE AND GAP DISTRIBUTION

To what extent is a k-bell particle, participating in the occupation process with relative probability p_k, represented in the saturation configuration of a $1 \times n$ array? This most interesting question, on the basis of the foregoing results, is quite easily answered. Among the particles composing the saturation coverage of a $1 \times n$ lattice space a k-bell particle is found with relative frequency

$$p_{k,n}^* = \frac{b_n^k}{\Sigma_{j=q}^r b_n^j}, \quad k = q,\ldots,r, \tag{33}$$

which may be computed utilizing Eq. (8). By means of Eqs. (2), (28), and (33),

$$p_k^* \equiv \lim_{n\to\infty} p_{k,n}^* = \frac{\lambda_k}{\Sigma_{i=q}^r \lambda_i}, \quad k = q,\ldots,r, \tag{34}$$

reconfirming Eq. (4). In the following section we shall give some numerical results concerned with Eq. (34) to get an idea of the magnitude of the deviation of p_k^* from p_k.

Another question of interest is the distribution of vacant spaces on a saturated $1 \times n$ array: The relative probability of meeting with an m-gap is

$$q_{m,n} = \frac{c_n^m}{\Sigma_{i=0}^{q-1} c_n^i}, \quad m = 0,1,\ldots,q-1, \tag{35}$$

and is computable from Eq. (9) for any n. As n tends to infinity,

$$q_m \equiv \lim_{n\to\infty} q_{m,n} = \frac{\rho_m}{\Sigma_{i=0}^{q-1} \rho_i}, \quad m = 0,\ldots,q-1, \tag{36}$$

where ρ_i is given by Eqs. (20) and (29).

VII. EXAMPLES

In this section we consider a few special cases. The numerical calculations of some of the previously studied quantities may serve to get some deeper insight into the peculiarities of model II.

A. Some two-type models

(i) *The case $q=2$, $r=3$:* Since here $p_2 + p_3 = 1$, the model is completely determined by a single parameter p_3, say. We obtain from Eqs. (2), (8), (22), and (28)

$$\lim_{n\to\infty} b_n^2/n \equiv \lambda_2(p_3)$$
$$= 2\exp\{-2-p_3\}\int_0^1 (t^2 - p_3 t^3) \times \exp\{2t + p_3 t^2\}dt \tag{37}$$

and

$$\lim_{n\to\infty} b_n^3/n \equiv \lambda_3(p_3) = 2p_3\exp\{-2-p_3\} \times \int_0^1 t^3 \exp\{2t + p_3 t^2\}dt. \tag{38}$$

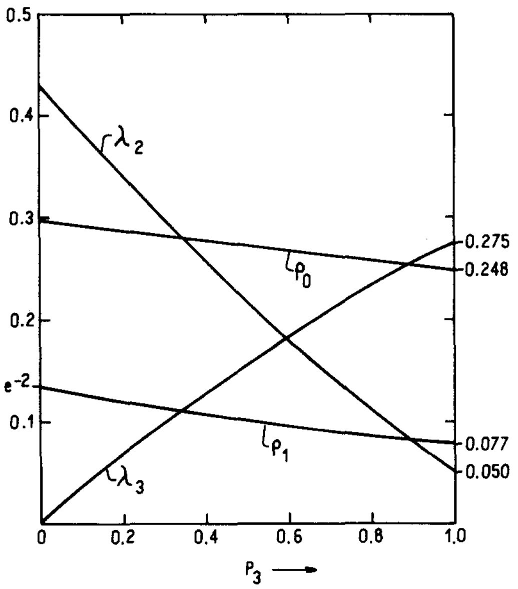


FIG. 1. Plots of λ_2 and λ_3 [Eqs. (37) and (38)] and ρ_0 and ρ_1 [Eqs. (39) and (40)] in the two-type model $q = 2$, $r = 3$ as functions of p_3.

Similarly, from Eqs. (3), (9), (20), and (29) it is seen that

$$\lim_{n\to\infty} c_n^0/n \equiv \rho_0(p_3) = 2\exp\{-2-p_3\}\int_0^1 (1-t) \times [1+2p_3t^2]\exp\{2t+p_3t^2\}dt \tag{39}$$

and

$$\lim_{n\to\infty} c_n^1/n \equiv \rho_1(p_3) = 2\exp\{-2-p_3\}\int_0^1 (1-t)t \times \exp\{2t+p_3t^2\}dt. \tag{40}$$

Figure 1 shows λ_2, λ_3, ρ_0, and ρ_1 as functions of p_3. We observe that λ_2 decreases almost linearly (with increasing p_3) to the limit value 0.050. This, clearly, is the fraction of two-gaps generated by the initially preponderating trimers and finally filled up by dimers. The fact that ρ_1 diminishes as $p_3 \to 1-$ is to be expected since the overall saturation coverage grows bigger when trimers become more frequent (see Fig. 3 in Ref. 12).

In the first two rows of Table I are given, for some values of $p \equiv p_3$, the relative frequencies $q_0(p) = \rho_0(p)/[\rho_0(p)+\rho_1(p)]$ and $q_1(p) = \rho_1(p)/[\rho_0(p)+\rho_1(p)]$ of a zero-gap and a one-gap, respectively.

(ii) *The case $q=4$, $r=5$:* In this case we see from Eqs. (20) and (27) that, for $m = 0,1,2,3$,

$$\lim_{n\to\infty} c_n^m/n \equiv \rho_m(p_5) = 2\exp\{-3-\tfrac{2}{3}-\tfrac{1}{2}p_5\} \times \int_0^1 (1-t)\psi_m(t)\exp\{2t+t^2+\tfrac{2}{3}t^3 + \tfrac{1}{2}p_5t^4\}dt,$$

with $\psi_0(t) = 1+2p_5t^4$, $\psi_1(t) = t$, $\psi_2(t) = t^2$, and $\psi_3(t) = t^3$. For some values of $p \equiv p_5$, the relative frequencies

$$q_m(p_5) = \frac{\rho_m(p_5)}{\Sigma_{i=0}^3 \rho_i(p_5)}, \quad m = 0,1,2,3 \tag{41}$$

of an m-gap are shown in the last four rows of Table I. As $p_5 \to 1-$, the frequency of m-gaps, $m = 1,2,3$, decreases; this means that the saturation coverage augments.

It is worthwhile to note that the values in the first column of Table I correspond to one-type models and, henceforth, coincide with those given by Mackenzie.[2]

B. The three-type model $q = 2$, $r = 4$

Due to the condition $p_2+p_3+p_4 = 1$ (with $p_2 > 0$) the model depends on two parameters p_3 and p_4, say. From

$$\lim_{n\to\infty} b_n^k/n \equiv \lambda_k(p_3,p_4) = 2\exp\{-2-p_3-\tfrac{5}{3}p_4\} \times \int_0^1 [p_kt^4+(1-t)f_k(t)]\exp\{2t + (p_3+p_4)t^2+\tfrac{2}{3}p_4t^3\}dt, \quad k = 2,3,4,$$

with $f_2(t) = t^2+[(1-p_3-p_4)/(1-p_4)]t^3$, $f_3(t) = [p_3/(1-p_4)]t^3$, and $f_4(t) = 0$, we calculated [see Eq. (34)]

$$p_k^*(p_3,p_4) = \frac{\lambda_k(p_3,p_4)}{\Sigma_{i=2}^4 \lambda_i(p_3,p_4)}, \quad k = 2,3,4. \tag{42}$$

Figure 2(a) shows, for some selected values of p_2, p_2^* as a function of p_4. As might be expected, whatever may be the values of p_2 and p_4, p_2^* exceeds p_2. What is remarkable is, in the case $p_2 = 0.02$, the strong increase of p_2^* as p_4 approaches

TABLE I. Relative probabilities $q_m(p)$ [Eqs. (36) and (41)] of an m-gap in three different two-type models ($q = j$, $r = j+1$; $j = 2,3,4$) for various values of $p = p_r$, $r = 3, 4, 5$.

	$p = 0.0$	$p = 0.2$	$p = 0.4$	$p = 0.6$	$p = 0.8$	$p = 1.0-$
$q = 2, r = 3$						
$q_0(p)$	0.6870	0.7042	0.7204	0.7357	0.7500	0.7633
$q_1(p)$	0.3130	0.2958	0.2796	0.2643	0.2500	0.2367
$q = 3, r = 4$						
$q_0(p)$	0.5390	0.5578	0.5758	0.5930	0.6093	0.6249
$q_1(p)$	0.2796	0.2672	0.2553	0.2440	0.2333	0.2231
$q_2(p)$	0.1813	0.1750	0.1689	0.1631	0.1574	0.1520
$q = 4, r = 5$						
$q_0(p)$	0.4500	0.4678	0.4850	0.5016	0.5176	0.5330
$q_1(p)$	0.2515	0.2421	0.2331	0.2245	0.2162	0.2083
$q_2(p)$	0.1714	0.1661	0.1610	0.1561	0.1514	0.1469
$q_3(p)$	0.1272	0.1239	0.1208	0.1177	0.1147	0.1119

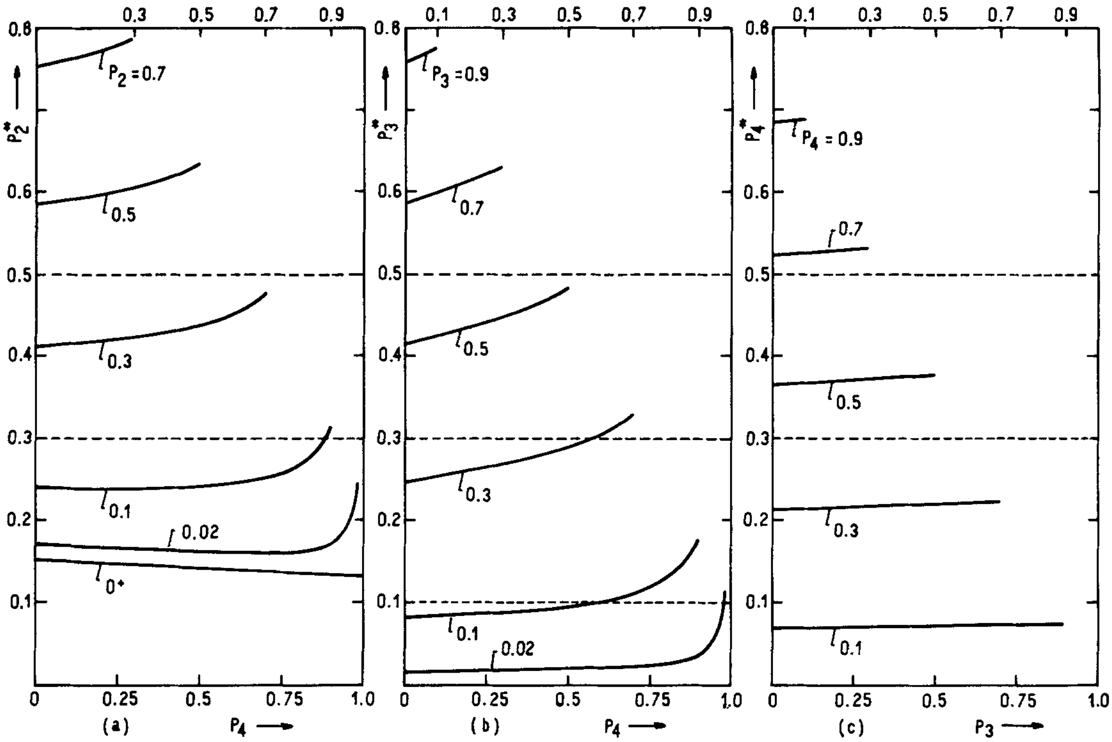


FIG. 2. Relative probabilities p_2^*, p_3^*, and p_4^* [Eq. (42)] in the three-type model $q=2$, $r=4$. (a) p_2^* as a function of p_4 for several values of p_2. (b) p_3^* as a function of p_4 for various values of p_3. (c) p_4^* as a function of p_3 for some values of p_4.

0.98. Clearly, when $p_4 \to 0.98-$, trimers become even more rare than dimers and these will then land in three-gaps, too. In the absence of trimers $p_2^*(0,0.98)=0.247$, i.e., the "output" probability p_2^* is more than 12 times larger than p_2, the "input" probability. This, clearly, is even more marked in the case $p_2 = 0+$ [see the corresponding plot of p_2^* in Fig. 2(a)].

In Fig. 2(b), for some values of p_3, p_3^* is plotted as a function of p_4. It is seen that p_3^* may be smaller and larger than p_3 and also equal to it. The latter situation takes place when $p_4 \simeq 0.59$, independently (!), as it seems, of the value of p_3.

Not surprisingly, the situation in Fig. 2(c) is contrary to that of Fig. 2(a): $p_4 > p_4^*$ for any p_3 and p_4.

C. A multitype model

In the $(r-1)$-type model $q=2$, $p_2 = p_3 = \cdots = p_r = 1/(r-1)$, all $r-1$ kinds of particles are equally frequent. For $r = 2,\ldots,11$, Table II contains the corresponding "output" probabilities p_k^*, $k=2,\ldots,r$, which manifest the smaller particles' "advantage" over larger ones. It is particularly noteworthy that, in the case $r=11$, 45% of the (ten types of) particles contributing to the saturation coverage of an infinite lattice space are dimers or trimers (on the average). On the other hand, as is easily checked from the last row in Table II, on the average only 21.86% of the total number of nonvacant compartments is occupied by dimers or trimers. The corresponding occupation percentage for 10- and 11-bell particles is almost identical: 21.49%.

VIII. CONCLUSIONS

We have analyzed the saturation configuration arising when a one-dimensional lattice space is filled sequentially at random with particles of random length. Recursion relationships for the mean number of k-bell particles placed and for

TABLE II. Relative probabilities p_k^* [Eq. (34)] in $(r-1)$-type models with $q=2$, $p_2=p_3=\cdots=p_r=1/(r-1)$, $r=2,\ldots,11$.

r	p_2^*	p_3^*	p_4^*	p_5^*	p_6^*	p_7^*	p_8^*	p_9^*	p_{10}^*	p_{11}^*
2	1.000									
3	0.585	0.415								
4	0.456	0.300	0.244							
5	0.394	0.246	0.193	0.167						
6	0.359	0.214	0.164	0.139	0.123					
7	0.336	0.194	0.146	0.121	0.106	0.096				
8	0.321	0.180	0.133	0.109	0.094	0.085	0.078			
9	0.310	0.170	0.123	0.100	0.086	0.077	0.070	0.065		
10	0.301	0.162	0.116	0.093	0.079	0.070	0.064	0.059	0.055	
11	0.294	0.156	0.110	0.088	0.074	0.065	0.059	0.054	0.051	0.048

the mean number of m-gaps have been derived and used to determine particle size and gap distribution.

ACKNOWLEDGMENTS

The author thanks Dr. E. E. Mola and J. L. Vicente for some helpful comments.

INIFTA is sponsored by the Universidad Nacional de La Plata, the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), and the Comisión de Investigaciones Científicas (Provincia de Buenos Aires).

This work was supported in part by CONICET.

[1]E. S. Page, J. R. Stat. Soc. Ser. B **21**, 364 (1959).

[2]J. K. Mackenzie, J. Chem. Phys. **37**, 723 (1962).

[3]R. B. McQuistan and D. Lichtman, J. Math. Phys. **9**, 1680 (1968).

[4]F. Gornick and J. L. Jackson, J. Chem. Phys. **39**, 1150 (1963).

[5]M. Gordon and I. H. Hillier, J. Chem. Phys. **38**, 1376 (1963).

[6]E. R. Cohen and H. Reiss, J. Chem. Phys. **38**, 680 (1963).

[7]T. H. K. Barron and E. A. Boucher, Trans. Faraday Soc. **66**, 2320 (1970).

[8]E. A. Boucher, J. Chem. Soc. Faraday Trans. 1 **68**, 2281 (1972).

[9]C. W. Lewis, J. Polym. Sci. Part. A 2 **10**, 377 (1972).

[10]E. A. Boucher, J. Chem. Phys. **59**, 3848 (1973).

[11]E. A. Boucher, J. Chem. Soc. Faraday Trans. 2 **72**, 1697 (1976).

[12]B. Mellein and E. E. Mola, J. Math. Phys. **26**, 514 (1985).

[13]P. Ney, Ph.D. thesis, Columbia University, 1960.

[14]E. A. Boucher and R. M. Nisbet, Chem. Phys. Lett. **40**, 61 (1976).

[15]B. Mellein, J. Math. Phys. (submitted).

[16]E. C. Titchmarsh, *The Theory of Functions* (Oxford U. P., Oxford, 1939).

Nonlocal Lie–Bäcklund transformations of the massive Thirring model

Paul H. M. Kersten and Ruud Martini
Twente University of Technology, Department of Applied Mathematics, P. O. Box 217, 7500 AE Enschede, The Netherlands

(Received 18 October 1984; accepted for publication 7 December 1984)

Previously we obtained second- and third-order Lie–Bäcklund transformations of the massive Thirring model. Now by introduction of nonlocal variables we obtain two (x,t)-dependent Lie–Bäcklund transformations. These nonlocal Lie–Bäcklund transformations act as generating operators on first- and second-order Lie–Bäcklund transformations, and we conjecture that they lead to two infinite hierarchies of commuting Lie–Bäcklund transformations.

I. INTRODUCTION AND GENERAL

In a recent paper[1] we studied Lie–Bäcklund transformations of the massive Thirring model.[2] This model is described by the following system of partial differential equations:

$$\begin{aligned}
&-u_{1x}+u_{1t}=mv_2-(u_2^2+v_2^2)v_1,\\
&u_{2x}+u_{2t}=mv_1-(u_1^2+v_1^2)v_2,\\
&v_{1x}-v_{1t}=mu_2-(u_2^2+v_2^2)u_1,\\
&-v_{2x}-v_{2t}=mu_1-(u_2^2+v_1^2)u_2.
\end{aligned}\tag{1.1}$$

We introduced the ideal I in $\mathbf{R}^{10}=\{(x,t,u_1,\ldots,v_2,u_{1x},\ldots,v_{2x})\}$ generated by four differential one-forms

$$\begin{aligned}
\alpha_1&=du_1-u_{1x}\,dx-G(1,1)dt,\\
\alpha_2&=du_2-u_{2x}\,dx-G(2,1)dt,\\
\alpha_3&=dv_1-v_{1x}\,dx-G(3,1)dt,\\
\alpha_4&=dv_2-v_{2x}\,dx-G(4,1)dt,
\end{aligned}\tag{1.2}$$

where $G(*,1)$ is obtained by solving for $u_{1t},\ldots,v_{2t}$ from (1.1).

The vector field V, defined by

$$V=V^3\,\partial_{u_1}+V^4\,\partial_{u_2}+V^5\,\partial_{v_1}+V^6\,\partial_{v_2}+\mathrm{pr},\tag{1.3}$$

where "pr" represents the prolongation of V^3, is a Lie–Bäcklund transformation for (1.1) if

$$\mathscr{L}_V I\subset D^jI,\tag{1.4}$$

where D^jI is the j-times prolonged ideal I (see Ref. 3).

We introduced a grading for (1.1) by setting

$$\begin{aligned}
&\deg(x)=-2,\quad \deg(t)=-2,\\
&\deg(u_i)=\deg(v_j)=1(i,j=1,2),\quad \deg(m)=2,
\end{aligned}\tag{1.5}$$

and obtained eight vector fields $X_1,\ldots,X_8$

$$\begin{aligned}
X_1^3&=\tfrac12(-mv_2+v_1R_2),\\
X_1^4&=\tfrac12(2u_{2x}-mv_1+v_2R_1),\\
X_1^5&=\tfrac12(mu_2-u_1R_2),\\
X_1^6&=\tfrac12(2v_{2x}+mu_1-u_2R_1),\\
X_2^3&=\tfrac12(2u_{1x}+mv_2-v_1R_2),\quad X_2^4=\tfrac12(mv_1-v_2R_1),\\
X_2^5&=\tfrac12(2v_{1x}-mu_2+u_1R_2),\quad X_2^6=\tfrac12(-mu_1+u_2R_1),\\
X_3^3&=u_{1x}(x+t)+mv_2x+\tfrac12u_1-xv_1R_2,\\
X_3^4&=v_{2x}(-x+t)+mv_1x-\tfrac12u_2-xv_2R_1,\\
X_3^5&=v_{1x}(x+t)-mu_2x+\tfrac12v_1+xu_1R_2,\\
X_3^6&=v_{2x}(-x+t)-mu_1x-\tfrac12v_2+xu_2R_1,\\
X_4^3&=v_1,\quad X_4^4=v_2,\quad X_4^5=-u_1,\quad X_4^6=-u_2,
\end{aligned}\tag{1.6}$$

$$\begin{aligned}
X_5^3&=\tfrac14\{2u_{2x}(-m+2v_1v_2)-4v_{2x}u_2v_1-mv_2(R_1+R_2)\\
&\quad-2mv_1R+v_1(R_2^2+2R_1R_2)\},\\
X_5^4&=\tfrac14\{-4v_{2xx}+2u_{1x}(-m+2u_1u_2)+4u_{2x}(R_1+R_2)\\
&\quad+4v_{1x}u_2v_1-mv_1(R_1+R_2)-2mv_2R\\
&\quad+v_2(R_1^2+2R_1R_2)\},\\
X_5^5&=\tfrac14\{2v_{2x}(-m+2u_1u_2)-4u_{2x}u_1v_2+mu_2(R_1+R_2)\\
&\quad+2mu_1R-u_1(R_2^2+2R_1R_2)\},\\
X_5^6&=\tfrac14\{4u_{2xx}+2v_{1x}(-m+2v_1v_2)+4v_{2x}(R_1+R_2)\\
&\quad+4u_{1x}u_1v_2+mu_1(R_1+R_2)+2mu_2R\\
&\quad-u_2(R_1^2+2R_1R_2)\}.
\end{aligned}\tag{1.7}$$

In (1.6) and (1.7) we introduced R, R_1, R_2 by

$$R=u_1u_2+v_1v_2,\quad R_1=u_1^2+v_1^2,\quad R_2=u_2^2+v_2^2.\tag{1.8}$$

The vector field X_6 is obtained from X_5 by the following transformation:

$$T{:}u_1\to u_2,\quad u_2\to u_1,\quad v_1\to v_2,\quad v_2\to v_1,\quad \partial_x\to-\partial_x,\tag{1.9a}$$

$$\begin{aligned}
X_6^3&=-T(X_5^4),\quad X_6^4=-T(X_5^3),\\
X_6^5&=-T(X_5^6),\quad X_6^6=-T(X_5^5).
\end{aligned}\tag{1.9b}$$

The explicit form of the vector fields X_7, X_8 is given for the sake of completeness in the Appendix.

In Sec. II we shall introduce nonlocal variables p_0,p_1,p_2 by prolonging the ideal D^jI with the potential forms P_0, P_1, P_2 associated with the vertical vector fields X_1, X_2, X_4. The condition (1.4) then generalizes to

$$\mathscr{L}_V I\subset\langle D^jI,P_0,P_1,P_2\rangle,\tag{1.10}$$

where $\langle D^jI,P_0,P_1,P_2\rangle$ represents the ideal generated by D^jI,P_0,P_1,P_2.

The condition (1.10) is equivalent to the one obtained by Krasilshchik and Vinogradov.[4] By assuming V to be dependent on p_0,p_1,p_2 as well, we derive two nonlocal Lie–Bäcklund transformations.

In order to compute the generalized Lie bracket[4] we first have to derive the nonlocal components of the vector fields $X_1,\ldots,X_6$, which have to satisfy the condition

$$\mathscr{L}_V P_i\subset\langle D^jI,P_0,P_1,P_2\rangle\quad(i=0,\ldots,2).\tag{1.11}$$

Finally, we give the commutators of the Lie–Bäcklund transformations, the results of which are given in Sec. III.

 0022-2488/85/071775-04$02.50

II. NONLOCAL LIE-BÄCKLUND TRANSFORMATIONS FOR THE MASSIVE THIRRING MODEL

In this section we construct nonlocal Lie–Bäcklund transformations for the massive Thirring model.

First of all we introduce a Lagrangian L for the massive Thirring model, i.e.,

$$L(u_1,\ldots,v_2,u_{1x},\ldots,v_{2t}) = \tfrac{1}{2}\{-u_1v_{1x}+u_{1x}v_1+u_2v_{2x}-u_{2x}v_2 + u_1v_{1t}-u_{1t}v_1+u_2v_{2t}-u_{2t}v_2\} + m(u_1u_2+v_1v_2)-\tfrac{1}{2}(u_1^2+v_1^2)(u_2^2+v_2^2). \tag{2.1}$$

A straightforward computation shows that the Euler–Lagrange equations associated with (2.1) are just the system of partial differential equations (1.1).

Application of Noether's theorem[5] to the infinitesimal symmetries

$$\begin{aligned} &O_1=\partial_x,\quad O_2=\partial_t,\\ &O_3=v_1\,\partial_{u_1}+v_2\,\partial_{u_2}-u_1\,\partial_{v_1}-u_2\,\partial_{v_2},\\ &O_4=t\,\partial_x+x\,\partial_t-\tfrac{1}{2}u_1\,\partial_{u_1}+\tfrac{1}{2}u_2\,\partial_{u_2}\\ &\qquad-\tfrac{1}{2}v_1\,\partial_{v_1}+\tfrac{1}{2}v_2\,\partial_{v_2}, \end{aligned} \tag{2.2}$$

which are equivalent[3] to $X_1,\ldots,X_4$ (1.6), leads to the following conserved vectors[5]:

$$\begin{aligned} A_1^x&=\tfrac{1}{2}\{u_1v_{1x}-u_{1x}v_1-u_2v_{2x}+u_{2x}v_2+R_1R_2\},\\ A_1^t&=\tfrac{1}{2}\{-u_1v_{1x}+u_{1x}v_1-u_2v_{2x}+u_{2x}v_2\},\\ A_2^x&=\tfrac{1}{2}\{u_1v_{1x}-u_{1x}v_1+u_2v_{2x}-u_{2x}v_2\},\\ A_2^t&=\tfrac{1}{2}\{-u_1v_{1x}+u_{1x}v_1+u_2v_{2x}\\ &\qquad-u_{2x}v_2-R_1R_2+2mR\},\\ A_3^x&=\tfrac{1}{2}\{R_1-R_2\},\quad A_3^t=-\tfrac{1}{2}\{R_1+R_2\},\\ A_4^x&=\tfrac{1}{2}x\{u_1v_{1x}-u_{1x}v_1+u_2v_{2x}-u_{2x}v_2\}\\ &\qquad+\tfrac{1}{2}t\{u_1v_{1x}-u_{1x}v_1-u_2v_{2x}+u_{2x}v_2+R_1R_2\},\\ A_4^t&=\tfrac{1}{2}x\{-u_1v_{1x}+u_{1x}v_1+u_2v_{2x}-u_{2x}v_2-R_1R_2\\ &\qquad+2mR\}+\tfrac{1}{2}t\{-u_1v_{1x}+u_{1x}v_1\\ &\qquad-u_2v_{2x}+u_{2x}v_2\}, \end{aligned} \tag{2.3}$$

where R, R_1, R_2 are defined by (1.8).

Our first attempt, without success, in searching for a generating Lie–Bäcklund transformation was an (x,t)-dependent local Lie–Bäcklund transformation of degree 2, because X_1, X_2 (1.6) are of degree 2, X_5, X_6 (1.7) being of degree 4. We were motivated by the form of the nonlocal Lie–Bäcklund transformations for the Korteweg–deVries (KdV) equation[4,6] ($u_t=uu_x+u_{xxx}$)

$$\psi=\{t(u_{xxxxx}+\tfrac{5}{3}u_{xxx}u+\tfrac{10}{3}u_{xx}u_x+\tfrac{5}{6}u_xu^2)+x(\tfrac{1}{3}u_{xxx} + \tfrac{1}{3}u_xu)+\tfrac{4}{3}u_{xx}+\tfrac{4}{9}u^2+\tfrac{1}{9}u_xp_{-1}\}\partial_u+\text{pr}, \tag{2.4}$$

where

$$p_{-1}=\int_{-\infty}^{x}u\,dx. \tag{2.5}$$

First, note that $\int_{-\infty}^{\infty}u\,dx$ is just a conserved quantity[3] for the KdV equation, or

$$u\,dx+(u_{xx}+\tfrac{1}{2}u^2)dt \tag{2.6}$$

is a conserved current of the KdV equation. Second, the generating (local) Lie–Bäcklund transformations of Burgers' equation[7,8]

$$\phi=\{x(2u_xu+2u_{xx})+t(4u_{xxx}+6u_{xx}u+6u_x^2 + 3u_xu^2)+u^2\}\partial_u+\text{pr} \tag{2.7}$$

and of the classical Boussinesq equation[9] ($u_t=uv_x+u_xv + \sigma v_{xxx};\ v_t=u_x+v_xv$)

$$\begin{aligned} Z_1&=\{t(\sigma u_{xxx}+\tfrac{3}{2}\sigma v_{xxx}v+3\sigma v_{xx}v_x+\tfrac{3}{4}u_x(v^2+2u)\\ &\quad+\tfrac{3}{2}v_xvu)+\tfrac{1}{2}x(\sigma v_{xxx}+u_xv+uv_x)+\tfrac{3}{2}\sigma v_{xx}+uv\}\partial_u\\ &\quad+\{t(\sigma v_{xxx}+\tfrac{3}{2}u_xv+\tfrac{3}{4}v_x(v^2+2u)+\tfrac{1}{2}x(u_x+v_xv))\\ &\quad+v^2/4+u\}\partial_v+\text{pr} \end{aligned} \tag{2.8}$$

(which is obtained by the action of the generating operator $\mathscr{D}$ on the scaling) are linear in x and t; while the coefficients of x,t are Lie–Bäcklund transformations themselves.

Motivated by these observations we introduce nonlocal variables p_0, p_1, p_2 by the potential forms

$$P_0=dp_0-p_0^1\,dx-p_0^2\,dt,\quad P_1=dp_1-p_1^1\,dx-p_1^2\,dt,$$
$$P_2=dp_2-p_2^1\,dx-p_2^2\,dt, \tag{2.9a}$$

where

$$\begin{aligned} &p_0^1=-A_3^t,\quad p_1^1=-(A_1^t+A_2^t),\quad p_2^1=-(A_1^t-A_2^t),\\ &p_0^2=+A_3^x,\quad p_1^2=A_1^x+A_2^x,\quad p_2^2=A_1^x-A_2^x. \end{aligned} \tag{2.9b}$$

We now construct the ideal of differential forms I'

$$I'=\langle D^3I,P_0,P_1,P_2\rangle \tag{2.10}$$

and impose the condition

$$\mathscr{L}_V I\subset\langle D^3I,P_0,P_1,P_2\rangle, \tag{2.11a}$$

which does lead to conditions on the local components of the vector field V, components which are supposed to be dependent on

$$p_0,\ p_1,\ p_2,\ x,\ t,\ u_1,\ldots,\ v_2,\ldots,\ u_{1xx},\ldots,\ v_{2xx}. \tag{2.11b}$$

The resulting conditions on the local components of V are similar to the conditions obtained by prolongation of the total derivative operators D_x, D_t, i.e., $\tilde{D}_x$, $\tilde{D}_t^4$.

Motivated by the results for the KdV, Burgers', and classical Boussinesq equations we search for a Lie–Bäcklund transformation

$$V=x\,\text{LB}_1+t\,\text{LB}_2+C, \tag{2.12}$$

where LB_1 and LB_2 are (x,t)-independent Lie–Bäcklund transformations of degree $\leqslant 4$, while C has to be of degree 2. (Note that the components have to be of degrees 5 and 3, respectively.) Since in this specific problem the mass m is of degree 2, we take LB_1, LB_2 to be linear combinations of $X_1,\ldots,X_6$ [(1.6) and (1.7)] whereas the C components (2.12) are supposed to be linear in u_{1x}, u_{2x}, v_{1x}, v_{2x}.

Substitution of (2.12) into the overdetermined system of partial differential equations obtained from (2.11) and solving the resulting system leads to *two* (x,t)-dependent nonlocal Lie–Bäcklund transformations, i.e.,

$$Z_1^3 = v_1 p_2 + x[-2X_5^3 - m^2 v_1] + t[2X_5^3] + \tfrac{1}{2} m u_2,$$
$$Z_1^4 = v_2 p_2 + x[-2X_5^4 - m^2 v_2] + t[2X_5^4] + \tfrac{3}{2} m u_1 + 3v_{2x} - \tfrac{3}{2} R_1 u_2 - \tfrac{1}{2} R_2 u_2,$$
$$Z_1^5 = -u_1 p_2 + x[-2X_5^5 + m^2 u_1] + t[2X_5^5] + \tfrac{1}{2} m v_2,$$
$$Z_1^6 = -u_2 p_2 + x[-2X_5^6 + m^2 u_2] + t[2X_5^6] + \tfrac{3}{2} m v_1 - 3u_{2x} - \tfrac{3}{2} R_1 v_2 - \tfrac{1}{2} R_2 v_2, \tag{2.13}$$
$$Z_2^3 = v_1 p_1 + x[-2X_6^3 + m^2 v_1] + t[-2X_6^3] + \tfrac{3}{2} m u_2 - 3v_{1x} - \tfrac{3}{2} R_2 u_1 - \tfrac{1}{2} R_1 u_1,$$
$$Z_2^4 = v_2 p_1 + x[-2X_6^4 + m^2 v_2] + t[-2X_6^4] + \tfrac{1}{2} m u_1,$$
$$Z_2^5 = -u_1 p_1 + x[-2X_6^5 - m^2 u_1] + t[-2X_6^5] + \tfrac{3}{2} m v_2 + 3u_{1x} - \tfrac{3}{2} R_2 v_1 - \tfrac{1}{2} R_1 v_1,$$
$$Z_2^6 = -u_2 p_1 + x[-2X_6^6 - m^2 u_2] + t[-2X_6^6] + \tfrac{1}{2} m v_1.$$

Note that the local components of the vector fields Z_1 and Z_2 do not depend on the nonlocal variable p_0. From now on we discard p_0 from our considerations.

III. THE ACTION OF THE VECTOR FIELDS Z_1 AND Z_2 ON $X_1,\ldots, X_6$

In order to derive the action of the vector fields Z_1 and Z_2 on the vector fields $X_1,\ldots, X_6$ [(2.13), (1.6), and (1.7)] we have to extend the Lie bracket in a way analogous to Krasilshchik and Vinogradov.[4] The nonlocal components of the vector fields $X_1,\ldots, X_6$ are obtained by prolongation, expressed by the condition

$$\mathscr{L}_{X_j} P_i \subset \langle D^3 I, P_1, P_2\rangle, \quad (i=1,2) \quad (j=1,\ldots,6). \tag{3.1}$$

This condition is equivalent to the condition that the Lie derivative $\mathscr{L}_{X_j}$ of the potential equations

$$p_{1x} + (A_1^t + A_2^t) = 0, \quad p_{1t} - (A_1^x + A_2^x) = 0,$$
$$p_{2x} + (A_1^t - A_2^t) = 0, \quad p_{2t} - (A_1^x - A_2^x) = 0 \tag{3.2}$$

is zero subject to (1.1) and (3.2) and their differential consequences.[5] We shall not take into account integration constants arising from condition (3.11). They refer to symmetries $\partial_{p_1}, \partial_{p_2}$ of (1.1) and (3.2). The computation of (3.1) leads to the following nonlocal components of the vector fields $X_1,\ldots, X_6$:

$$X_j = X_j^1 \partial_{p_1} + X_j^2 \partial_{p_2} + X_j^3 \partial_{u_1} + X_j^4 \partial_{u_2} + X_j^5 \partial_{v_2} + X_j^6 \partial_{v_2} + \mathrm{pr} \quad (j=1,\ldots,6), \tag{3.3a}$$

$$X_1^1 = -\tfrac{1}{2} mR,$$
$$X_1^2 = -v_2 u_{2x} + u_2 v_{2x} + \tfrac{1}{2} mR - \tfrac{1}{2} R_1 R_2,$$
$$X_2^1 = -v_1 u_{1x} + u_1 v_{1x} - \tfrac{1}{2} mR + \tfrac{1}{2} R_1 R_2,$$
$$X_2^2 = \tfrac{1}{2} mR,$$
$$X_3^1 = \tfrac{1}{2}(x+t)(+2u_1 v_{1x} - 2v_1 u_{1x} + R_1 R_2) - mtR + p_1,$$
$$X_3^2 = \tfrac{1}{2}(x+t)(-2u_2 v_{2x} + 2v_2 u_{2x} + R_1 R_2) + mtR - p_2,$$
$$X_4^1 = 0, \quad X_4^2 = 0,$$
$$X_5^1 = -\tfrac{1}{2} m v_1 u_{2x} + \tfrac{1}{2} m u_1 v_{2x} - \tfrac{1}{4} mR(R_1 + R_2) + \tfrac{1}{4} m^2 (R_1 + R_2),$$
$$\begin{aligned} X_5^2 = {} & + u_{2xx} u_2 + v_{2xx} v_2 - u_{2x}^2 - v_{2x}^2 \\ & - \tfrac{1}{2} m u_2 v_{1x} + m v_1 u_{2x} + \tfrac{1}{2} m v_2 u_{1x} - m u_1 v_{2x} \\ & - u_{2x}(R_2 v_2 + 2R_1 v_2) + v_{2x}(R_2 u_2 + 2R_1 u_2) \\ & - \tfrac{1}{4} m^2 (R_1 + R_2) + \tfrac{3}{4} mR(R_1 + R_2) \\ & + \tfrac{1}{2} R_1 R_2 (R_1 + R_2), \end{aligned} \tag{3.3b}$$
$$\begin{aligned} X_6^1 = {} & - u_{1xx} u_1 - v_{1xx} v_1 + v_{1x}^2 + u_{1x}^2 \\ & - \tfrac{1}{2} m v_{2x} u_1 + m u_{1x} v_2 + \tfrac{1}{2} m u_{2x} v_1 - m v_{1x} u_2 \\ & - u_{1x} v_1 (R_1 + 2R_2) + v_{1x} u_1 (R_1 + 2R_2) \\ & + \tfrac{1}{4} m^2 (R_1 + R_2) - \tfrac{3}{4} mR(R_1 + R_2) \\ & - \tfrac{1}{2} R_1 R_2 (R_1 + R_2), \end{aligned}$$
$$X_6^2 = -\tfrac{1}{2} m v_2 u_{1x} + \tfrac{1}{2} m u_2 v_{1x} + \tfrac{1}{4} mR(R_1 + R_2) - \tfrac{1}{4} m^2 (R_1 + R_2),$$

while the p_1 component of Z_1 and the p_2 component of Z_2 are given by

$$Z_1^1 = \tfrac{1}{2}(x-t)\{-2m u_1 v_{2x} + 2m v_1 u_{2x} + (-m^2 + mR)(R_1 + R_2)\} - \tfrac{1}{2} m u_1 v_2 + \tfrac{1}{2} m u_2 v_1,$$
$$Z_2^2 = \tfrac{1}{2}(x+t)\{-2m u_2 v_{1x} + 2m v_2 u_{1x} + (m^2 - mR)(R_1 + R_2)\} + \tfrac{1}{2} m u_1 v_2 - \tfrac{1}{2} m u_2 v_1. \tag{3.4}$$

Computation of the generalized Lie bracket then leads to the following results:

$$[Z_1, X_1] = -\tfrac{1}{2} m^2 X_4 - 2X_5, \quad [Z_2, X_1] = \tfrac{1}{2} m^2 X_4,$$
$$[Z_1, X_2] = -\tfrac{1}{2} m^2 X_4, \quad [Z_2, X_2] = \tfrac{1}{2} m^2 X_4 - 2X_6,$$
$$[Z_1, X_3] = Z_1, \quad [Z_2, X_3] = -Z_2, \tag{3.5}$$
$$[Z_1, X_4] = 0, \quad [Z_2, X_4] = 0,$$
$$[Z_1, X_5] = 4X_7 - 2m^2 X_1 - m^2 X_2, \quad [Z_2, X_5] = m^2 X_1,$$
$$[Z_1, X_6] = m^2 X_2, \quad [Z_2, X_6] = 4X_8 - m^2 X_1 - 2m^2 X_2,$$

while

$$[Z_1, Z_2] = -2m^2 X_3.$$

Transformation of the basis vector fields by[1]

$$Y_1 = X_1, \quad Y_2 = X_2, \quad Y_3 = X_3, \quad Y_4 = X_4,$$
$$Y_5 = X_5 + (m^2/4) X_4, \quad Y_6 = X_6 - (m^2/4) X_4, \tag{3.6}$$
$$Y_7 = Y_7 - (m^2/2) X_1 - (m^2/4) X_2,$$
$$Y_8 = X_8 - (m^2/4) X_1 - (m^2/2) X_2$$

yields the following commutators:

$$\begin{aligned}
&[Z_1, Y_1] = -2Y_5, && [Z_2, Y_1] = \tfrac{1}{2} m^2 Y_4,\\
&[Z_1, Y_2] = -\tfrac{1}{2} m^2 Y_4, && [Z_2, Y_2] = -2Y_6,\\
&[Z_1, Y_3] = Z_1, && [Z_2, Y_3] = -Z_2,\\
&[Z_1, Y_4] = 0, && [Z_2, Y_4] = 0,\\
&[Z_1, Y_5] = 4Y_7, && [Z_2, Y_5] = m^2 Y_1,\\
&[Z_1, Y_6] = m^2 Y_2, && [Z_2, Y_6] = 4Y_8,
\end{aligned} \tag{3.7}$$

while

$$[Z_1, Z_2] = -2m^2 Y_3.$$

From (3.7) we conclude that Z_1 acts as a generating operator on Y_1, Y_5, while Z_2 acts as a generating operator on Y_2, Y_6. The action of Z_1 on Y_2, Y_6 is of decreasing nature, just as Z_2 acts on Y_1, Y_5.

We suspect that the vector fields Z_1 and Z_2 generate a hierarchy of commuting Lie–Bäcklund transformations. At this moment we do not have a general proof of this fact.

Remark: in (3.7) only Z_1^1, Z_2^2 are given, necessary to compute the generalized Lie bracket

$$[Z_1, Z_2] = -2m^2 Y_3.$$

We should mention that Z_1 does not admit a prolongation Z_1^2, while Z_2 does not admit a Z_2^1 prolongation in this formulation (3.1). They probably do admit a prolongation in a more general formulation, taking into account higher-order nonlocal variables related to the Lie–Bäcklund transformations Y_5, Y_6 (1.7). We hope to study this problem in future work.

APPENDIX: LIE-BÄCKLUND TRANSFORMATIONS X_7, X_8

The vector field X_7 is given[1] by

$$\begin{aligned}
X_7^3 = \tfrac{1}{8}\{&8u_{2xx}u_2v_1 + 4v_{2xx}(2v_1v_2 - m) - 4u_{2x}^2v_1 + 4u_{2x}(m(R_1 + R_2 + v_1^2 + v_2^2) - 3v_1v_2(R_1 + R_2)) - 4v_{2x}^2v_1\\
&+ 4v_{2x}(-(u_1v_1 + u_2v_2)m + 3u_2v_1(R_1 + R_2)) + 4u_{1x}mR - 2m^2v_1(R_1 + R_2) - 4v_2m^2R + 4v_1mR(R_1 + 2R_2)\\
&+ v_2m(R_1^2 + 4R_1R_2 + R_2^2) - v_1(R_2^3 + 6R_2^2R_1 + 3R_2R_1^2)\},\\
X_7^4 = \tfrac{1}{8}\{&8u_{2xxx} + 12v_{2xx}(R_1 + R_2) + 8u_{1xx}u_1v_2 + 4v_{1xx}(2v_1v_2 - m) - 12u_{2x}^2v_2 + 24u_{2x}v_{2x}u_2\\
&+ 2u_{2x}(10mR - 3R_1^2 - 12R_1R_2 - 3R_2^2) + 12v_{2x}^2v_2 + 24v_{2x}u_{1x}u_1 + 24v_{2x}v_{1x}v_1 + 8u_{1x}^2v_2\\
&+ 4u_{1x}(m(R_1 + R_2 + u_1^2 + u_2^2) - 3u_1u_2(R_1 + R_2)) + 8v_{1x}^2v_2 + 4v_{1x}(m(u_1v_1 + u_2v_2) - 3u_2v_1(R_1 + R_2)) - 4m^2v_1R\\
&- 2m^2v_2(R_1 + R_2) + mv_1(R_2^2 + 4R_1R_2 - R_1^2) + 4mv_2R(R_2 + 2R_1) - v_2(R_1^3 + 6R_1^2R_2 + 3R_1R_2^2)\},\\
X_7^5 = \tfrac{1}{8}\{&-8v_{2xx}v_2u_1 - 4v_{2xx}(u_1u_2 - m) + 4v_{2x}^2u_1 + 4v_{2x}(m(R_1 + R_2 + u_1^2 + u_2^2) - 3u_1u_2(R_1 + R_2)) + 4u_{2x}^2u_1\\
&+ 4u_{2x}(-(u_1v_1 + u_2v_2)m + 3v_2u_1(R_1 + R_2)) + 4v_{1x}mR + 2m^2u_1(R_1 + R_2) + 4u_2m^2R - 4u_1mR(R_1 + 2R_2)\\
&- u_2m(R_1^2 + 4R_1R_2 + R_2^2) + u_1(R_2^3 + 6R_2^2R_1 + 3R_2R_1^2)\},\\
X_7^6 = \tfrac{1}{8}\{&8v_{2xxx} - 12u_{2xx}(R_1 + R_2) + 8v_{1xx}u_2v_1 - 4u_{1xx}(2u_1u_2 - m) + 12v_{2x}^2u_2 - 24u_{2x}v_{2x}v_2\\
&+ 2v_{2x}(10mR - 3R_1^2 - 12R_1R_2 - 3R_2^2) - 12u_{2x}^2u_2 - 24u_{2x}v_{1x}v_1 + 24u_{2x}u_{1x}u_1\\
&- 8v_{1x}^2u_2 + 4v_{1x}(m(R_1 + R_2 + v_1^2 + v_2^2) - 3v_1v_2(R_1 + R_2)) - 8u_{1x}^2u_2 + 4u_{1x}(m(u_1v_1 + u_2v_2) - 3u_1v_2(R_1 + R_2))\\
&+ 4m^2u_1R + 2m^2u_2(R_1 + R_2) - mu_1(R_2^2 + 4R_1R_2 + R_1^2) - 4mu_2R(R_2 + 2R_1)\\
&+ u_2(R_1^3 + 6R_1^2R_2 + 3R_1R_2^2)\},
\end{aligned} \tag{A1}$$

while X_8 is obtained from X_7 by transformation (1.9a) and

$$X_8^3 = -T(X_7^4),\ X_8^4 = -T(X_7^3),\quad X_8^5 = -T(X_7^6),\ X_8^6 = -T(X_7^5). \tag{A2}$$

[1] P. H. M. Kersten and R. Martini, J. Math. Phys. **26**, 822 (1985).
[2] D. J. Kaup and A. C. Newell, Lett. Nuovo Cimento **20**, 325 (1977).
[3] F. A. E. Pirani, D. C. Robinson, and W. F. Shadwick, "Local Jet Bundle Formulation of Bäcklund Transformations," in *Mathematical Physics Studies* (Reidel, Dordrecht, Netherlands, 1979).
[4] I. S. Krasilshchik and A. M. Vinogradov, Acta Appl. Math. **2**, 79 (1984).
[5] R. L. Anderson and N. H. Ibragimov, Lie–Bäcklund Transformations in Applications, SIAM Studies in Applied Mathematics (SIAM, Philadelphia, 1979).
[6] H. M. M. Ten Eikelder, Ph.D. thesis, University of Eindhoven, The Netherlands, 1984.
[7] P. H. M. Kersten, "Lie–Bäcklund symmetries for the massive Thirring model and Burgers' equation," Memorandum 469, Twente University of Technology, The Netherlands, 1984.
[8] A. M. Vinogradov, Acta Appl. Math. **2**, 21 (1984).
[9] M. Ito, Phys. Lett. A **104**, 248 (1984).

Coulomb potential envelopes for a relativistic fermion in a central field

Richard L. Hall
Department of Mathematics, Concordia University, 1455 de Maisonneuve Blvd. W., Montreal, Quebec, Canada H3G 1M8

(Received 10 October 1984; accepted for publication 15 February 1985)

We consider a hydrogenlike system in which the Coulomb potential is replaced by the more general central potential $V(r) = vf(r/b) = vg(-b/r)$, where g is monotone increasing and convex. The method of potential envelopes is applied to this problem and approximations are obtained for the energy trajectories based on the expression $\epsilon_{nj} = \min_{u \in (0,1)} [D_{nj}(u) - uD'_{nj}(u) + vf\{-1/bD'_{nj}(u)\}]$, where $D_{nj}(u)$ is the known exact trajectory function for the hydrogenic atom. General formulas are given for linear combinations of power-law potentials and the log potential. Some graphical results are presented in the case of the Coulomb-plus-linear potential $f(r) = -\alpha/r + \beta r$.

I. INTRODUCTION

We study the bound-state energies of a single fermion which moves in an attractive central potential and obeys the Dirac equation. The potential $V(r)$ is the time component of a Lorentz four-vector, just like the Coulomb potential of the hydrogenic problem. The only Lorentz-scalar term in the Hamiltonian is the term proportional to the constant rest mass m. We suppose that the potential $V(r)$ is a smooth transformation of the Coulomb potential $-a/r$ and therefore has the representation

$$V(r) = V_0 f(r/a) = V_0 g(-a/r), \quad r = |\mathbf{r}|, \tag{1.1}$$

where the transformation function g is monotone increasing and convex on $(-\infty, 0)$. This class of potentials includes, for example, combinations of Coulomb, linear, and logarithmic components such as

$$V(r) = -A/x + Bx + C\log(x), \quad x = r/a, \tag{1.2}$$

which are of interest in connection with the construction of quark models for baryons. In the relativistic domain such central vector potentials do not lead to confinement. Suitable increasing scalar potentials with $m = m(r)$ are sufficient for confinement but we do not treat these in the present study: in this paper m is a constant equal to the rest mass of the particle.

The close relationship of the general class (1.1) of potentials to the Coulomb potential allows us to take advantage of the well-known exact solution to Dirac's equation for the hydrogenic atom. Every smooth potential of the form (1.1) is the envelope of a family of Coulomb potentials, each having the form

$$f^{(t)}(x) = -P(t)/x + Q(t), \quad P(t) > 0, \quad Q(t) > 0, \tag{1.3}$$

where t is a parameter which we shall take to be the point of contact between the tangential potential $f^{(t)}(x)$ and the original potential $f(x)$. The convexity of g implies that $f^{(t)}(x) \leqslant f(x)$, for each $t \in (0, \infty)$. Since the Dirac eigenvalue problems generated by the envelope "components" $f^{(t)}(x)$ have exact solutions, we can try to exploit these to obtain approximations to the Dirac eigenvalues generated by $f(x)$. This application of "envelope representations" in the analysis of potentials is called the method of potential envelopes. It was first introduced[1] as a tool for the nonrelativistic N-body problem and has subsequently been further refined[2] by the use of "kinetic potentials." The main purpose of the present article is to explore the use of this geometrical approach in the study of relativistic problems. In this endeavor, two new difficulties immediately arise.

The first of these is a mathematical problem. We should like to conjecture that if $f^{(t)}(x) \leqslant f(x)$, then the eigenvalues associated with $f^{(t)}$ are, one by one, lower than those associated with f. However, if we open standard modern reference works such as Reed–Simon[3] or Thirring,[4] we do not find the equivalent of the Rayleigh–Ritz theorem for the Dirac Hamiltonian. The reason for this is that the Dirac operator is not bounded below and so we do not have the equivalent of the variational principle for eigenvalues that is so useful for the corresponding nonrelativistic problem. Elementary variational arguments leading to the Dirac equation may be found, for example, in articles by Swirles and by Hartree.[5] Throughout the present article we keep the Lorentz scalar term (which is proportional to the mass m) constant. It is clear that the eigenvalues of our problem are monotonic in any added *constant* term in the potential $V(r)$. In particular, the spectrum of the soluble hydrogenic problem associated with the potential $V(r) = -A/r + B$ is monotone in the constant B, and we know that the discrete energies are also monotone in the positive constant A: within this very restricted family of potentials, therefore, if $V(r)$ is increased by adjusting either A or B, then the discrete eigenvalues increase. Kato has shown[6] that the Dirac Hamiltonian is essentially self-adjoint for hydrogenlike potentials that are not too strong (i.e., $Z < 137 \times 2/\pi$). Wightman has shown[7] that the Dirac Hamiltonian has some nice stability properties under smooth local perturbations. We can add to this the very weak argument that we have not been able to find a counterexample to the conjecture as a result of some numerical investigations. However, all these observations do not bring us close to a theorem and indeed it may turn out that no such general theorem is possible which would include all the problems we treat. Our policy will therefore be to use the envelope method simply as a guide which will lead us towards an approximation for the eigenvalues based on the

 0022-2488/85/071779-07$02.50

known hydrogenic spectrum. We do not claim that the approximate energies are lower bounds to the discrete eigenvalues.

We now come to our second difficulty. In our proposed application of the potential envelope method the envelope components $f^{(t)}(x)$ have the term $-P(t)/x$. Now in order to use the known exact solution for the hydrogen atom we must have $P(t) < P_0$, where the constant P_0 depends on the units we employ but corresponds to "$Z = 137$" (or more strictly, "$Z = 137\times 2/\pi$") in the atomic case. This means, for example, that even if we apply our results to a pure linear potential, the coupling constant should not be too large. The representation we use for the potential therefore automatically eliminates both singularities stronger than $1/r$, and large coupling to increasing potentials. It is perhaps mildly curious that a restriction which starts out merely as a pathology of the representation turns out also to be consistent with what is likely to be physically meaningful in a non-field-theoretic framework.

II. DIMENSIONS AND SCALING

Rather than set various physical quantities initially to the value 1, we adopt explicit dimensionless variables according to the following policy. All energies are measured in units of mc^2, where m is the rest mass of the particle, and all lengths are measured in terms of the Compton wavelength $\hbar/mc$. Thus if E is the total energy of the particle which moves in the potential

$$V(r) = V_0 f(r/a), \tag{2.1}$$

then we define the following dimensionless variables:

$$\epsilon = E/mc^2,\quad v = V_0/mc^2,\quad b = amc/\hbar,$$
$$z = rmc/\hbar,\quad x = r/a = z/b. \tag{2.2}$$

In terms of these variables we now write the coupled radial equations (we follow, essentially, the notation of the book by Messiah,[8] p. 928):

$$\left\{\frac{-d}{dz} + \frac{\tau k}{z}\right\}\psi_2(z) = \left\{\epsilon - 1 - vf\left(\frac{z}{b}\right)\right\}\psi_1(z),$$
$$\left\{\frac{d}{dz} + \frac{\tau k}{z}\right\}\psi_1(z) = \left\{\epsilon + 1 - vf\left(\frac{z}{b}\right)\right\}\psi_2(z), \tag{2.3}$$

where the parity of the spinor Ψ^P_{jM} constructed from the large and small radial factors $z^{-1}\psi_1(z)$ and $z^{-1}\psi_2(z)$ is given by $P = (-1)^{j+(1/2)\tau}$, and $\tau = \pm 1$. We shall refer to the allowed values of $\epsilon = \epsilon_{nj}$ in (2.3) as "Dirac eigenvalues." Our radial quantum number $n = 1,2,3...$ indicates, for a given j, the nth solution of (2.3) for the radial functions (ψ_1,ψ_2). The eigenvalues are ordered according to $\epsilon_{n'j} \geqslant \epsilon_{nj}$, $n' > n$, and both possible values of the parity parameter τ are allowed in the collection of ordered eigenvalues. Although, of course, the parity $P = (-1)^{j+(1/2)\tau}$ is a constant of the motion its usefulness to us is limited by our present methodology in which the more general problem is analysed in terms of the hydrogenic atom. With this labeling convention, the degeneracy of the eigenvalue ϵ_{nj} is always exactly $(2j+1)$. The familiar hydrogen-atom eigenvalues will be given below in the same notation.

It is clear from (2.3) that, unlike the corresponding nonrelativistic problem, we cannot in general rescale the equations in order to express the energies ϵ as functions of v alone: this is because the solution to the relativistic problem depends on c whose value in turn depends on the length units used. Hence, for the general problem we must write

$$\epsilon = F(v,b). \tag{2.4}$$

For pure power laws with potential shapes

$$f(x) = \mathrm{sgn}(q)x^q,\quad q \geqslant -1,\quad q \neq 0, \tag{2.5}$$

we have the functional form

$$\epsilon = F(\gamma),\quad \text{where } \gamma = vb^{-q}, \tag{2.6}$$

whereas, for the log potential

$$f(x) = \log(x), \tag{2.7}$$

we obtain

$$\epsilon = F(v) - v\log(b). \tag{2.8}$$

The functions $F = F_{nj}$ above are, of course, all different. We have not been able to learn any more about these functions by scaling arguments although we feel that such reasoning should yield more information, as indeed it does in the nonrelativistic case.[2] For a combination of the Coulomb and linear potentials we shall write the potential in the form

$$f(x) = -A/x + Bx,\quad \text{with } v = 1. \tag{2.9}$$

The radial equations (2.3) then imply the relation

$$\epsilon = F(\alpha,\beta),\quad (\alpha,\beta) = (Ab, B/b). \tag{2.10}$$

Again, it does not appear to be possible to simplify (2.10) further by the use of general scaling arguments. However, in the nonrelativistic limit the usual scaling law must emerge: we shall return to this point later when we discuss nonrelativistic limits in Sec. III.

A function which will be frequently used in this article is the F function for the Coulomb potential $f(x) = -1/x$. In this case we let $vb = u$ so that $\epsilon = F(vb) = F(u) = D(u)$ is given[8] (with our convention for the radial quantum number n) by the expression

$$\begin{aligned}\epsilon_{nj} &= D_{nj}(u) = D(u)\\ &= [1 + u^2\{\nu - k + (k^2 - u^2)^{1/2}\}^{-2}]^{-1/2},\\ k &= j + \tfrac{1}{2},\quad \nu = I(n/2) + k,\end{aligned} \tag{2.11}$$

where $I(X)$ is the greatest integer $\leqslant X$ and ν is the symbol which we shall use for the principal quantum number of the hydrogenic atom, and $u < 1$ (the "$Z < 137$" limit for relativistic hydrogenic atoms). We recall that all the eigenvalues so labeled have degeneracy exactly $(2j+1)$. We are not employing a parity label: the parity P of the spinor corresponding to the (hydrogenic) eigenvalue ϵ_{nj} is given by $\tau = (-1)^n$ and $P = (-1)^{j+(1/2)\tau}$. The imposition of this relationship between τ and n (which, by our definition, "counts" *all* the radial states with label JM) is possible because of the degeneracies of the hydrogen spectrum. In the nonrelativistic limit the "large component" of the Dirac spinor becomes dominant and we shall denote the orbital angular momentum label of the sperical harmonic in this component by l. In terms of l the parity of the spinor may be written $P = (-1)^l$. In the nonrelativistic limit one also uses a radial quantum number

$n_r = 1,2,3,\ldots$. We now give the explicit relationships between all these various quantum numbers for the hydrogenic spectrum:

$$\nu = I(n/2) + j + \tfrac{1}{2}$$
$$= \{n - I(n/2)\} + \{j + \tfrac{1}{2} + 2I(n/2) - n\} \qquad (2.12)$$
$$= n_r + l, \quad k = j + \tfrac{1}{2}.$$

It follows from (2.12) that $l = j + \frac{1}{2}$ when n is even and $l = j - \frac{1}{2}$ when n is odd.

We shall also need the derivative $D'(u)$ of the function $D(u)$. This function is given by

$$D'(u) = -u\{\nu - k + k^2(k^2 - u^2)^{-1/2}\}$$
$$\times [u^2 + \{\nu - k + (k^2 - u^2)^{1/2}\}^2]^{-3/2} < 0. \qquad (2.13)$$

We note in passing that it follows from (2.13) that $D''(u) < 0$ for $u \in (0,k)$. In summary, then, $D(u)$ is a positive, monotone decreasing, and concave function of $u \in (0,k)$; $D(u)$ is physically meaningful only for $u < 1$. A useful approximation for $D(u)$ for $u^2 \ll 1$ is given by the following (well-known) partial Taylor expansion in terms of u^2:

$$D(u) \sim 1 - (u^2/2\nu^2)[1 + \{\nu/k - \tfrac{3}{4}\}u^2/\nu^2],$$
$$\nu = I(n/2) + k. \qquad (2.14)$$

III. NONRELATIVISTIC LIMITS

Nonrelativistic scaling laws must emerge from the relativistic formulas in the appropriate limit as $c \to \infty$. We now look at this question. Schrödinger's equation for the central field problem may be written as

$$-(\hbar^2/2m)\Delta\psi + V_0 f(r/a)\psi = E^*\psi, \qquad (3.1)$$

where E^* is the energy and the asterisk is a label which we shall use for nonrelativistic quantities. In terms of the dimensionless variables defined by

$$v^* = 2mV_0a^2/\hbar^2 = 2vb^2, \quad \epsilon^* = 2mE^*a^2/\hbar^2, \qquad (3.2)$$

Schrödinger's equation then takes the standard form

$$\{-\Delta + v^*f(x)\}\psi = \epsilon^*\psi. \qquad (3.3)$$

The functions which give the energy trajectories[1,2] are then defined by

$$\epsilon^* = F^*(v^*). \qquad (3.4)$$

From (3.2) and the corresponding equation (2.2) for the relativistic case it is clear that the required limit is as follows:

$$c \to \infty \Rightarrow 2\{\epsilon - 1\}b^2 \to \epsilon^*. \qquad (3.5)$$

The details will, of course, depend on the potential shape f. We first look at the power-law potentials $f(x) = \mathrm{sgn}(q)x^q$, $q \geqslant -1, q \neq 0$, for which we have shown in Sec. II that $\epsilon = F(\gamma)$, where $\gamma = vb^{-q}$. Because $q \geqslant -1$ we have that $c \to \infty \Rightarrow b \to \infty \Rightarrow \gamma = \gamma^* b^{-(2+q)} \to 0$ for a given value of γ^*. The expression which should approach ϵ^* in this case is therefore given by

$$2\{\epsilon - 1\}b^2 = \{v^*\}^{2/(2+q)}[2\{F(\gamma) - 1\}/(2\gamma)^{2/(2+q)}]. \qquad (3.6)$$

Meanwhile, the nonrelativistic scaling law[1] is given by $F^*(v^*) = F^*(1)\{v^*\}^{2/(2+q)}$. It follows therefore that the functions F and F^* are related by

$$F^*(1) = \lim_{\gamma \to 0} [2(F(\gamma) - 1)/(2\gamma)^{2/(2+q)}]. \qquad (3.7)$$

As an illustration of the limit (3.7) we consider the Coulomb case $q = -1$. The expression on the right-hand side of (3.7) becomes $\frac{1}{2}[(D(\gamma) - 1)\gamma^{-2}]$ whose limiting value $-1/4\nu^2$ is immediately obtained from Eq. (2.14). Thus we have $\epsilon^* = -\{v^*\}^2/4\nu^2$, the familiar hydrogenic energy formula.

For the logarithmic potential $f(x) = \log(x)$ we know from (2.8) that the relativistic energies are given by $\epsilon = f(v) - v\log(b)$. The nonrelativistic energy trajectories, on the other hand, are known[9] to have the form $\epsilon^* = F^*(v^*) = -\frac{1}{2}v^* \log\{v^*\} + v^*F^*(1)$. Since $c \to \infty \Rightarrow v \to 0$, we obtain for the logarithmic potential the limit

$$F^*(1) = \lim_{v \to 0}[\{F(v) - 1\}v^{-1} + \tfrac{1}{2}\log(2v)]. \qquad (3.8)$$

Values for $F^*_{nj}(1)$ may be found in the third article of Ref. 2.

Lastly we look at the Coulomb-plus-linear potential with shape $f(x) = -A/x + Bx$, and $v = 1$. For the relativistic problem we defined $\alpha = Ab$ and $\beta = B/b$, and we found that $\epsilon = F(\alpha,\beta)$. Setting $v = 1$ means that $V_0 = mc^2$. Consequently, if the nonrelativistic Hamiltonian H^*, whose eigenvalues ϵ^* we seek, is written in the form

$$H^* = -\Delta - \alpha^*/x + \beta^*x, \qquad (3.9)$$

then we must have the following correspondences between the potential coefficients:

$$\alpha^* = 2Ab^2 = 2\alpha b, \quad \beta^* = 2Bb^2 = 2\beta b^3. \qquad (3.10)$$

The eigenvalues of H^* satisfy[2,10] the following scaling law:

$$\epsilon^* = F^*(\alpha^*,\beta^*) = \{\alpha^*\}^2F^*(1,\lambda), \quad \lambda = \beta^*\{\alpha^*\}^{-3}. \qquad (3.11)$$

Now the relativistic quantity which, according to (3.5), goes over to ϵ^* as $c \to \infty$ is given, in the present example, by

$$2\{\epsilon - 1\}b^2 = \{\alpha^*\}^2[\tfrac{1}{2}\{F(\alpha^*/2b, \beta^*/2b^3) - 1\}$$
$$\times (\alpha^*/2b)^{-2}]. \qquad (3.12)$$

It follows from (3.11) and (3.12) that the limit scaling law for the relativistic energy function F is given by (3.12) together with the following limit relation:

$$F^*(1,\lambda) = \lim_{\alpha \to 0}[\tfrac{1}{2}\{F(\alpha, 4\lambda\alpha^3) - 1\}/\alpha^2]. \qquad (3.13)$$

The limit scaling laws which we have found are the closest we could come to the usual scaling laws which are obtained in the nonrelativistic domain. The approximate eigenvalues which we shall find for the relativistic and nonrelativistic problems are also related by these same limit scaling laws.

IV. THE POTENTIAL ENVELOPE METHOD

The following is a self-contained presentation of a particular application of the potential envelope method. We do not, however, make the claim that our approximate energies are lower bounds to the exact energies which would certainly be the case for the corresponding Schrödinger problem. A more complete description of this method, which is not limited only to families of Coulomb potentials, may be found in Refs. 1 and 2. We assume now that the potential shape $f(x)$ is given as a smooth, monotone-increasing, and convex transformation g of the hydrogenic shape $-1/x$. Thus we have

$$f(x) = g(-1/x). \tag{4.1}$$

We get an entirely analogous theory if we assume that g is concave: for the corresponding Schrödinger problem in this case, instead of lower bounds, the resulting eigenvalue formulas yield upper bounds. For a full discussion of this and some nonrelativistic examples, we refer the reader to our earlier articles.[1,2,10]

From now on we shall assume that g is convex on $(-\infty, 0)$. With the aid of calculus we can easily derive the following potential inequality from (4.1):

$$f(x) \geqslant f^{(t)}(x) = P(t)\{-1/x\} + Q(t),$$
$$\text{for all } x \in (0,\infty) \text{ and each fixed } t \in (0,\infty), \tag{4.2}$$

where

$$P(t) = t^2 f'(t) > 0, \quad Q(t) = f(t) + tf'(t). \tag{4.3}$$

We say that $f = \underset{t}{\text{Envelope}}\,\{f^{(t)}\}$ and we call this an "envelope representation" of the potential f. In the present application we may speak of a "Coulomb envelope representation" since, up to an additive constant, $f^{(t)}$ is a Coulomb potential.

If, for suitable values of v, b, and j, solutions exist to the coupled Dirac radial equations (2.3), then those values of $\epsilon_{nj} = F_{nj}(v,b)$ for which $\int_0^\infty \{\psi_1^2(z) + \psi_2^2(z)\}dz < \infty$ are called Dirac eigenvalues. We recall that we have chosen to enumerate all such eigenvalues, for a given j and m_j, and both values ± 1 of τ, by means of our radial quantum number n; there is no parity label. We now suppose that the potential shape f is such that the Coulomb coefficient $P(t)$ in its envelope representation satisfies $vbP(t) < 1$ (this is the "$Z < 137$" restriction; it is not required for *all* t, but only for those values of t that we eventually use). Under these conditions we see from (2.3) and (2.8) that the envelope component $f^{(t)}$ leads to the energy formulas

$$f^{(t)} \to \epsilon_{nj}^{(t)} = F_{nj}^{(t)}(v,b) = D_{nj}(vP(t)b) + vQ(t). \tag{4.4}$$

Our energy approximation ϵ_{nj}^A is then obtained by finding the envelope with respect to t of the family of functions $\epsilon_{nj}^{(t)}$ in (4.4). Thus we have

$$\epsilon_{nj}^A = F_{nj}^A(v,b) = \min_t \epsilon_{nj}^{(t)}. \tag{4.5}$$

If we now set

$$u = vbP(t) = vbt^2 f'(t) = vbg'(-1/t), \tag{4.6}$$

then we see that u is a monotone increasing function of t because g is convex ($g'' > 0$). We now find the critical point of $\epsilon_{nj}^{(t)}$ by differentiating (4.3) and canceling g'' to give $btD'_{nj}(u) = -1$. The recipe for F_{nj}^A may therefore be rewritten in terms of the variable u and we have

$$F_{nj}^A(v,b) = \min_{u \in (0,1)} [D_{nj}(u) - uD'_{nj}(u) + vf\{-1/bD'_{nj}(u)\}]. \tag{4.7}$$

Thus the approximate energy functions F_{nj}^A are expressed as a Legendre transformation involving the potential shape f and the energy functions D_{nj} of the hydrogenic atom. Since we know from Eq. (2.13) that $D_{nj}(u)$ is concave (i.e., $D''_{nj}(u) < 0$), we can differentiate the energy expression in (4.7), cancel $D''_{nj}(u)$, and again obtain, now in terms of u, the necessary condition for an extreme value

$$u/vb = \{bD'_{nj}(u)\}^{-2} f'\{-1/bD'_{nj}(u)\} = g'\{bD'_{nj}(u)\}. \tag{4.8}$$

Equation (4.8) is, of course, just (4.6) with t set to the critical value $t = \{-bD'_{nj}(u)\}^{-1}$. Since g is convex by hypothesis, and we know that D_{nj} is concave for $u \in (0,k)$, we deduce that the right-hand side of (4.8) is a decreasing function of u. Also $D'_{nj}(0) = 0$. Therefore sufficient conditions for the existence of a solution to (4.8) for $u \in (0,1)$ are

$$g'(0-) > 0 \quad \text{and} \quad v < [bg'\{bD'_{nj}(1)\}]^{-1}. \tag{4.9}$$

In this sense, the coupling constant v should not be too large for, if it is, the Coulomb envelope representation is pushed into the "$Z > 137$" region. For $j = \frac{1}{2}$, $D'_{nj}(1) = \infty$ and there is usually a solution in this case for all v. More details will be presented when we consider some examples in the next section. Values of $D'_{nj}(1)$ for $j > \frac{1}{2}$ are shown in Table I.

Further simplification of the general situation is possible if we define the "Coulomb kinetic potentials"[2] $\mathbf{h}_{nj}(s)$ by another Legendre transformation:

$$s = D_{nj}(u) - uD'_{nj}(u) \geqslant 1,$$
$$\mathbf{h}_{nj}(s) = D'_{nj}(u), \quad \mathbf{h}'_{nj}(s) = -1/u. \tag{4.10}$$

This transformation is possible because we know from (2.13) that $D_{nj}(u)$ is concave for $u \in (0,k)$. In terms of these new variables we have

$$D_{nj}(u) = \min_{s>1} [s + u\mathbf{h}_{nj}(s)]. \tag{4.11}$$

Thus, in this formulation, the energy trajectories $D_{nj}(u)$ are obtained from the potential $h(x) = -1/x$ in two stages:

$$h(x) \to \mathbf{h}_{nj}(s) \to D_{nj}(u). \tag{4.12}$$

In these terms our general problem is represented by the chain

$$f(x) = g\{h(x)\} \to \mathbf{f}_{nj}^{(b)}(s) \simeq g\{b\,\mathbf{h}_{nj}(s)\} \to F_{nj}^A(v,b). \tag{4.13}$$

Energy trajectories are therefore generated by Legendre transformations (4.11) of the corresponding kinetic potentials: if the potential is now transformed by g, then the new approximate kinetic potentials are obtained by applying the same transformation g to the old kinetic potentials. Equation (4.7) now becomes

TABLE I. Values of $D'_{nj}(1)$. Table of values of $|D'_{nj}(1)|$ which are used for indicating the allowed ranges of the potential parameters. The values for even $n > 1$ are the same as those corresponding to $n + 1$. For $j = \frac{1}{2}$ this quantity is unbounded.

n	$j=\frac{3}{2}$	$j=\frac{5}{2}$	$j=\frac{7}{2}$	$j=\frac{9}{2}$	$j=\frac{11}{2}$
1	0.288 68	0.117 85	0.064 55	0.040 82	0.028 17
3	0.134 39	0.067 50	0.041 68	0.028 49	0.020 76
5	0.074 71	0.043 22	0.029 00	0.020 97	0.015 92
7	0.046 93	0.029 89	0.021 29	0.016 05	0.012 58
9	0.032 03	0.021 85	0.016 27	0.012 68	0.010 19

$$F^A_{nj}(v,b) = \min_{s>1} [s + vg\{b\,\mathbf{h}_{nj}(s)\}]$$

$$= \min_{s>1} [s + vf\{-1/b\,\mathbf{h}_{nj}(s)\}]. \quad (4.14)$$

From the nonrelativistic version of this theory[1,2] we know that, in the nonrelativistic limit, the positive parameter s here represents the rest energy (= 1) plus the mean kinetic energy of the particle. Some aspect of this physical interpretation must remain in the relativistic case, too: we notice that all parameters introduced by the potential appear only in the term after s. The kinetic potentials $\mathbf{h}_{nj}(s)$ for the relativistic hydrogenic atom appear to be algebraically rather complicated except at the bottom of each angular-momentum subspace ($n = 1$, that is to say, $\nu = k = j + \frac{1}{2}$), where we find from (4.10) that they have the delightfully simple explicit form

$$\mathbf{h}_{1j}(s) = -\{s^2 - 1\}^{1/2}/(j + \tfrac{1}{2}). \quad (4.15)$$

While the implicit equation (4.7) is adequate for practical purposes, the formulation (4.14) in terms of kinetic potentials is both more elegant and more suitable for the general study of these geometrical approximation methods.

V. COMBINATIONS OF POWERS AND THE LOG POTENTIAL

We suppose that the potential shape has the form

$$f(x) = \sum_q [A^q \operatorname{sgn}(q) x^q] + \lambda \log(x),$$
$$A^{(q)} \geqslant 0, \quad \lambda \geqslant 0, \quad q > -1, \quad q \neq 0, \quad (5.1)$$

where, as before, $x = r/a = z/b$, and we set $v = 1$. The potential $vf(z/b)$ which is used in the coupled radial equations (2.3) is then explicitly given by

$$vf\left(\frac{z}{b}\right) = \sum_q [\gamma^{(q)} \operatorname{sgn}(q) z^q] + \lambda \log\left(\frac{z}{b}\right),$$
$$\gamma^{(q)} = A^{(q)} b^{-q}. \quad (5.2)$$

We see that the transformation function $g(X) = f(-1/X)$ is monotone increasing and convex on $(-\infty, 0)$. The arguments of Sec. II immediately show that the trajectory functions for these potentials have the general form

$$\epsilon_{nj} = F_{nj}(\{\gamma^{(q)}\}, \lambda) - \lambda \log(b). \quad (5.3)$$

If now, for conciseness, we omit the subscripts $\{nj\}$ and the argument u of $D(u)$ then the general energy approximation equations (4.7) and (4.8) take the form

$$\epsilon^A = D - uD' + \sum_q \operatorname{sgn}(q) \gamma^{(q)} |D'|^{-q} - \lambda \log\{b\,|D'|\}, \quad (5.4)$$

$$1 = \sum_q |q| \gamma^{(q)} \{u|D'|^{(1+q)}\}^{-1} + \lambda \{u|D'|\}^{-1}. \quad (5.5)$$

We must be able to solve (5.4) and (5.5) for $u \in (\gamma^{(-1)}, 1)$. The general condition (4.9) for the existence of such a critical point becomes

$$1 > \sum_q |q| \gamma^{(q)} |D'_{nj}(1)|^{-(1+q)} + \lambda\, |D'_{nj}(1)|^{-1}. \quad (5.6)$$

It is clearly necessary that $\gamma^{(-1)} < 1$ and we therefore assume that this is so from now on. For $j = \frac{1}{2}$ we have $|D'_{n(1/2)}(1)| = \infty$ so that (5.6) can always be satisfied in these cases. In all other cases (5.6) implies upper limits on the allowed magnitudes of the component coefficients $\{\gamma^{(q)}\}$ and λ. One way to check the condition is to use Table I which gives values of $|D'_{nj}(1)|$ for $j > \frac{1}{2}$.

VI. THE COULOMB-PLUS-LINEAR POTENTIAL: A BETTER APPROXIMATION

The energy approximation for the Coulomb-plus-linear potential is included in Eqs. (5.4) and (5.5) for general linear combinations of power laws and the log potential. In this section we apply an approximation method which we have developed[2,10] for the corresponding nonrelativistic problem. In order to explain how this works, we first have to get the parametric equations for ϵ^A in a form which is homogeneous in the coupling constants α and β. We shall then look at the nonrelativistic limit and proceed by analogy.

Since the potential is a combination of two powers we can set $v = b = 1$ and without loss of generality write the potential function appearing in the radial equations (2.3) in the form

$$f(z) = -\alpha/z + \beta z, \quad \alpha \geqslant 0, \quad \beta \geqslant 0. \quad (6.1)$$

Comparison with other work is most likely to focus on this potential and we therefore remind the reader, with the aid of (2.2), that (6.1) corresponds precisely to the explicit central potential

$$V(r) = -\alpha\{\hbar c\} r^{-1} + \beta\{m^2 c^3/\hbar\} r. \quad (6.2)$$

Equations (5.4) and (5.5) now become

$$\epsilon^A = D - uD' - \alpha|D'| + \beta\,|D'|^{-1}, \quad (6.3)$$

$$1 = \alpha/u + \beta\{u|D'|^2\}^{-1}. \quad (6.4)$$

By multiplying the term D in (6.3) by the representation of 1 in (6.4) we obtain

$$\epsilon^A = \alpha D/u + \beta\{D + 2u|D'|\}\{u|D'|^2\}^{-1}. \quad (6.5)$$

In this form the parametric equations clearly indicate how ϵ^A varies with α and β. If $\beta = 0$, then $u = \alpha$ and we have $\epsilon = D(\alpha)$ which result is, of course, exact. If $\alpha = 0$, we obtain an approximation for the pure linear potential.

Now we look at the nonrelativistic limit in which $D(u) = 1 - u^2/2\nu^2$ and therefore

$$\epsilon^A - 1 = -\alpha t/2\nu + 3\beta\nu/2t; \quad (6.6)$$

$$1 = \alpha/t + \beta\nu t^3, \quad t = u/\nu. \quad (6.7)$$

If we now follow Sec. II and use the correspondences $\alpha^* = 2\alpha$, $\beta^* = 2\beta$, and $\epsilon^* = 2(\epsilon - 1)$, we obtain from Eqs. (6.6) and (6.7) exactly the nonrelativistic *lower bound* given by Eqs. (3.4) and (3.5) of Ref. 10. However, we also found in that work that a much better approximation (but no longer a bound) is obtained if the factors ν in the linear terms (associated with β) are transformed by

$$\nu = (n_r + l) \rightarrow \mu = (An_r + l - C), \quad (6.8)$$

where $(A,C) = (1.794, 0.418)$. This approximation was constructed essentially by regarding the linear potential as a "mean" between the Coulomb and the harmonic-oscillator potential; for the latter potential $(A,C) = (2,\frac{1}{2})$. The idea of a

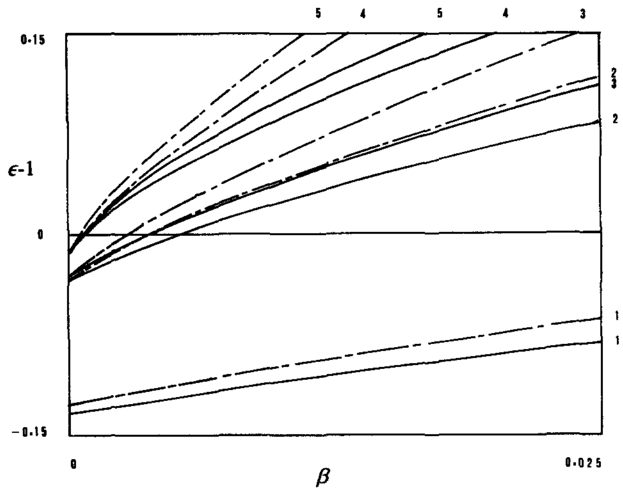


FIG. 1. Approximate energy trajectories $\epsilon_{nj}(\alpha,\beta)$ with $\alpha=\frac{1}{2}$ and $j=\frac{1}{2}$ for the Coulomb-plus-linear potential $f(r)=-\alpha/r+\beta r$. The values of n are indicated at the ends of the graphs which are shown as — - — for the nonrelativistic problem and in full line for the corresponding relativistic problem.

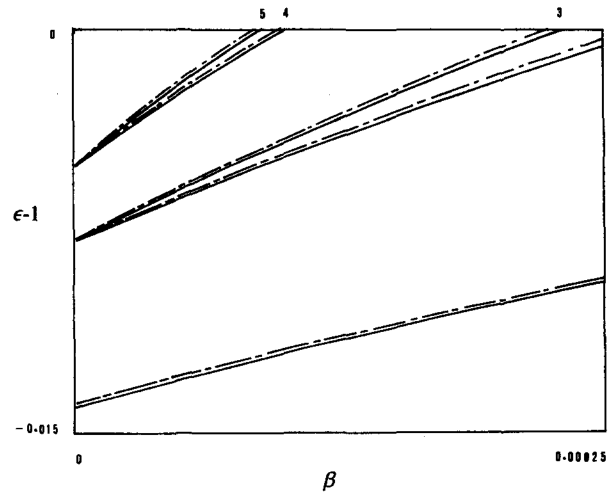


FIG. 3. Approximate energy trajectories $\epsilon_{nj}(\alpha,\beta)$ with $\alpha=\frac{1}{2}$ and $j=\frac{5}{2}$ for the Coulomb-plus-linear potential $f(r)=-\alpha/r+\beta r$. The values of n are indicated at the ends of the graphs which are shown as — - — for the nonrelativistic problem and in full line for the corresponding relativistic problem.

"mean" is more transparent in terms of kinetic potentials which we have not yet fully developed for relativistic problems. However, at present we simply translate (6.8) in terms of the relativistic quantum numbers $\{n,j\}$, and then apply the result in (6.4) and (6.5). From Eq. (2.12) we infer that the expression for μ should be

$$\mu=A\{n-I(n/2)\}+\{j+\tfrac{1}{2}+2I(n/2)-n\}-C,$$
$$(A,C)=(1.794,\,0.418). \tag{6.9}$$

Finally, we see that the recipe for the improved approximate energy ϵ^A is provided by the rule $\beta\rightarrow\beta\mu/\nu$. The simplest way to express this is to return to the use of u as a parameter and from (6.4) and (6.5) we obtain

$$\epsilon^A_{nj}=D_{nj}(u)+2(u-\alpha)|D'_{nj}(u)|, \tag{6.10}$$

$$\beta=(u-\alpha)\nu\mu^{-1}|D'_{nj}|^2,\quad u\in(\alpha,1). \tag{6.11}$$

As we have often mentioned, β cannot be too large. In fact we require

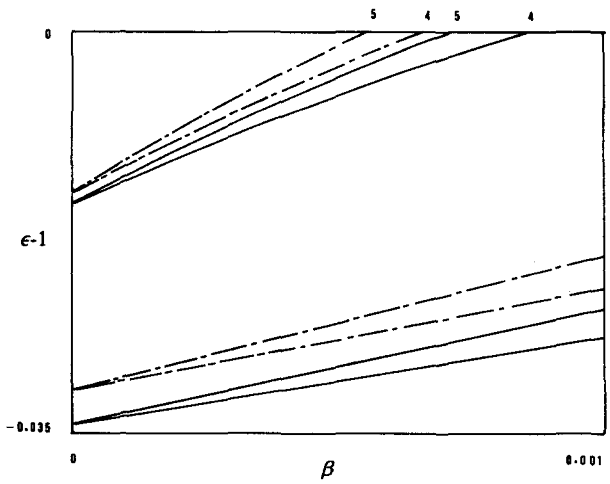


FIG. 2. This is a magnified picture of the energy trajectories $\epsilon_{n(1/2)}(\frac{1}{2},\beta)$ for $n=2$–5 shown more fully in Fig. 1. In this diagram one can clearly see the splitting of the Coulomb degeneracy as the linear potential measured by β is turned on. Nonrelativistic: — - —; relativistic ——.

$$\beta<(1-\alpha)\nu\mu^{-1}|D'_{nj}(1)|^2. \tag{6.12}$$

In Figs. 1–3 we exhibit some graphs of $\epsilon^A_{nj}(\beta)$ for $\alpha=\frac{1}{2}, j=\frac{1}{2}$ and $\frac{5}{2}$, and $n=1$–5. We have chosen $\alpha=\frac{1}{2}$ for these illustrations because the system is then already very relativistic even with $\beta=0$ for it corresponds to "$Z=137/2$" for a hydrogenlike atom. Figure 2 is a magnified picture of the start of the $n=2$–5 curves shown more fully in Fig. 1. As the linear field measured by β is switched on, one can easily see the splitting of the Coulomb degeneracy. In Fig. 3 the case $j=\frac{5}{2}$ is illustrated. In all three diagrams the nonrelativistic results for which $D_{nj}(u)=1-u^2/2\nu^2$ are shown as broken lines whereas the corresponding relativistic results are shown as continuous curves. Parametric equations, of course, are very convenient for plotting graphs.

VII. CONCLUSION

We have looked at the problem of a single fermion moving in a central field and obeying Dirac's equation. The problem is a generalization of the hydrogenic problem in which the potential is a convex transformation of the Coulomb potential: the Lorentz-scalar term, the rest mass is constant. This problem was already formulated in 1928 and today numerical solutions can always be found with the aid of a computer. In this article we have explored the use of geometrical technique to try to solve the problem approximately in analytical terms. The most specific results are the parametric equations (6.10) and (6.11) for the Coulomb-plus-linear potential. With the aid of these equations one can easily obtain a first approximation to the energy spectrum. The Klein-Gordon equation can, of course, be treated in exactly the same fashion.

The geometrical approach has become very important recently particularly because of the growing interest in nonlinear phenomena such as solitary waves. The general aim here is to exploit a known exact solution by looking at transformations which leave the essential form of the solution invariant. Envelope representations allow one to study smooth transformations of a given soluble problem and this leads to a kind of global approximation theory.

ACKNOWLEDGMENTS

The author would like to thank the Department of Physics of The National Taiwan University in Taipei for its hospitality during a visit when this work was carried out.

He would also like to thank the Natural Sciences and Engineering Research Council Canada for partial financial support of this work by Grant No. A3438.

[1]R. L. Hall, Phys. Rev. D **22**, 2062 (1980).

[2]R. L. Hall, J. Math. Phys. **24**, 324 (1983); Phys. Rev. D **27**, 2379 (1983); J. Math. Phys. **25**, 2708 (1984).

[3]M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV: Analysis of Operators* (Academic, New York, 1978).

[4]W. Thirring, *A Course in Mathematical Physics 3: Quantum Mechanics of Atoms and Molecules* (Springer, New York, 1981).

[5]B. Swirles, Proc. R. Soc. London Ser. A **152**, 625 (1935); D. R. Hartree, Repts. Prog. Phys. **XI**, 113 (1946).

[6]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1980), p.307.

[7]A. S. Wightman, *The Dirac Equation*, in the collection entitled *Aspects of Quantum Theory*, edited by A. Salam and E. P. Wigner (Cambridge U. P., Cambridge, 1972), p.95.

[8]A. Messiah, *Quantum Mechanics, Volume II* (North-Holland, Amsterdam, 1961). The Dirac Hamiltonian with central potentials is discussed on p.924.

[9]R. L. Hall and C. Kalman, Phys. Lett. B **83**, 80 (1979).

[10]R. L. Hall, Phys. Rev. D **30**, 433 (1984). In this reference the radial quantum number n_r is written as "n," and the coefficients A and C in the quantity μ are written "$2A$" and "$2C$." In some expressions involving $\mu = (2An + l - 2C)$, the quantum number n has accidentally been left out.

Explicit forms of Bäcklund transformations for the Grassmannian and CP^{N-1} sigma models

R. Sasaki
Research Institute for Theoretical Physics, Hiroshima University, Takehara, Hiroshima 725, Japan

(Received 17 July 1984; accepted for publication 22 February 1985)

A fairly wide class of classical solutions of the Euclidean two-dimensional Grassmannian and CP^{N-1} sigma models has been constructed explicitly and elementarily by the present author. Starting from these classical solutions we derive the explicit forms of the Bäcklund transformations following Harnad *et al.* Properties of the Bäcklund transformations for the Grassmannian models are discussed. In particular, a simple interpretation of the Bäcklund transformations for the CP^{N-1} model is obtained. Generalization of these results to the noncompact Grassmannian sigma models is straightforward.

I. INTRODUCTION

Exactly integrable field theoretical models in two dimensions such as the sine–Gordon theory, the massive Thirring models, and the O(n) sigma model[1] have provided an interesting theater for discussing the complicated nonlinear interactions in particle physics. Among them a simple and important class of models is given by the SL(N,C) principal chiral model[2] and its reductions to the various subgroups or to the Riemannian symmetric spaces.[3] These are two-dimensional free scalar field theories with their field variables constrained in a Lie group or in a symmetric space. A salient feature of these models is the existence of the Bäcklund transformations[4] (sometimes abbreviated as BT's). Namely, if a solution is known a new solution of the same nonlinear equation can be generated by a transformation which involves only the linear algebraic procedure.[5] The Bäcklund transformation method, however, has had so far only limited success because of the lack of general enough starting solutions and/or the solutions of the corresponding linear scattering problems.

The situation has changed recently for the Grassmannian sigma models. The Grassmannian sigma models are obtained from the SL(N,C) principal chiral model by reduction to the complex Grassmann manifolds. Because of the many common features with the four-dimensional gauge theories, e.g., the built-in non-Abelian gauge structure, the model has attracted physicists' attention. Recently a fairly wide class of classical solutions of the two-dimensional Euclidean Grassmannian sigma models was constructed explicitly and elementarily by the present author.[6] Moreover, the linear scattering problems for these solutions can also be solved explicitly. So we are now in a position to be able to construct the explicit forms of the Bäcklund transformations. In this paper we construct and discuss the explicit forms of the simplest type of the Bäcklund transformations for the Grassmannian and the CP^{N-1} sigma models.[7] A particularly simple interpretation of the BT for the CP^{N-1} model is obtained. This we believe will set an interesting first step towards understanding the relationship between the BT method and our method of solution, which so far appear to be uncorrelated.

This paper is organized as follows. In Sec. II we recapitulate the derivation of the Grassmannian and the CP^{N-1} sigma models through the reduction of the principal chiral model. This is mainly for the purpose of introducing appropriate notation and for self-containedness. In Sec. III we summarize some of the recent results about explicit solutions of the Grassmannian sigma models[6] together with the associated linear scattering problem. In Sec. IV the simplest type of Bäcklund transformations is constructed explicitly starting from the solutions shown in Sec. III. In Sec. V, the BT for the CP^{N-1} model is discussed in some detail and a simple interpretation of the Bäcklund transformation is obtained. In Sec. VI we discuss the generalization of the above results to the noncompact Grassmannian models. The Appendix is devoted to a short summary of the formal Bäcklund transformation theory formulated by Harnad *et al.*[5]

II. PRINCIPAL CHIRAL MODELS AND SIGMA MODELS

In this section we recapitulate the derivation of the Grassmannian and CP^{N-1} models through the reduction of the principal chiral model. The principal chiral models are the simplest examples of completely integrable relativistic field theories. They are two-dimensional free scalar fields on Lie groups. A typical example is the SL(N,C) principal chiral model defined by the Lagrangian

$$L = \mathrm{Tr}[(\partial_\mu g)(\partial_\mu g^{-1})], \quad \mu = 1,2, \tag{2.1}$$

in which $g = g(x)$ is an element of SL(N,C) and $x = (x_1,x_2)\in R^2$ are the coordinates of the two-dimensional Euclidean space. The equation of motion obtained by the Euler derivative of the above Lagrangian reads

$$\partial_\mu \partial_\mu g - (\partial_\mu g)g^{-1}(\partial_\mu g) = 0. \tag{2.2}$$

Due to the geometric construction of the model, it exhibits a high degree of symmetry. For example, it has an infinite number of conservation laws[8] and above all the equation of motion (2.2) can be interpreted as the integrability condition for the following set of linear equations:

$$\frac{\partial}{\partial x_+}\psi(x;\lambda) = \frac{A}{1+\lambda}\psi(x;\lambda),$$
$$\frac{\partial}{\partial x_-}\psi(x;\lambda) = \frac{B}{1-\lambda}\psi(x;\lambda). \tag{2.3}$$

 0022-2488/85/071786-07$02.50

Here ψ is an $N\times N$ matrix function of x and λ,

$$A=\left(\frac{\partial}{\partial x_+}g\right)g^{-1},\quad B=\left(\frac{\partial}{\partial x_-}g\right)g^{-1},\quad x_\pm=x_1\pm ix_2, \tag{2.4}$$

and λ is an arbitrary complex (spectral) parameter.

The SL(N,C) principal chiral model is interesting not only for its own sake but also for the large number of integrable subsystems contained in it. The process of finding suitable submanifolds of SL(N,C) on which the restriction of the equation of motion (2.2) is again integrable is called the reduction. It is known that when the submanifold is another Lie group,[2] e.g., SU(n), SO(n), or a Riemannian symmetric space[3] G/H, the reduced system is integrable. In the present paper we focus our attention to the reduction of Eq. (2.2) to the complex Grassmannian manifold and to the complex projective space as a special case of the former. The resulting field theoretical model is called the complex Grassmannian sigma model and the CP^{N-1} model, respectively. The noncompact version of these models will also be discussed briefly. As will be discussed in the next section a quite general class of explicit solutions for these models can be constructed elementarily.

The complex Grassmannian manifold, to be denoted as $G(N,m)$ hereafter, is a typical example of a Riemannian symmetric space. It is defined as a quotient space G/H

$$G(N,m)=\mathrm{SU}(m+n)/\mathrm{S}(\mathrm{U}(m)\times\mathrm{U}(n)),\quad N=m+n. \tag{2.5}$$

Namely the isometry group G is SU($m+n$) and the isotropy subgroup is S(U(m)$\times$U(n)),

$$G=\mathrm{SU}(m+n),\quad H=\mathrm{S}(\mathrm{U}(m)\times\mathrm{U}(n)).$$

The Cartan immersion of G/H in G is defined by an involutive automorphism σ (see Ref. 9),

$$i;\quad G/H\rightarrow G,\quad i(gH)\mapsto\sigma(g)g^{-1}. \tag{2.6}$$

In the present case the automorphism, to be denoted as σ_1, is given by

$$\sigma_1(g)=I_{n,m}gI_{n,m},\quad I_{n,m}=\begin{pmatrix}I_n & 0\\ 0 & -I_m\end{pmatrix}, \tag{2.7}$$

in which $I_n(I_m)$ is the $n\times n$ ($m\times m$) unit matrix. We need another constraint determining the unitary group SU(N) ($N=m+n$) as a subgroup of SL(N,C),

$$\mathrm{SU}(N)=\{g\in\mathrm{SL}(N,C)|\sigma_2(g)=g\},\quad\sigma_2(g)=(g^\dagger)^{-1}. \tag{2.8}$$

The images of the immersion is a totally geodesic submanifold $\Sigma\subset\mathrm{SL}(N,C)$,

$$\Sigma=\mathrm{Im}\,i=\{g\in\mathrm{SL}(N,C)|\sigma_1(g)=g^{-1},\ \sigma_2(g)=g\}. \tag{2.9}$$

The complex projective space is a special case of the Grassmann manifold $G(N,m)$ for $m=1$,

$$CP^{N-1}=G(N,1)=\mathrm{SU}(N)/[\mathrm{SU}(N-1)\times\mathrm{U}(1)]. \tag{2.10}$$

From now on we fix the integer N and m and consider the principal chiral model restricted on Σ. Let us introduce a simplified notation for $I_{n,m}$ in Eq. (2.7),

$$t\equiv I_{n,m}. \tag{2.11}$$

For an arbitrary element $g\in\Sigma$, tg is Hermite and its square is the unit matrix

$$(tg)^\dagger=tg,\quad(tg)^2=1. \tag{2.12}$$

If we define an $N\times N$ matrix P by

$$P=\tfrac{1}{2}(1-tg)\quad\text{or}\quad g=t(1-2P), \tag{2.13}$$

we find that P is a projector

$$P=P^\dagger=P^2 \tag{2.14}$$

and

$$\mathrm{Tr}\,P=m. \tag{2.15}$$

The last equation follows from the continuity of the trace. It is easy to see that if $g\in\Sigma$ is a solution of Eq. (2.2), then P satisfies the following equation[10]:

$$[\partial_\mu\,\partial_\mu P,P]=0, \tag{2.16}$$

and vice versa. This is the equation of motion of the $G(N,m)$ model. If Eq. (2.13) is substituted to the linear scattering equation (2.3) of the principal chiral model, we obtain

$$\partial_+\psi=\frac{2t\,[\partial_+P,P\,]t}{1+\lambda}\,\psi,\quad\partial_-\psi=\frac{2t\,[\partial_-P,P\,]t}{1-\lambda}\,\psi, \tag{2.17}$$

$$\partial_+=\frac{\partial}{\partial x_+},\quad\partial_-=\frac{\partial}{\partial x_-}.$$

The compatibility (integrability) condition of the above equations is the $G(N,m)$ equation (2.16), which shows that the reduction of the principal SL(N,C) chiral model to the complex Grassmannian sigma model retains the complete integrability. For another formulation of the $G(N,m)$ model, please refer to Refs. 3 and 6.

III. SOLUTIONS FOR THE COMPLEX GRASSMANNIAN SIGMA MODELS

In this section we summarize some of the recent results about explicit solutions for the complex Grassmannian sigma models and the CP^{N-1} model.[6] As is explained in the previous section the $G(N,m)$ model is essentially a massless (complex scalar) free field theory with a geometrical constraint, namely a harmonic map to the Grassmann manifold. Therefore it is quite natural to expect that the solutions should somehow be related to holomorphic or antiholomorphic functions of $x_+(=x_1+ix_2)$, since they solve the massless free field equation in two Euclidean dimensions. In fact the simplest solutions of the CP^1 model[11] and the CP^{N-1} model,[12] the instanton solutions, are constructed explicitly in terms of holomorphic functions. Our solution method can be regarded as a natural generalization of them.

The $G(N,m)$ model is described by an $N\times N$ projection matrix P, $\mathrm{Tr}\,P=m$, Eq. (2.14), with the equation of motion

$$[\partial_+\,\partial_-P,\,P\,]=0. \tag{3.1}$$

The CP^{N-1} model is included as a special case of $m=1$. A fairly general class of solutions (generic solutions) are constructed as follows. Let us introduce m linearly and functionally independent holomorphic N-component vectors

$$f_1(x_+),f_2(x_+),\dots,f_m(x_+),\quad\partial_-f_i=0, \tag{3.2}$$

and define another set of N-component vectors $f_{m+1},\dots,f_N$ by differentiating (3.2) with respect to x_+:

$$f_{m+1} = \partial_+ f_1, f_{m+2} = \partial_+ f_2, \ldots, f_{2m} = \partial_+ f_m,$$
$$f_{2m+1} = \partial_+^2 f_1, \ldots, f_{3m} = \partial_+^2 f_m, \ldots, f_N. \tag{3.3}$$

For the time being we assume that the vectors $f_1,\ldots,f_N$ are linearly independent spanning the N-dimensional complex space C^N. Then we orthonormalize the vectors $F_1,\ldots,f_N$ by the Gram–Schmidt procedure in this order and obtain an orthonormal basis of C^N,

$$e_1, e_2, \ldots, e_N, \quad e_i^\dagger \cdot e_k = \delta_{ik}. \tag{3.4}$$

By picking up *m consecutive orthonormal vectors*, we define the following $N \times m$ matrices:

$$\begin{aligned} X_{(1)} &= (e_1, e_2, \ldots, e_m), \\ X_{(2)} &= (e_2, e_3, \ldots, e_{m+1}), \\ &\vdots \\ X_{(N-m+1)} &= (e_{N-m+1}, \ldots, e_N). \end{aligned} \tag{3.5}$$

Each of them gives a solution of the $G(N,m)$ model through

$$\begin{aligned} &P = P_{(j)} = X_{(j)} X_{(j)}^\dagger, \quad j = 1,2,\ldots, N-m+1, \\ &[\partial_+ \partial_- P_{(j)}, P_{(j)}] = 0, \end{aligned} \tag{3.6}$$

which is characterized by the arbitrary input holomorphic vectors $f_1,\ldots,f_m$. In particular, the first one $P_{(1)}$ is an instanton[13] and the last is an anti-instanton. The proof is quite elementary so we show only its outline. For more details and for other types of solutions we refer to Refs. 6 and 14. Let us show that $P_{(j)}$ is a solution. First we introduce an auxiliary matrix variable $Q_{(j)}$ by

$$Q_{(j)} = \sum_{k=1}^{j-1} e_k e_k^\dagger, \tag{3.7}$$

which is a projector orthogonal to $P_{(j)}$,

$$Q_{(j)}^\dagger = Q_{(j)} = Q_{(j)}^2, \quad Q_{(j)} P_{(j)} = P_{(j)} Q_{(j)} = 0. \tag{3.8}$$

Hereafter the suffix j of $P_{(j)}$ and $Q_{(j)}$ is fixed and omitted for simplicity. Because of the way of constructing orthonormal vectors (3.4) from the holomorphic vectors (3.2) and (3.3), the projectors P and Q satisfy the following equations:

$$\begin{aligned} &(\partial_- Q) Q = Q \partial_+ Q = 0, \\ &P \partial_- Q = (\partial_+ Q) P = Q \partial_+ P = (\partial_- P) Q = 0, \end{aligned} \tag{3.9}$$

and

$$\begin{aligned} &P \partial_+ Q = -(\partial_+ P) Q = \partial_+ Q, \\ &(\partial_- Q) P = -Q(\partial_- P) = \partial_- Q. \end{aligned} \tag{3.10}$$

By combining these we obtain

$$(\partial_- P) P + \partial_- Q = 0, \tag{3.11a}$$

$$P(\partial_+ P) + \partial_+ Q = 0. \tag{3.11b}$$

From the combination $\partial / \partial x_+ [(3.11a)] - \partial / \partial x_- [(3.11b)]$, we get the desired relationship

$$[\partial_+ \partial_- P, P] = 0.$$

For the above solution P of the $G(N,m)$ model, the corresponding linear scattering equation (2.17) can be solved easily. First let us note that the commutators $[\partial_\pm P, P]$ in Eq. (2.17) are linearized by using Eqs. (3.11a) and (3.11b),

$$\begin{aligned} &[\partial_+ P, P] = \partial_+ (P + 2Q), \\ &[\partial_- P, P] = -\partial_- (P + 2Q). \end{aligned} \tag{3.12}$$

Therefore we look for a particular solution of Eq. (2.17) in the form

$$\psi = tU, \quad U(x;\lambda) = 1 + \alpha P + \beta Q, \tag{3.13}$$

with unknown complex constants α and β. An elementary calculation using the projection properties of P, Q and Eqs. (3.9)–(3.11) shows that[15]

$$\begin{aligned} &U(x;\lambda) = 1 + [2/(\lambda - 1)] P + [4\lambda / (\lambda - 1)^2] Q, \\ &U(x;\lambda)^{-1} = 1 - [2/(\lambda + 1)] P - [4\lambda / (\lambda + 1)^2] Q \end{aligned} \tag{3.14}$$

solves the equation. A general solution of Eq. (2.17) is obtained from (3.14) by multiplying an arbitrary x-independent (but λ-dependent) matrix from the right. We impose

$$\psi(x;0) = g(x), \quad \psi(x;\infty) = 1, \tag{3.15}$$

as boundary conditions for $\psi(x,\lambda)$, to fix the constant matrix and obtain

$$\begin{aligned} t\psi(x;\lambda) &= \left(1 + \frac{2}{\lambda - 1} P + \frac{4\lambda}{(\lambda - 1)^2} Q\right)\left(\frac{1 - \lambda t}{1 - \lambda}\right), \\ \psi(x;\lambda)^{-1} t &= \left(\frac{1 + \lambda t}{1 + \lambda}\right)\left(1 - \frac{2}{\lambda + 1} P - \frac{4\lambda}{(\lambda + 1)^2} Q\right). \end{aligned} \tag{3.16}$$

It is easy to check that this solution fulfills the invariance condition due to the reduction[5]

$$\psi(\lambda) = g t \psi(1/\lambda) t, \tag{3.17a}$$

$$\psi(\lambda) = [\psi(-\bar{\lambda})^\dagger]^{-1}. \tag{3.17b}$$

For later calculation we introduce an arbitrary "initial" point x_0 and the normalized solution ψ_0 as

$$\begin{aligned} \psi_0(x, x_0; \lambda) &\equiv \psi(x;\lambda) \psi(x_0;\lambda)^{-1}, \\ &= tU(x;\lambda) U(x_0;\lambda)^{-1} t, \end{aligned} \tag{3.18}$$

with $\psi_0(x_0, x_0; \lambda) = 1$.

IV. BÄCKLUND TRANSFORMATIONS

In this section we construct the explicit forms of the Bäcklund transformations for the complex Grassmannian sigma models starting from the known solutions of the models given in the previous section. In their beautiful paper, Harnad *et al.*[5] have shown the linearization of the multi-Bäcklund transformations. Namely, a new solution defined by application of an arbitrary number of BT's to the original one can be constructed in a purely linear algebraic way if the solutions of the linear scattering equation for the original one are known. In the Appendix we summarize some of their main results appropriate for us. However, in applying their method to calculate the explicit forms of new solutions there arises a technical difficulty, i.e., inverting a big matrix Γ_{ij}, Eq. (A8), at least $K \times K$ for a K-tuple BT, which is highly nontrivial.

Therefore in this paper we restrict ourselves to the simplest case, namely the minimal ($K = 2$) Bäcklund transformation that maps a solution of the $G(N,m)$ model to another of the same model. We introduce another simplifying assumption that the residues of the dressing matrix are of rank 1 [see Eq. (4.3)]. First we construct the dressing matrix of Zakharov and Makhailov[2] after Harnad *et al.* In the case $K=2$ the dressing matrix $\chi(\lambda)$ and its inverse $\chi^{-1}(\lambda)$ have two poles which we denote as

$$\lambda_1=\xi,\quad \lambda_2=\lambda_{\hat 1}=1/\xi,\quad \text{for } \chi(\lambda),$$
$$\mu_1=\lambda_{\bar 1}=-\bar\xi,\quad \mu_2=\lambda_{\hat{\bar 1}}=-1/\bar\xi,\quad \text{for } \chi^{-1}(\lambda). \tag{4.1}$$

Here $\xi\,(\neq 0)$ is an arbitrary complex number. We assume for simplicity that none of the poles in Eq. (4.1) coincide. In order to calculate the residues Q_i, R_i of the dressing matrix

$$\chi(\lambda)=1+Q_1/(\lambda-\xi)+Q_2/(\lambda-1/\xi),$$
$$\chi^{-1}(\lambda)=1+R_1/(\lambda+\bar\xi)+R_2/(\lambda+1/\bar\xi), \tag{4.2}$$

their rank must be specified. Here we consider also the simplest situation, namely,

$$\operatorname{rank}(Q_i)=\operatorname{rank}(R_i)=1,\quad i=1,2. \tag{4.3}$$

Then the matrices $\Gamma_{ij},\Gamma^\dagger_{ij}$, Eq. (A8), are 2×2 and they can be easily inverted. The residues Q_i and R_i are expressed as

$$Q_i=X_iF_i^\dagger,\quad R_i=H_iK_i^\dagger,\quad i=1,2, \tag{4.4}$$

in which F_i and H_i are $N\times 1$ matrices

$$F_i=(\psi_0(x;\lambda_i)^\dagger)^{-1}f_i,\quad H_i=\psi_0(x;\mu_i)h_i. \tag{4.5}$$

The $N\times 1$ matrices X_i and K_i are determined from the linear equation (A7). Since the "initial" values f_i and h_i obey the restriction (A12) due to the reduction, they are linearly dependent and can be expressed by any one of them, say f_1,

$$h_1=f_1a_1,\quad h_2=(g_0^{-1})^\dagger tf_1a_2,\quad f_2=(g_0^{-1})^\dagger tf_1a_3, \tag{4.6}$$

in which $g_0=g(x_0)$ and a_i is an arbitrary complex number. By using the involution properties of ψ, Eq. (A9), and the expression (3.18) for the normalized solution ψ_0, we obtain

$$F_1=tU(x;-\bar\xi)U(x_0;-\bar\xi)^{-1}tf_1,$$
$$F_2=tU(x;-1/\bar\xi)U(x_0;-\bar\xi)^{-1}tf_1a_3,$$
$$H_1=tU(x;-\bar\xi)U(x_0;-\bar\xi)^{-1}tf_1a_1,$$
$$H_2=tU(x;-1/\bar\xi)U(x_0;-\bar\xi)^{-1}tf_1a_2. \tag{4.7}$$

Therefore if we introduce a new constant vector ($N\times 1$ matrix) u by

$$u\equiv U(x_0;-\bar\xi)^{-1}tf_1, \tag{4.8}$$

the explicit x_0 dependence is wiped out.[16] Since the overall scale of u is irrelevant for later calculation we may assume without loss of generality that u is a unit vector

$$u^\dagger u=1. \tag{4.9}$$

By using the involution properties of ψ again together with another simplifying notation

$$h=h(x;\xi)\equiv U(x;-\bar\xi)u, \tag{4.10}$$

we obtain

$$F_1=th,\quad H_1=a_1th,\quad F_2=a_3gh,\quad H_2=a_2gh, \tag{4.11}$$

in which $g=g(x)=t(1-2P)$. Then the matrix Γ_{ij}, (A8) reads

$$\Gamma_{11}=a_1\beta/(\xi+\bar\xi),\quad \Gamma_{12}=a_2\gamma/(\xi+1/\bar\xi),$$
$$\Gamma_{21}=a_1\bar a_3\gamma/(1/\xi+\bar\xi),\quad \Gamma_{22}=a_2\bar a_3\beta/(1/\xi+1/\bar\xi), \tag{4.12}$$

in which β and γ are real numbers

$$\beta=\beta(x;\xi)\equiv h^\dagger h>0,\quad \gamma=\gamma(x;\xi)\equiv h^\dagger tgh. \tag{4.13}$$

By solving the linear equation (A7), we obtain the residues

$$Q_1=\frac{1}{|\xi|^2\Delta}\left(\frac{|\xi|^2\beta}{\xi+\bar\xi}thh^\dagger t-\frac{\xi\gamma}{1+|\xi|^2}ghh^\dagger t\right),$$
$$Q_2=\frac{1}{|\xi|^2\Delta}\left(\frac{\beta}{\xi+\bar\xi}ghh^\dagger g^\dagger-\frac{\bar\xi\gamma}{1+|\xi|^2}thh^\dagger g^\dagger\right),$$

and

$$R_1=-Q_1^\dagger,\quad R_2=-Q_2^\dagger, \tag{4.14}$$

in which

$$\Delta\equiv\beta^2/(\xi+\bar\xi)^2-\gamma^2/(1+|\xi|^2)^2. \tag{4.15}$$

It should be noted that the dependence on the arbitrary numbers a_1, a_2, and a_3 cancels out.

A new solution $\tilde g$ is obtained from

$$\tilde g=\chi(0)g,\quad \chi(0)=1-Q_1/\xi-\xi Q_2. \tag{4.16}$$

It is straightforward to check that $\tilde g$ is in Σ, i.e.,

$$(\tilde g)^\dagger=(\tilde g)^{-1}\quad\text{and}\quad t\tilde gt=(\tilde g)^{-1}. \tag{4.17}$$

So we express the new solution $\tilde g$ in terms of a projector $\widetilde P$,

$$\tilde g=t(1-2\widetilde P),$$
$$\widetilde P=P+\frac{1}{2|\xi|^2\Delta}\left\{\frac{\beta}{\xi+\bar\xi}[\bar\xi hh^\dagger(1-2P)+\xi(1-2P)hh^\dagger]-\frac{\gamma}{1+|\xi|^2}[(1-2P)hh^\dagger(1-2P)+|\xi|^2hh^\dagger]\right\}. \tag{4.18}$$

This is the explicit form of the new solution $\widetilde P$. A projector is characterized by its image. For the solution $P=P_{(f)}$ it is spanned by the matrix $X=X_{(f)}$ given in Eq. (3.5). For the new solution $\widetilde P$ its image is spanned by an $N\times m$ matrix $\widetilde X$,

$$\widetilde X=\alpha X+hh^\dagger X,\quad \widetilde P\widetilde X=\widetilde X, \tag{4.19}$$

in which α is a complex number

$$\alpha=\alpha(x;\xi)\equiv-\xi\beta/(\xi+\bar\xi)+\gamma/(1+|\xi|^2). \tag{4.20}$$

The new solution depends on the arbitrary complex number ξ and on the arbitrary constant vector u, Eq. (4.8). In order to clarify the meaning of these parameters we consider an extreme limit of the BT given by $|\xi|\to\infty$. In this limit we have

$$h\to h_0+O(1/|\xi|),\quad \beta\to\beta_0+O(1/|\xi|),$$
$$\gamma\to\gamma_0+O(1/|\xi|),\quad \Delta\to\beta_0^2/(\xi+\bar\xi)^2+O(1/|\xi|^3), \tag{4.21}$$

in which h_0 and β_0 are x independent whereas γ_0 is x dependent. The new solution $\widetilde P$ in this limit reads

$$\widetilde P=P+[(\xi+\bar\xi)^2/2|\xi|^2\beta_0]\{(\beta_0-\gamma_0)h_0h_0^\dagger-2\beta_0(\bar\xi h_0h_0^\dagger P+\xi Ph_0h_0^\dagger)\}+O(1/|\xi|). \tag{4.22}$$

If we introduce a constant SU(N) matrix S

$$S=1-[(\xi+\bar\xi)/\xi]\pi,\quad \pi=h_0h_0^\dagger/\beta_0,\quad SS^\dagger=1, \tag{4.23}$$

the new solution $\widetilde{P}$ is expressed as

$$\widetilde{P} = SPS^\dagger + O(1/|\xi|). \tag{4.24}$$

Namely, the Bäcklund transformation in this limit reduces to a trivial SU(N) transformation which is parametrized by ξ and u. Similar observation can be made for the other extreme case of $|\xi| \to 0$. Those $O(1/|\xi|)$ terms in Eq. (4.24) constitute the nontrivial and essential part of the Bäcklund transformation, which will be discussed in some detail for the case of the CP^{N-1} model.

V. CP^{N-1} CASE

Let us examine the new solution $\widetilde{P}$, Eq. (4.18), in more detail for the CP^{N-1} model. The complex projective space is a special case of the complex Grassmannian manifold $G(N,m)$ for $m = 1$. The dynamical variable of the CP^{N-1} model is again a projector P with Tr $P = 1$, which can also be expressed by a complex N-component unit vector z as

$$P = zz^\dagger, \quad z^\dagger \cdot z = 1. \tag{5.1}$$

The equation of motion is (2.16) and its solutions given in Sec. III are quite simple. In this case we start from an arbitrary N-component holomorphic vector f and its derivatives

$$f, \partial_+ f, \partial_+^2 f, \ldots, \partial_+^{N-1} f, \quad \partial_- f = 0, \tag{5.2}$$

and obtain an orthonormal basis of C^N

$$e_1, e_2, \ldots, e_N, \quad e_i^\dagger e_k = \delta_{ik}, \tag{5.3}$$

by using the Gram–Schmidt procedure in this order. By picking up any member of the orthonormal vectors (5.3), say e_j, we get a solution of the CP^{N-1} model,

$$P = P_{(j)} = e_j e_j^\dagger, \quad j = 1,2,\ldots,N. \tag{5.4}$$

Therefore we also call e_j a solution of the CP^{N-1} model.

If we start from a solution e_j and apply the Bäcklund transformation of the previous section, we find from Eq. (4.19) that

$$z' = \alpha e_j + hh^\dagger e_j \tag{5.5}$$

is again a solution up to the overall normalization. By using the explicit form of h

$$h = (1 - [2/(\bar{\xi} + 1)]\, P - [4\bar{\xi}/(\bar{\xi} + 1)^2] Q) u, \tag{5.6}$$

and

$$(e_j^\dagger u)(u^\dagger e_j) = u^\dagger P_{(j)} u = u^\dagger P u, \tag{5.7}$$

which is valid only for the CP^{N-1} model, we find

$$z' = \frac{\bar{\xi}(1-\xi)(1+\xi)}{(\xi + \bar{\xi})(1 + |\xi|^2)} \left(1 - \frac{4(\xi + \bar{\xi})(1 + |\xi|^2)}{(\xi + 1)^2(\bar{\xi} + 1)^2} u^\dagger Q u\right) e_j + \left(u - \frac{4\bar{\xi}}{(\bar{\xi} + 1)^2} Q u\right) \frac{\xi - 1}{\xi + 1} u^\dagger e_j. \tag{5.8}$$

The new solution z' (up to normalization) can be expressed in a simple form as

$$z' = S e_j - b(\xi)(u^\dagger Q u - Q u u^\dagger) e_j, \tag{5.9}$$

in which S is a constant $N \times N$ matrix depending on ξ and u,

$$S = S(\xi; u) = 1 - a(\xi) u u^\dagger, \tag{5.10}$$

and

$$a(\xi) = \xi(\xi + \bar{\xi})(1 + |\xi|^2)/|\xi|^2(\xi + 1)^2,$$
$$b(\xi) = 4(\xi + \bar{\xi})(1 + |\xi|^2)/(\xi + 1)^2(\bar{\xi} + 1)^2. \tag{5.11}$$

The matrix S is *not unitary for finite* ξ,

$$S^\dagger S = SS^\dagger = 1 - b(\xi) u u^\dagger, \tag{5.12}$$

but in the limit of $|\xi| \to \infty$ or 0, it is unitary.

The meaning of the transformation (5.9) is clear when we consider the $j = 1$ case, namely the instanton solution

$$e_1 = f/|f|, \quad Q = 0. \tag{5.13}$$

Then the new solution is

$$z' = Sf/|f|,$$

in which Sf is again a holomorphic vector and the new solution after normalization is again an instanton

$$z' = Sf/|Sf|, \quad \widetilde{P} = (Sf)(Sf)^\dagger/|Sf|^2. \tag{5.14}$$

Since the transformation $f \to Sf$ is not unitary, the new solution $\widetilde{P}$, Eq. (5.14), cannot be obtained from the old P by a unitary transformation, therefore it is nontrivial.

The generalization of the above argument to the noninstanton solutions is straightforward. It is summarized as the following theorem.

Theorem: The Bäcklund transformation of the solution e_j, Eq. (5.4), of the CP^{N-1} model is given by $\tilde{e}_j$, which is the jth member of the new orthonormal basis

$$\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_N. \tag{5.15}$$

They are obtained from the holomorphic vectors

$$\tilde{f} \equiv Sf, \quad \partial_+ \tilde{f}, \partial_+^2 \tilde{f}, \ldots, \partial_+^{N-1} \tilde{f}, \quad \partial_- \tilde{f} = 0, \tag{5.16}$$

by the Gram–Schmidt procedure.

A few comments are in order. Of course the "Bäcklund transformation" in the theorem means those restricted ones considered in the previous section. It is expected that a similar theorem holds for the multi-Bäcklund transformations [but still keeping the restriction (4.3)], since a multi-BT is essentially the same as a repeated minimal BT as is shown by Harnad *et al.*[5] On the other hand, it is known[17] that any finite action solution of the CP^{N-1} model can be written as one of the e_i's in (5.3) for some particular choice of the meromorphic vector f. And our result is consistent with it. So it is quite plausible that the above theorem holds for the most general Bäcklund transformations with certain modifications for the form of the matrix S. When we turn to the $G(N,m)$ model ($m \geq 2$) the situation is less transparent. It is as yet open whether the class of solutions of the $G(N,m)$ model considered in Sec. III and its generalizations[6] are closed or not under the Bäcklund transformations.

VI. NONCOMPACT MODELS

The generalization of the above argument to the noncompact Grassmannian sigma models[18] is straightforward. So we sketch the outline. The noncompact Grassmannian sigma model is a massless free field theory constrained on a manifold

$$\mathrm{SU}(m,n)/\mathrm{S}(\mathrm{U}(m) \times \mathrm{U}(n)), \tag{6.1}$$

which can be realized as a submanifold of SL(N,C),

$$\tilde{\Sigma} = \{g \in \mathrm{SL}(N,C) | g = t(g^\dagger)^{-1}t,\ g^{-1} = tgt\}. \qquad (6.2)$$

From the second condition we can introduce a projector P

$$g = t(1-2P), \quad P^2 = P, \quad \mathrm{Tr}\, P = m, \qquad (6.3)$$

whose Hermitian conjugation is determined by the first condition

$$P = tP^\dagger t. \qquad (6.4)$$

By substituting Eq. (6.3) to the principal chiral model equation (2.2), we get the same equation of motion as the compact model

$$[\partial_\mu \partial_\mu P, P] = 0. \qquad (6.5)$$

The construction of solutions is almost the same as before. We start from the set of holomorphic vectors given in Eqs. (3.2) and (3.3)

$$f_1,\ldots,f_m,\ldots,f_N, \qquad (6.6)$$

and orthonormalize them by the Gram–Schmidt procedure with respect to the inner product

$$(a,b) = a^\dagger t b, \qquad (6.7)$$

to obtain an orthonormal basis of C^N

$$e_1,e_2,\ldots,e_N, \quad (e_i,e_k) = e_i^\dagger t e_k = \delta_{ik}. \qquad (6.8)$$

By picking up m consecutive orthonormal vectors, we define as in Eq. (3.5) the $N \times m$ matrices

$$X_{(1)}, X_{(2)},\ldots,X_{(N-m+1)}. \qquad (6.9)$$

Each of them gives a solution of the noncompact Grassmannian sigma model through

$$P = P_{(j)} = tX_{(j)}X^\dagger_{(j)}. \qquad (6.10)$$

It is obvious that Eq. (6.10) satisfies the condtions (6.3) and (6.4). The proof that $P = P_{(j)}$ solves the equation of motion goes almost parallel with the compact case, except that the definition of Q is modified as

$$Q = Q_{(j)} = \sum_{k=1}^{j-1} te_k e_k^\dagger. \qquad (6.11)$$

Moreover, the linear scattering equation for the noncompact Grassmannian sigma model has exactly the same form as the compact one, Eq. (2.17), due to the same relation between g and P. So the linear scattering equation is solved by the same expression, Eq. (3.14), as the compact case. Therefore, one can also calculate the explicit forms of the Bäcklund transformations for the noncompact model, which we do not repeat here.

One of the important differences between the compact and noncompact models is the frequent occurrence of singularities in the solutions of the noncompact models. Since the inner product (6.7) defines a norm which is not positive definite, the normalization procedure introduces singularities wherever the norm of a vector vanishes. Therefore, even if we start from meromorphic vectors $f_1,\ldots,f_N$, the obtained solutions are not of finite action. This is also related to the nonexistence of topological invariants for the noncompact model.

APPENDIX: SUMMARY OF FORMAL CONSTRUCTION OF THE BT

The theory of multi-Bäcklund transformations for the principal chiral model together with various reductions has been formulated by Harnad *et al.*[5] Here we summarize some of their results appropriate for us. We start from a solution $g = g(x)$ of the principal $\mathrm{SL}(N,C)$ chiral model (2.1) together with the corresponding linear scattering equation

$$\begin{aligned}\partial_+\psi(x;\lambda) &= [(\partial_+ g)g^{-1}/(1+\lambda)]\psi(x;\lambda),\\ \partial_-\psi(x;\lambda) &= [(\partial_- g)g^{-1}/(1_-\lambda)]\psi(x;\lambda).\end{aligned} \qquad (\mathrm{A}1)$$

We try to construct a new solution $\tilde{g}$ through the dressing matrix $\chi(\lambda)$ introduced by Zakharov and Mikhailov,[2]

$$\tilde{g}(x) = \chi(\lambda = 0)g(x). \qquad (\mathrm{A}2)$$

If we assume that the dressing matrix has only simple poles,

$$\chi(\lambda) = 1 + \sum_{i=1}^{K} \frac{Q_i}{\lambda - \lambda_i}, \quad \chi^{-1}(\lambda) = 1 + \sum_{i=1}^{K} \frac{R_i}{\lambda - \mu_i}, \qquad (\mathrm{A}3)$$

the dressing matrix method of Zakharov and Mikhailov becomes equivalent to the multi-Bäcklund transformation method of Harnad *et al.* and the residues Q_i and R_i can be calculated from the solutions of the linear scattering problem (A1) through purely linear algebraic relations.

Theorem (Harnad et al.[5]): The residues Q_i and R_i are of the form

$$\begin{aligned}&Q_i = X_i F_i^\dagger, \quad R_i = H_i K_i^\dagger,\\ &X_i,F_i \in C^{N\times q_i}, \quad H_i,K_i \in C^{N\times r_i},\\ &q_i = \mathrm{rank}\, Q_i, \quad r_i = \mathrm{rank}\, R_i,\end{aligned} \qquad (\mathrm{A}4)$$

where the rectangular matrices F_i and H_i are determined from their "initial" values f_i and h_i by

$$F_i = (\psi_0(\lambda_i)^\dagger)^{-1} f_i, \quad H_i = \psi_0(\mu_i)h_i. \qquad (\mathrm{A}5)$$

In Eq. (A5), $\psi_0(\lambda) \equiv \psi_0(x;\lambda)$ is a solution of Eq. (A1) with the "initial" condition

$$\psi_0(x_0;\lambda) = 1, \qquad (\mathrm{A}6)$$

for an arbitrarily chosen "initial" point x_0. Another set of rectangular matrices X_i and K_i are solutions to the linear system

$$\sum_{i=1}^{K} X_i \Gamma_{ij} = H_j, \quad \sum_{i=1}^{K} K_i \Gamma^\dagger_{ij} = -F_j, \qquad (\mathrm{A}7)$$

with

$$\begin{aligned}\Gamma_{ij} &= F_i^\dagger H_j/(\lambda_i - \mu_j), \quad \text{if } \lambda_i \neq \mu_j,\\ \Gamma_{ii} &= -F_i^\dagger \psi_0'(\lambda_i)\psi_0^{-1}(\mu_i)H_i + f_i^\dagger c_i h_i, \quad \text{if } \lambda_i = \mu_i,\end{aligned} \qquad (\mathrm{A}8)$$

where $c_i \in C^{N\times N}$ is arbitrary and $f_i^\dagger h_i = 0$ in the latter case. It should be remarked that calculation of Q_i and R_i involves inverting the matrix Γ, a $(\Sigma_i q_i)\times(\Sigma_i r_i)$ matrix, which is highly nontrivial.

Next we proceed to the case of the complex Grassmannian sigma models. As is shown in Sec. II, when the reduction condition (2.9) is imposed on the linear scattering equation (2.3) it reduces to Eq. (2.17). This also imposes the following invariance conditions on the solutions of Eq. (2.17):

$$\psi(\lambda) = g\sigma_1(\psi(1/\lambda)) = gt\psi(1/\lambda)t,$$
$$\psi(\lambda) = \sigma_2(\psi(-\bar{\lambda})) = (\psi(-\bar{\lambda})^\dagger)^{-1}. \tag{A9}$$

These are translated into the invariance conditions on the dressing matrix χ

$$\chi(\lambda) = \tilde{g}t\chi(1/\lambda)tg^{-1}, \quad \chi(\lambda) = (\chi(-\bar{\lambda})^\dagger)^{-1}. \tag{A10}$$

These conditions are necessary and sufficient to ensure that the transformed solution $\tilde{g}$, Eq. (A2), belongs again to the Grassmann manifold or to Σ in Eq. (2.9). The conditions (A10) imply that a pole λ_i in $\chi(\lambda)$ should be accompanied by a pole $1/\lambda_i$ in $\chi(\lambda)$ and that it also requires a pair of poles $-\bar{\lambda}_i$ and $-1/\bar{\lambda}_i$ in $\chi^{-1}(\lambda)$. Let us denote them as

$$\lambda_i, \quad \lambda_{\hat{i}} \equiv 1/\lambda_i, \quad \text{for poles in } \chi,$$
$$\lambda_{\bar{i}} \equiv -\bar{\lambda}_i, \quad \lambda_{\hat{\bar{i}}} \equiv -1/\bar{\lambda}_i, \quad \text{for poles in } \chi^{-1}. \tag{A11}$$

The above conditions (A10) for χ also impose constraints on F_i and H_i in Eq. (A4), which are satisfied for all x provided the following conditions are met at $x = x_0$ by their initial values f_i and h_i. They are

$$\sigma_1; \quad [(g_0^{-1})^\dagger t f_i] = [f_{\hat{i}}], \quad [g_0 t h_i] = [h_{\hat{i}}],$$
$$\sigma_2; \quad [f_i] = [h_{\bar{i}}], \quad [f_{\hat{i}}] = [h_{\hat{\bar{i}}}], \tag{A12}$$

in which $[f_i] = [h_{\bar{i}}]$ means

$$f_i A_i = h_{\bar{i}},$$

for some constant matrix A_i and $g_0 = g(x_0)$.

[1]K. Pohlmeyer, Commun. Math. Phys. **46**, 207 (1976).
[2]V. E. Zakharov and A. V. Mikhailov, Sov. Phys. JETP **47**, 1017 (1978).
[3]H. Eichenherr and M. Forger, Nucl. Phys. B **155**, 381 (1979).
[4]A. T. Ogielski, M. K. Prasad, A. Sinha, and L. L. C. Wang, Phys. Lett. B **91**, 387 (1980).
[5]J. Harnad, Y. Saint-Aubin, and S. Snider, Commun. Math. Phys. **92**, 329 (1984).
[6]R. Sasaki, Phys. Lett. B **130**, 69 (1983); Zeits. Phys. C **24**, 163 (1984).
[7]V. L. Golo and A. M. Perelomov, Phys. Lett. B **79**, 112 (1978); H. Eichenherr, Nucl. Phys. B **146**, 215 (1978).
[8]E. Brezin, C. Itzykson, J. Zinn-Justin, and J. B. Zuber, Phys. Lett. B **82**, 442 (1979).
[9]S. Helgason, *Differential Geometry, Lie Groups and Symmetric Spaces* (Academic, New York, 1978).
[10]$[A,B]$ denotes the commutator of two matrices A and B.
[11]A. A. Belavin and A. M. Polyakov, JETP Lett. **22**, 245 (1975).
[12]A. D'Adda, P. Di Vecchia, and M. Lüscher, Nucl. Phys. B **146**, 63 (1978).
[13]A. J. MacFarlane, Phys. Lett. B **82**, 239 (1979).
[14]K. Fujii, T. Koikawa, and R. Sasaki, Progr. Theor. Phys. **71**, 338 (1984); K. Fujii and R. Sasaki, Progr. Theor. Phys. **71**, 831 (1984).
[15]A. M. Din and W. J. Zakrzewski, Nucl. Phys. B **237**, 461 (1984).
[16]Here we have suppressed the explicit ξ dependence of u, which is irrelevant so far as the minimal BT is concerned.
[17]A. M. Din and W. J. Zakrzewski, Nucl. Phys. B **174**, 397 (1980).
[18]A. C. Davis, A. J. MacFarlane, and J. W. van Holten, Phys. Lett. B **125**, 151 (1983); J. Ellis, C. Kounnas, and D. V. Nanopoulos, Nucl. Phys. B **247**, 373 (1984).

On the axial gauge: Ward identities and the separation of infrared and ultraviolet singularities by analytic regularization

Hoong-Chien Lee[a] and Michael S. Milgram[b]
Atomic Energy of Canada Limited, Chalk River Nuclear Laboratories, Chalk River, Ontario, Canada K0J 1J0

(Received 9 October 1984; accepted for publication 1 February 1985)

It is shown that the method of analytically regulating Yang–Mills theories in the axial gauge preserves gauge invariance. Two- and three-point Ward identities are computed and verified at the one-loop level. The method also permits a convenient and gauge-invariant separation of infrared and ultraviolet singularities in the axial gauge. In the axial gauge the renormalization constants $Z_3 = Z_1 = 1 + 11g^2C_2/(48\pi\epsilon^2)$, leading to a β function which is identical to that computed in the covariant ξ gauges.

I. INTRODUCTION

Recently, an analytic method[1,2] has been shown to be a powerful and elegant means to implement the principal-value prescription[3] for regulating and evaluating Feynman integrals occurring in Yang–Mills theories in the axial gauges.[4] Such gauges are defined by the constraint $n \cdot A^a = 0$, where A^a_μ is the vector gauge field, n_μ is an arbitrary constant vector, and the superscript a is an index for the gauge group. In particular, analytic representations[1,2] have been found for two-point functions in the general axial gauge ($n^2 \neq 0$) (Ref. 4), the light-cone gauge ($n^2 = 0$) (Ref. 5), and the special gauge defined by $p \cdot n = 0$ ($n^2 \neq 0$) (Ref. 2), where p_μ is the external momentum in the two-point functions.

Although there is a widely held belief (possibly caused by using the work "analytic" to describe different and inequivalent regularization methods[6]) that "analytic regularization does not preserve gauge invariance," it will be demonstrated that the contrary is true for the new analytic method. The preservation of gauge invariance depends on the fact that the method preserves such algebraic properties as commutativity and associativity of operations in the Feynman integrals and that in the appropriate limit the new method, for both the infinite and regular parts of a Feynman integral, yields results that are identical to those obtainable from dimensional regularization. For the light-cone gauge, this has already been demonstrated by verifying two- and three-point Ward identities[7] at the one-loop level.[8]

In Sec. II, we extend this study by verifying these identities in the general axial gauge ($n^2 \neq 0$). Here again, we find that the preservation of algebraic properties mentioned earlier is sufficient to guarantee that Ward identities will be upheld, even before the Feynman integrals involved in the identities are evaluated.

The capability of distinguishing infrared from ultraviolet singularities is another one of the appealing properties of analytic regularization. The lack of this capability in dimensional regularization[9] has been the cause of considerable inconvenience in practical calculations using that technique. In Sec. III, using integrals appearing in the three-point Ward identities as examples, we show how the analytic method can easily be employed to separate the two types of singularities. We also show that when these two types of singularities are not distinguished, the cancellation between the two is the direct cause of the vanishing of some tadpoles.[10] In other words, if infrared and ultraviolet singularities are separated, then not all tadpoles vanish.

In Sec. IV we show that the β function, which can be computed from the renormalization constants Z_3 and Z_1, associated with the self-energy and the three-vertex, respectively, are identical in the axial gauge and the covariant ξ gauges to lowest order. Furthermore, in contrast to the ξ gauges, the equality $Z_1 = Z_3$ in the axial gauge allows the β function to be derived directly from the self-energy.

In Sec. V, we compare our results for the axial gauge with those obtained previously for the light-cone gauge.[8] Briefly, other than being ghost-free, the axial gauge shares the properties of the covariant gauges,[11] but does not have the pecularities possessed by the light-cone gauge. On the other hand, computations in the light-cone gauge are much less tedious.

In the following, we briefly review the analytic representation for the "two-point" integrals—integrals with one external momentum—defined by

$$S_{2\omega}(p,n;\kappa,\mu,\nu,s) \equiv \int d^{2\omega}q\,[(p-q)^2]^\kappa (q^2)^\mu (q\cdot n)^{2\nu+s}, \tag{1}$$

where $s = 0$ or 1 and ω, κ, μ, and ν are continuous variables. Feynman integrals in four-dimensional Euclidean space (Minkowski space is reached by analytic continuation) that are sometimes divergent and therefore ill defined, correspond to those in (1) when κ, μ, and ν are integers and $\omega = 2$; these form a subset of (1) which we call primal integrals. Methods of analytic regularization[1] were used to find a representation for (1) in terms of a Meijer G function[12]

[a] Theoretical Physics Branch.
[b] Applied Mathematics Branch.

 0022-2488/85/071793-12$02.50

$$
\begin{aligned}
S_{2\omega}(p,n;\kappa,\mu,\nu,s) &= \frac{\pi^2(p^2)^{\omega+\kappa+\mu+\nu}(n^2)^{\nu}(p\cdot n)^s\Gamma(s+\nu+\frac{1}{2})}{\Gamma(-\kappa)\Gamma(-\mu)\Gamma(-\nu)\Gamma(2\omega+\kappa+\mu+2\nu+s)} \\
&\quad\times G^{2,3}_{3,3}\left(y\,\middle|\,\begin{matrix}1-\omega-\mu-\nu-s,1+\omega+\kappa+\mu+\nu,1+\nu;\\ 0,\omega+\kappa+\nu;1/2-s\end{matrix}\right), \quad |y|<1, \\
&= \frac{\pi^2(p^2)^{\omega+\kappa+\mu}(p\cdot n)^{2\nu+s}\Gamma(s+\nu+\frac{1}{2})}{\Gamma(-\kappa)\Gamma(-\mu)\Gamma(-\nu)\Gamma(2\omega+\kappa+\mu+2\nu+s)} \\
&\quad\times G^{3,2}_{3,3}\left(1/y\,\middle|\,\begin{matrix}1+\nu,1-\omega-\kappa;1/2+s+\nu\\ 0,-\omega-\kappa-\mu,\omega+\mu+2\nu+s;\end{matrix}\right), \quad |y|>1,
\end{aligned} \tag{2}
$$

where $y\equiv(p\cdot n)^2/(p^2n^2)$. The right-hand side of (2) is a well-defined, analytic function of all its variables and, when $n^2\neq 0$, has at most simple poles in the (ω,κ,μ,ν) plane.

The evaluation of any primal integral in terms of the independent variables y, p, n and infinitesimals ϵ_i, which label the singularities, now becomes a well-defined mechanistic process, which is discussed elsewhere.[1] Tables of primal integrals have also been prepared.[13]

II. WARD IDENTITIES

We shall verify both the one-loop radiative corrections to the two-point Ward identity

$$p_\mu \Pi^t_{\mu\nu}(p)=\Pi^t_{\nu\mu}(p)p_\mu=0, \tag{3}$$

as well as a special case of the three-point Ward identity

$$
\begin{aligned}
ip_\lambda\Gamma^{abc}_{\lambda\mu\nu}(p,-p,0) &\equiv if^{abc}p_\lambda\Gamma_{\lambda\mu\nu}(p,-p,0) \\
&= gf^{abc}\Pi^t_{\mu\nu}(p),
\end{aligned} \tag{4}
$$

where $\Pi^t_{\mu\nu}(p)$ is the self-energy less its $O(g^0)$ longitudinal term

$$\Pi^t_{\mu\nu}(p)=\Pi_{\mu\nu}(p)-(i/\xi)n_\mu n_\nu, \tag{5}$$

g is the Yang–Mills coupling, f^{abc} are the structure constants of the gauge group, and ξ is the parameter of the gauge-fixing Lagrangian $-(1/2\xi)(n\cdot A^a)^2$; the ghost-free axial gauge is realized in the limit $\xi\to 0$. For a derivation of the identities (3) and (4) see Ref. 8.

Equation (3) expresses the well-known notion that $\Pi^t_{\mu\nu}(p)$ is transverse to p_μ. Because the subtracted term in (5) appears only in the zeroth order (in g), (3) also implies that all radiative corrections to $\Pi_{\mu\nu}$ are transverse. The transversility of $\Pi^t_{\mu\nu}$ may be expressed by writing

$$\Pi^t_{\mu\nu}(p)=-i[\Pi_0(p)P_{\mu\nu}+\Pi_1(p)N_{\mu\nu}], \tag{6}$$

where $P_{\mu\nu}$ and $N_{\mu\nu}$ are two linearly independent tensors transverse to p_μ:

$$P_{\mu\nu}\equiv p^2\delta_{\mu\nu}-p_\mu p_\nu, \tag{7a}$$

$$N_{\mu\nu}\equiv p_\mu p_\nu-(p_\mu n_\nu+p_\nu n_\mu)p^2/p\cdot n+n_\mu n_\nu p^4/(p\cdot n)^2. \tag{7b}$$

Our first task is to verify that, to one-loop order (see Fig. 1), $\Pi^t_{\mu\nu}$ indeed has the form (6), and to also compute the coefficients $\Pi^{(1)}_0$ and $\Pi^{(1)}_1$ in the limit $\xi=0$. (The superscript denotes that these are one-loop results.)

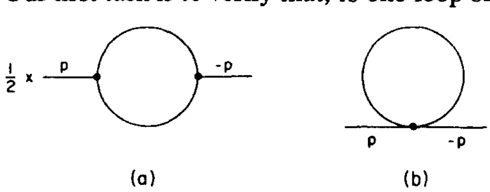


FIG. 1. Diagrams for $O(g^2)$ self-energy. Part (b) is tadpolelike and vanishes only in the limit (10).

The symbol $\Gamma_{\lambda\mu\nu}(p,-p,0)$ in (4) represents the three-vertex with one external line carrying zero momentum; this special momentum configuration reduces the three-vertex to a two-point function. Feynman diagrams for the one-loop radiative corrections to $\Gamma_{\lambda\mu\nu}$ are shown in Fig. 2.

We now digress to explain how it is possible to evaluate the tensor $\Pi^t_{\mu\nu}$, which clearly depends on integrals with nonscalar integrands, using only the scalar integral (1). We parenthetically note that it would have been impossible to derive a viable analytic regularization with generalized exponents, if it were necessary to find representations for integrals with tensorial integrands.

There are four linearly independent, symmetric, rank-two tensors in the axial gauge: $\delta_{\mu\nu}, p_\mu p_\nu, p_\mu n_\nu+p_\nu n_\mu$, and $n_\mu n_\nu$. Therefore any symmetric rank-two tensor such as $\Pi^t_{\mu\nu}$ can be expressed as

$$
\begin{aligned}
\Pi^t_{\mu\nu} = \tfrac{1}{2}[&A_1\delta_{\mu\nu}+A_2 p_\mu p_\nu/p^2 \\
&+A_3(p_\mu n_\nu+p_\nu n_\mu)/p\cdot n \\
&+p^2A_4n_\mu n_\nu/(p\cdot n)^2]/(\zeta-1),
\end{aligned} \tag{8}
$$

where A_i are scalar functions of p^2, n^2, and $p\cdot n$, and $\zeta=1/y$. These functions can be found by contracting $\Pi^t_{\mu\nu}$ as follows:

$$a_1=\Pi^t_{\mu\nu}\delta_{\mu\nu}, \tag{9a}$$

$$a_2=\Pi^t_{\mu\nu}p_\mu p_\nu/p^2, \tag{9b}$$

$$a_3=\Pi^t_{\mu\nu}p_\mu n_\nu/p\cdot n, \tag{9c}$$

$$a_4=\Pi^t_{\mu\nu}n_\mu n_\nu p^2/(p\cdot n)^2, \tag{9d}$$

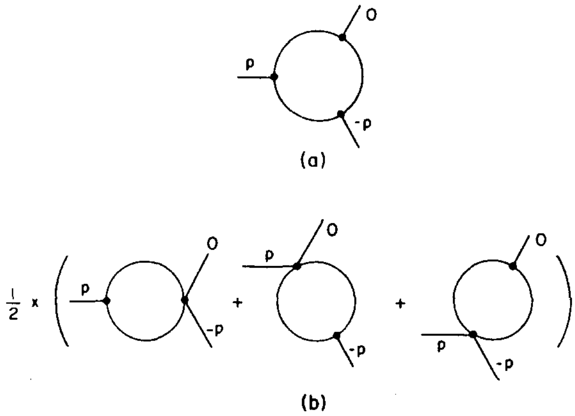


FIG. 2. Diagrams for $O(g^2)$ three-vertex. All momenta flow in. The last diagram in part (b) vanishes identically because of gauge group and Lorentz symmetry.

and solving the resulting set of equations. The relations between a_i and A_i are given in Appendix A. This technique can be generalized for computing tensors of any rank. Since any such tensor can be evaluated by calculating the corresponding scalars a_i, it follows that (for two-point functions) only scalar integrals of the type (1) need ever be computed.

We now return to follow the simpler of two paths. On this path all primal integrals are evaluated in a limit in which the infrared and ultraviolet singularities are not distinguished. (In the next section we consider another limit in which the two types of singularities are distinguished.) For now, we let

$$\kappa = K, \quad \mu = M, \quad \nu = N, \quad \omega = 2 + \epsilon, \tag{10}$$

where ϵ is small but finite. This limit is equivalent[1] to that obtainable with the principal-value prescription used in conjunction with dimensional regularization. We emphasize that with (10) and analytic regularization, all tadpole-like integrals, namely primal integrals satisfying either one or both of the conditions

$$\text{(i)} \quad K \geqslant 0, \tag{11a}$$

$$\text{(ii)} \quad M \geqslant 0, \quad N \geqslant 0, \tag{11b}$$

vanish. In comparison, dimensionless regularization, having only one generalized variable ω, is insufficiently powerful to regulate such integrals. Conventionally,[10] such integrals are simply assumed to be zero[14] in dimensional regularization.

Using the method described above, we find that $\Pi^i_{\mu\nu}$ indeed has the form of the right-hand side of (6) [this implies that the a_2 and a_3 of (9) vanish, or equivalently, $A_1 + A_2 = A_4 = -A_3$]. The A_1 and A_3 are listed in Table IV of Appendix B in terms of reduced primal integrals defined by

$$\widehat{S}(K,M,N,s) \equiv \pi^{-\omega}(p^2)^{-K-M-2}(p\cdot n)^{-2N-s} \times S_4(p,n;K,M,N,s). \tag{12}$$

Because of space limitations, all integrals satisfying (11) have been omitted from Table IV.

In the limit (10), using (2) we find

$$\Pi_0 = \frac{g^2C_2}{32\pi^2}\frac{1}{1-\zeta}\left[\frac{22}{3\hat{e}}(1-\zeta) - \ln\left(\frac{4}{\zeta}\right)(8 - 6\zeta + \zeta^2) - \frac{62}{9} + \frac{44\zeta}{9} + 2\zeta^2 + \left(\frac{8}{\zeta} - 8 + 2\zeta - \frac{\zeta^2}{2}\right)Z\right], \tag{13}$$

$$\Pi_1 = \frac{g^2C_2}{32\pi^2}\frac{1}{1-\zeta} \times\left[-\frac{10}{3} + 2\zeta + \ln\left(\frac{4}{\zeta}\right)\left(7 - \zeta - \frac{9}{1-\zeta}\right) - \frac{1}{2}\left(\frac{16}{\zeta} - 5 + \zeta - \frac{9}{1-\zeta}\right)Z\right], \tag{14}$$

where

$$1/\hat{e} \equiv 1/\epsilon + \ln p^2 + \gamma, \tag{15}$$

$\zeta = 1/y$, γ is Euler's constant, and Z is defined in Table II.

Noticeably absent from Π_1 is an infinite part. This will allow us to comment in Sec. III on the multiplicative renormalizability of the Yang–Mills theory in the axial gauge after having ascertained that the finiteness of Π_1 is not the result of a cancellation between the as yet undistinguished infrared and ultraviolet singularities.

We now consider the three-point Ward identity. To verify (4), it is sufficient to expand the left-hand side, which is again a symmetric rank-two tensor, as in (9). However, to learn more about the three-vertex $\Gamma_{\mu\nu\lambda}$, we shall instead calculate it explicitly. This vertex function has the symmetry relation

$$\Gamma_{\lambda\mu\nu}(p,-p,0) = -\Gamma_{\mu\lambda\nu}(-p,p,0), \tag{16}$$

showing that it can be expanded in terms of ten independent tensors $T^{(i)}_{\lambda\mu\nu}$ given in Table I:

$$\Gamma_{\lambda\mu\nu}(p,-p,0) = \frac{g}{2(\zeta-1)}\sum_{i=1}^{10} B_i\, T^{(i)}_{\lambda\mu\nu}. \tag{17}$$

In a manner similar to (9), we compute the ten scalar functions B_i by contracting both sides of (17) using the symbol b_i to label the contracted left-hand sides. The linear relations between the sets B_i and b_i are given in Appendix A. The formidable expressions for B_i, in the form of a sum of primal integrals with tadpoles omitted, are given in Table V of Appendix B. When the integrals are evaluated using (2) in the limit (10) the B_i's have values given in Table II.

To satisfy (4), the B_i's must first satisfy the "transversality" relations

$$B_3 + (B_6 + B_7)/(\zeta-1) = B_4 + (B_5 + B_8)/(\zeta-1) = -(B_6 + B_{10})/(\zeta-1) \tag{18a}$$

and

$$B_1 + 2B_2 + B_3 + B_4 = -(\zeta-1)(B_6 + B_7 + B_8 + B_9), \tag{18b}$$

TABLE I. The ten tensors $T^i_{\lambda\mu\nu}$ of (17) and the operators $O^i_{\lambda\mu\nu}$ of (A4).

i	$T^i_{\lambda\mu\nu}$	$O^i_{\lambda\mu\nu}$
1	$\delta_{\lambda\mu}\, p_\nu$	$\delta_{\lambda\mu}\, p_\nu/p^2$
2	$\delta_{\mu\nu}\, p_\lambda + \delta_{\lambda\nu}\, p_\mu$	$\delta_{\mu\nu}\, p_\lambda/p^2$
3	$p^2/p\cdot n\delta_{\lambda\mu}\, n_\nu$	$\delta_{\lambda\mu}\, n_\nu/(p\cdot n)$
4	$p^2/p\cdot n(\delta_{\mu\nu}\, n_\lambda + \delta_{\nu\lambda}\, n_\mu)$	$\delta_{\mu\nu}\, n_\lambda/p\cdot n$
5	$p^2/(p\cdot n)^2 n_\lambda n_\mu\, p_\nu/(\zeta-1)$	$n_\lambda n_\mu\, p_\nu/(p\cdot n)^2/(\zeta-1)$
6	$p^2/(p\cdot n)^2(n_\mu n_\nu\, p_\lambda + n_\nu\, n_\lambda\, p_\mu)/(\zeta-1)$	$n_\mu\, n_\nu\, p_\lambda/(p\cdot n)^2/(\zeta-1)$
7	$p_\lambda\, p_\mu\, n_\nu/(p\cdot n)/(\zeta-1)$	$p_\lambda\, p_\mu\, n_\nu/p^2/p\cdot n(\zeta-1)$
8	$(p_\mu\, p_\nu\, n_\lambda + p_\nu\, p_\lambda\, n_\mu)/(p\cdot n)/(\zeta-1)$	$p_\mu\, p_\nu\, n_\lambda/p^2/p\cdot n/(\zeta-1)$
9	$p_\lambda\, p_\mu\, p_\nu/p^2/(\zeta-1)$	$p_\lambda\, p_\mu\, p_\nu/p^4/(\zeta-1)$
10	$p^4/(p\cdot n)^3 n_\lambda n_\mu n_\nu/(\zeta-1)$	$p^2 n_\lambda n_\mu n_\nu/(p\cdot n)^3/(\zeta-1)$

TABLE II. Coefficients (in units of $g^2C_2/32\pi^2$) B_i of (17) after evaluating all the integrals in Table V.[a] Note: $1/\hat{\epsilon} = 1/\epsilon + \gamma + \log(p^2)$.

$$
\begin{aligned}
B_1 = {} & -\tfrac{88}{3}*1/\hat{e}*(1-1/y) - \tfrac{304}{9} - 20/y^2 + 376/(9y) \\
& + \log(4y)*(46 + 10/y^2 - 34/y - 18y/(1-y)) + Z*(25 - 32y + 5/y^2 - 15/y + 9y/(1-y)), \\
B_2 = {} & \tfrac{44}{3}*1/\hat{e}*(1 - 1y) - \tfrac{268}{9} + 4/y^2 + 124/(9y) \\
& + \log(4y)*(-2 - 2/y^2 + 10/y - 18y/(1-y)) + Z*(-11 - 1/y^2 + 3/y + 9y/(1-y)), \\
B_3 = {} & 32 + 12/y^2 - 32/y + \log(4y)*(-14 - 6/y^2 + 10/y + 18y/(1-y)) + z*(7 - 3/y^2 + 7/y - 9y/(1-y)), \\
B_4 = {} & 16 - 4/y + \log(4y)*(-14 + 2/y + 18y/(1-y)) + Z*(-5 + 16y + 1/y - 9y/(1-y)), \\
B_5 = {} & \tfrac{328}{3} + 20/y^2 - 208/(3y) + \log(4y)*(-70 - 10/y^2 + 50/y + 90y/(1-y)) \\
& + Z*(-37 + 64y - 5/y^2 + 23/y - 45y/(1-y)), \\
B_6 = {} & \tfrac{268}{3} + 8/y^2 - 112/(3y) + \log(4y)*(26 - 4/y^2 + 2/y + 90y/(1-y)) \\
& + Z*(-1 + 16y - 2/y^2 + 5/y - 45y/(1-y)), \\
B_7 = {} & -\tfrac{152}{3} - 12/y^3 + 40/y^2 - 112/(3y) + \log(4y)*(-90 + 6/y^3 - 14/y^2 + 38/y - 90y/(1-y)) \\
& + Z*(5 + 3/y^3 - 9/y^2 + 1/y + 45y/(1-y)), \\
B_8 = {} & -\tfrac{206}{3} - 12/y^2 + 116/(3y) + \log(4y)*(6 + 6/y^2 - 18/y - 90y/(1-y)) \\
& + Z*(13 - 32y + 3/y^2 - 11/y + 45y/(1-y)), \\
B_9 = {} & \tfrac{8}{3} + 12/y^3 - 208/(3y^2) + 344/(3y) + \log(4y)*(90 - 6/y^3 + 14/y^2 - 38/y + 90y/(1-y)) \\
& + Z*(-5 - 3/y^3 + 9/y^2 - 1/y - 45y/(1-y)), \\
B_{10} = {} & -96 - 12/y^2 + 48/y + \log(4y)*(6 + 6/y^2 - 18/y - 90y/(1-y)) \\
& + Z*(13 - 32y + 3/y^2 - 11/y + 45y/(1-y)).
\end{aligned}
$$

[a] In the above tables, we use the symbol Z to denote the infinite series

$$
\begin{aligned}
Z &= 2\sum_{l=0}^{\infty} \frac{(1)_l\, y^l}{(3/2)_l}\left[\ln y - \psi(1+l) + \psi\left(\frac{3}{2}+l\right)\right], \quad |y|\leqslant 1 \\
&= \frac{1}{2\sqrt{\pi}}\sum_{l=0}^{\infty}\frac{(\frac{1}{2})_l\, y^{-l-1}}{(1)_l}\left\{\left[\psi\left(\frac{1}{2}-l\right) - \psi(1+l) - \ln y\right]^2 + 2\psi'\left(\frac{1}{2}\right) - \psi'(1+l) - \psi'\left(\frac{1}{2}-l\right)\right\}, \quad |y|>1 \\
&= \sqrt{\pi}\, G^{2,3}_{3,3}\left(y \,\middle|\, \begin{matrix} 0,0,0; \\ 0,0;-\frac{1}{2} \end{matrix}\right), \quad \text{for all } y.
\end{aligned}
$$

which they do. Then they must satisfy

$$B_3 + (B_6 + B_7)/(\zeta - 1) = -A_3, \tag{18c}$$

$$B_2 + B_4 = -A_1, \tag{18d}$$

which they also do, both in terms of primal integrals (from Tables IV and V) and when the intervals are evaluated [from (13), (14), and Table II]. This concludes our demonstration by explicit computation that the two- and three-point Ward identities are satisfied in our method, in the limit (10).

We wish to emphasize a rather appealing feature of the analytic method, that is, its ability to display the equality (4) *before* the primal integrals on both sides of the equations have been evaluated[15] by any means of regularization. In our opinion, it is this feature that truly demonstrates the superior properties of the analytic method: it allows one to manipulate divergent integrals using the same formal rules of algebra as are used for finite integrals.

In Table II, the absence of infinite parts in B_i, $i = 3$–10, is crucial for multiplicative renormalizability. However, a final assessment of this desirable result again must be deferred until after we have separated the infrared and ultraviolet singularities in the primal integrals. This is done in the next section.

III. TADPOLES AND THE SEPARATION OF INFRARED AND ULTRAVIOLET SINGULARITIES

The capability of evaluating tadpole integrals[10] defined by the conditions (11) is special to analytic regularization. To illustrate why dimensional regularization is incapable of regulating a tadpole integral consider a simple case of (11) with $K = N = 0$, and M arbitrary

$$I(\omega) \equiv \int d^{2\omega}q(q^2)^M. \tag{19}$$

Clearly, if $\mathrm{Re}(\omega + M) \geqslant 0$, $I(\omega)$ is ultraviolet divergent and if $\mathrm{Re}(\omega + M) \leqslant 0$, $I(\omega)$ is infrared divergent. So regardless of the value of M, there does not exist a region in the entire complex ω plane in which $I(\omega)$ can be well defined. Therefore, as a function of ω alone, $I(\omega)$ cannot be evaluated by analytic continuation. In other words, the integral cannot be regulated by dimensional regularization.

In our analytic regularization, all two-point integrals (1), of which (19) is but a special case, are considered to be functions over the complex (ω,κ,μ,ν) (hyper) space, not merely as functions of ω. Although no region in the ω plane exists in which $I(\omega)$ is regular, there always exists a region in the (ω,κ,μ,ν) space in which the generalized integral is regular.

Therefore it is always possible to evaluate (1) by analytic continuation. This is why dimension regularization must set the right-hand side of (19) equal to zero (or some other value) by decree, whereas in analytic regularization, the value of any given primal integral with parameters K, M, and N is determined by the limiting processes $(\omega,\kappa,\mu,\nu) \to (\omega,K,M,N) \to (2,K,M,N)$. In the limiting process (10) used in the last section, all tadpole integrals vanish. See Ref. 1 for more details.

We now return to (19) and consider the special case $M = -2$. By power counting it is clear that the integral is both ultraviolet and infrared divergent when ω approaches 2. Why then should it be finite and zero in analytic regularization? The answer is simple: In the special limiting process (10) both the infinite and the finite parts associated, respectively, with the infrared and ultraviolet singularities exactly cancel! (See Table 3 of Ref. 13 for tabulation of this and other integrals.) That these two types of singularities can cancel each other is certainly consistent with the spirit of dimensional regularization because in this method all singularities must be expressed as $1/\epsilon$ poles and are therefore indistinguishable. Similarly, it has already been shown elsewhere[1] that the vanishing of many tadpole integrals defined by (11) is caused by the cancellation between infrared and ultraviolet singularities.[16]

In dimensional regularization, with sufficient knowledge of what the outcome should be, it is possible to separate infrared from ultraviolet singularities.[17] A common practice is to to assign a mass to each massless particle. Other than being cumbersome, this procedure is also very tricky and must be practiced with great care because it does not preserve gauge invariance.

In our analytic method, infrared and ultraviolet singularities can be easily separated by a judicious choice of limiting procedure. The simplest limit that serves this purpose (but does not distinguish the two types of infrared singularities[16]) is

$$\kappa = K+\rho, \quad \mu = M+\rho, \quad \nu = N,$$
$$\omega = 2+\epsilon, \quad \rho \to 0, \tag{20}$$

with ϵ small but finite. In this limit, ultraviolet singularities are characterized by the pole

$$1/\hat{\epsilon}_1 = -1/\epsilon_1 + \ln p^2 + \gamma, \quad \epsilon_1 = -2\rho - \epsilon, \tag{21a}$$

and both infrared singularities by

$$1/\hat{\epsilon}_0 = 1/\epsilon_0 + \ln p^2 + \gamma, \quad \epsilon_0 = \epsilon + \rho. \tag{21b}$$

Tadpole integrals that are both ultraviolet and infrared divergent are proportional to $1/\hat{\epsilon}_0 - 1/\hat{\epsilon}_1$. The limit (10) is a special case of (20) with $\rho = 0$, for in that limit such integrals vanish:

$$(1/\hat{\epsilon}_0 - 1/\hat{\epsilon}_1)_{\rho=0} = 1/\epsilon - 1/\epsilon = 0. \tag{22}$$

We now use limit (20) to evaluate the integrals in Tables IV and V. In this limit tadpole diagrams such as (b) of Fig. 1 contain terms that do not automatically vanish. However, the ultraviolet and infrared singularities of the regulated integrals describing Fig. (1b) cancel among themselves, so that in the limit (20) this diagram does vanish. Remarkably, the results for the remaining diagrams are identical to those in (13), (14), and Table II except that all pole terms $(1/\hat{\epsilon})$ therein are now replaced by ultraviolet poles $(1/\hat{\epsilon}_1)$; all infrared singularities have cancelled! We do not know whether this is a general result [i.e., whether we can use the simpler limit (10) in which all tadpole integrals vanish and treat all poles as ultraviolet singularities] or if not, what is the reason for this remarkable cancellation of infrared singularities at the one-loop level. We do note, however, that the Ward identities (18) do not appear to be manifestly satisfied prior to regularization, if all tadpole integrals and diagrams are retained.

IV. THE β FUNCTION

Our calculation shows that at the one-loop level, all infinite parts in $\Pi_{\mu\nu}$ and $\Gamma_{\lambda\mu\nu}$ are of ultraviolet origin, and of the operators generated by radiative correction, only those that appear in the original Lagrangian at the tree (no-loop) level—$\Pi_0 P_{\mu\nu}$ in $\Pi_{\mu\nu}$ and $B_i O^{(i)}_{\lambda\mu\nu}$, $i = 1,2,3$ in $\Gamma_{\lambda\mu\nu}$—have infinite parts. These results indicate that Yang–Mills theories are multiplicatively renormalizable in the axial gauge. That is, the infinite parts generated by radiative corrections can be absorbed into renormalization constants Z_1 and Z_3 that rescale the gauge field, the coupling constant, and the gauge parameter according to

$$A^a_\mu \to \bar{A}^a_\mu = Z_3^{-1/2} A^a_\mu, \tag{23a}$$

$$g \to \bar{g} = Z_1 Z_3^{-3/2} g \equiv Z_g\, g, \tag{23b}$$

so that the bare Lagrangian is

$$\mathscr{L}(\bar{A},\bar{g},\bar{\xi}) = -\tfrac{1}{4}[F^a_{\mu\nu}(\bar{A},\bar{g})]^2 - (1/2\bar{\xi})(n\cdot\bar{A}^a_\mu)^2, \tag{24}$$

and from (13) and Table II

$$Z_3^{\text{axial}} = Z_1^{\text{axial}} = 1 + \frac{g^2 C_2}{16\pi^2}\frac{11}{3}\left[\frac{1}{\epsilon} + \ln\left(\frac{p^2}{\lambda^2}\right) + \cdots\right], \tag{25}$$

where the explicit dependence on an arbitrary momentum scale λ is displayed. As expected, the two renormalization constants are identical in the axial gauge, an outcome which is not true in the covariant ξ gauges, where[18]

$$Z_1^\xi - 1 = \frac{g^2 C_2}{32\pi^2}\left(\frac{17}{6} - \frac{3}{2}\xi\right)\left[\frac{1}{\epsilon} + \cdots\right],$$
$$Z_3^\xi - 1 = \frac{g^2 C_2}{32\pi^2}\left(\frac{13}{3} - \xi\right)\left[\frac{1}{\epsilon} + \cdots\right]. \tag{26}$$

Although the renormalization constants are gauge dependent, the β function, or the logarithmic derivative of the renormalized coupling constant,

$$\beta(g) \equiv \lambda\frac{\partial g}{\partial\lambda} = 2gZ_g^{-1}\frac{\partial Z_g}{\partial(1/\epsilon)}, \tag{27}$$

should be gauge independent. This is indeed true since

$$Z_g^{\text{axial}} = (Z_3^{\text{axial}})^{-1/2} = Z_g^\xi = 1 - \frac{g^2 C_2}{32\pi^2}\frac{11}{3\epsilon} + O(g^4), \tag{28}$$

which leads to the equality

$$\beta(g)^{\text{axial}} = \beta(g)^\xi = -\frac{g^3 C_2}{16\pi^2}\frac{11}{3\epsilon} + O(g^5). \tag{29}$$

The point to be noted here is that unlike the ξ gauges, in the axial gauge the equality (25) allows the β function to be derived directly from radiative corrections to the self-energy.

V. AXIAL GAUGE VERSUS LIGHT-CONE GAUGE

A comparison of our result for the axial gauge with results obtained previously[2,8] for the principal-value prescription of the light-cone gauge ($n^2 = 0$) is summarized in Table III, with the following comments.

(i) The advantages for the light-cone gauge is the simplicity of the propagator and the extreme ease with which Feynman integrals can be evaluated. On the other hand, we emphasize that although integrals in the axial gauge are comparatively more cumbersome to compute, with the aid of analytic regularization, such computations are not the kind of brutal undertaking they used to be when the principal-value prescription was used. In any case, two-point integrals in both the axial and light-cone gauges have now been evaluated and tabulated.[13]

(ii) In this paper and in Ref. 8, analytic regularization has been employed to verify that two- and three-point Ward identities in Yang–Mills theories are true in both gauges.

(iii) A pecularity of the light-cone gauge is that some of the divergences generated by one-loop radiative corrections manifest themselves[8,19] as double poles $[O(1/\epsilon^2)]$. This effect is directly caused by the coalescence of ultraviolet divergences with one type of infrared divergence inherent in the analytic regularization of this gauge; only one other Lorentz invariant regularization of this gauge exists,[20] which manifests a nonlocal infinite part residual to the double pole. In this aspect the axial gauge is normal: one-loop corrections generate only single poles and local interactions.

(iv) We have shown that in the axial gauge infrared and ultraviolet singularities can be separated by letting the generalized integrals approach a given primal integral in an appropriate way [see (20)]. In contrast, for the same reason given in (iii) these singularities cannot be separated in the light-cone gauge. Indeed, when the limit (20) is used to evaluate the integrals in the three-point Ward identity, we find that the identity is no longer true. The only limit that we believe does not lead to any incorrect result in the light-cone gauge is (10), in which all tadpole-like integrals (defined by $K \geqslant 0$ and/or $M \geqslant 0$ in this gauge) vanish.

(v) In the axial gauge, infinite parts generated by one-loop corrections occur only in operators associated with the original Lagrangian at the tree (zero-loop) level. Therefore, as is well-known, the theory in this gauge is multiplicatively renormalizable. In contrast, in the light-cone gauge, new operators generated by radiative corrections also have infinite parts.[8,19,21] Consequently, the theory in this gauge, assuming it is renormalizable, is not multiplicatively renormalizable. The renormalization program in the light-cone gauge needs to be thoroughly studied.

Note added in proof: All comments in this paper pertaining to the peculiarities of the light-cone gauge refer to the principal-value prescription of that gauge. Recent calculations by the authors (Chalk River preprint CRNL-TP-85-II-11) have shown that the Mandelstam–Leibbrandt prescription of the gauge does not share such peculiarities; in particular it is one-loop renormalizable.

ACKNOWLEDGMENT

One of us (HCL) ackowledges the hospitality of National Taiwan University, at which part of this work was completed.

APPENDIX A: CALCULATION OF SCALAR FUNCTIONS A_i AND B_i

Define the scalar functions A_i by the general expansion for the self-energy (8) and compute the scalar quantities a_i defined in (9). Substituting (8) into (9) yields the linear relations

$$a_i = (U_\Pi^{-1})_{ij} A_j, \tag{A1}$$

which have inverse relations

$$A_i = (U_\Pi)_{ij}\, a_j. \tag{A2}$$

For general axial gauges, defining $\zeta = p^2 n^2/(p \cdot n)^2 = 1/y$, we find

$$U_\Pi = \frac{1}{\zeta - 1}\begin{bmatrix} (\zeta-1)^2 & -\zeta(\zeta-1) & 2(\zeta-1) & -(\zeta-1) \\ -\zeta(\zeta-1) & 3\zeta^2 & -6\zeta & \zeta+2 \\ \zeta-1 & -3\zeta & 2(\zeta+2) & -3 \\ -(\zeta-1) & \zeta+2 & -6 & 3 \end{bmatrix}, \tag{A3}$$

TABLE III. Comparison of axial and light-cone gauges in analytic regularization.

	Axial gauge	Light-cone gauge (principal-value prescription)
Evaluation of integrals	moderately easy	extremely easy
Preserves gauge invariance	yes	yes
Divergences at one loop	single poles	single and double poles
Infrared and ultraviolet singularities separable	yes [see (20)]	no
"New" operators in $\Pi_{\mu\nu}(p)$ at one loop contain infinite parts	no	yes
"New" operators in $\Gamma_{\lambda\mu\nu}(p, -p, 0)$ at one loop contain infinite parts	no	yes
Multiplicatively renormalizable	yes	no (see note added in proof)
Renormalization constants	$Z_1 = Z_3 = 1 + \frac{g^2 C_2}{16\pi^2} \cdot \frac{11}{3\epsilon}$	?
β function	$-\frac{g^2 C_2}{16\pi^2} \cdot \frac{11}{3}$	?

independently of the regularization method.

Similarly, for the vertex function written in (17), we compute the scalar quantities b_i defined by contracting

$$b_i = \sum_{\lambda,\mu,\nu} \Gamma_{\lambda\mu\nu}(p,-p,0) O^{(i)}_{\lambda\mu\nu}, \tag{A4}$$

where $O^{i}_{\lambda\mu\nu}$ are operators defined in Table I. This defines a linear relation between B_i and b_i which, upon inversion, yields

$$B_i = (U_\Gamma)_{ij}\, b_j. \tag{A5}$$

We find

$$U_\Gamma = \frac{1}{2(\zeta-1)} \times \begin{bmatrix} \zeta & 0 & -1 & 0 & -\zeta & -2 & \zeta & 2\zeta & -\zeta^2 & 1 \\ 0 & \zeta & 0 & -1 & -1 & -\zeta-1 & \zeta & 2\zeta & -\zeta^2 & 1 \\ -1 & 0 & 1 & 0 & 1 & 2 & -\zeta & -2 & \zeta & -1 \\ 0 & -1 & 0 & 1 & 1 & 2 & -1 & -\zeta-1 & \zeta & -1 \\ -\zeta & -2 & 1 & 2 & 3\zeta+2 & 10 & -\zeta-4 & -8\zeta-2 & \zeta(\zeta+4) & -5 \\ -1 & -\zeta-1 & 1 & 2 & 5 & 3\zeta+7 & -4\zeta-1 & -5\zeta-5 & \zeta(\zeta+4) & -5 \\ \zeta & 2\zeta & -\zeta & -2 & -\zeta-4 & -8\zeta-2 & \zeta(3\zeta+2) & 10\zeta & -5\zeta^2 & \zeta+4 \\ \zeta & 2\zeta & -1 & -\zeta-1 & -4\zeta-1 & -5\zeta-5 & 5\zeta & \zeta(3\zeta+7) & -5\zeta^2 & \zeta+4 \\ -\zeta^2 & -2\zeta^2 & \zeta & 2\zeta & \zeta(\zeta+4) & 2\zeta(\zeta+4) & -5\zeta^2 & -10\zeta^2 & 5\zeta^3 & -3\zeta-2 \\ 1 & 2 & -1 & -2 & -5 & -10 & \zeta+4 & 2\zeta+8 & -3\zeta-2 & 5 \end{bmatrix} \tag{A6}$$

again independently of the method of regularization.

APPENDIX B: REDUCTION TO PRIMAL INTEGRALS

In Tables IV and V we list the one-loop scalar functions A_i and B_i, defined, respectively, in (8) and (17) in terms of reduced primal integrals defined in (12). The A_i's and B_i's are given in units of $g^2C_2/16\pi^2$.

The tables were generated by evaluating the diagrams of Fig. 1, contracting as in (9) and (A4), and using the matrices (A3) and (A6). All contractions were simplified by reducing them to a sum of primal integrals using both the "shift" rule $q \rightarrow p-q$ in (1) where necessary, and partial fraction decomposition of integrals with multiple denominators. See Appendix C of Ref. 8 for more details.

Once the coefficients were reduced to a sum of primal integrals, the integrals themselves were reduced to a smaller set by using algebraic identities easily obtainable for cases with $N>0$. For example,

$$S(K,M,1,0) = (p\cdot n)^2 S(M,K,0,0) - 2p\cdot n S(M,K,0,1) + S(M,K,1,0). \tag{B1}$$

We emphasize that the legitimacy of this technique is founded on the fact that divergent integrals obey the usual rules of algebra, such as (B1). We also note that when regulating primal integrals, it is vital to preserve this property. This is true of both (1) and (20), but it is easy to invent limiting processes which do not preserve simple algebraic identities such as (B1).

Each of the primal integrals was then evaluated according to (2) by an algorithm described elsewhere.[13] Both limits (10) and (20) were investigated. To reduce the tables to manageable proportions, those integrals satisfying (11) are omitted here.

All calculations were performed with the algebraic manipulator SCHOONSCHIP,[22] except for the matrices (A3) and (A6) that were obtained using REDUCE.[23] The tables themselves were formatted, using an on-line editor and typewriter, directly from the computer output.

TABLE IV. Coefficients A_i of (8) in terms of primal integrals with tadpoles omitted.

$$\begin{aligned}
A_1 = {} & -(-1/y^3 + 2/y^2)*\widehat{S}(-1,0,-1,0) - (-2/y^3 + 6/y^2 - 8/y)*\widehat{S}(-1,0,-1,1) \\
& - (1/(2y^3))*\widehat{S}(-1,1,-1,0) - (1/y^3 - 2/y^2)*\widehat{S}(-1,1,-1,1) \\
& - (12 - 10/y)*\widehat{S}(-1,-1,0,0) - 8*\widehat{S}(-1,-1,1,0) - (1/(2y^3) - 4/y^2 + 4/y)*\widehat{S}(-1,-1,-1,0) \\
& - (-16 + 1/y^3 - 4/y^2 + 16/y)*\widehat{S}(-1,-1,-1,1), \\
A_2 = {} & +(3 - 1/y^3 + 1/y^2 - 1/y + 3y/(1-y))*\widehat{S}(-1,0,-1,0) \\
& + (-12 - 2/y^3 + 4/y^2 - 12y/(1-y))*\widehat{S}(-1,0,-1,1) \\
& + (\tfrac{3}{2} + 1/(2y^3) + 1/(2y^2) + 3/(2y) + 3y/(2(1-y)))*\widehat{S}(-1,1,-1,0) \\
& + (-3 + 1/y^3 - 1/y^2 - 3/y - 3y/(1-y))*\widehat{S}(-1,1,-1,1) \\
& + (10 - 10/y + 6y/(1-y))*\widehat{S}(-1,-1,0,0) + (8 + 24y/(1-y))*\widehat{S}(-1,-1,1,0) \\
& + (\tfrac{3}{2} + 1/(2y^3) - 7/(2y^2) + 3/(2y) + 3y/(2(1-y)))*\widehat{S}(-1,-1,-1,0) \\
& + (-9 + 1/y^3 - 3/y^2 + 11/y - 9y/(1-y))*\widehat{S}(-1,-1,-1,1), \\
A_4 = {} & -A_3 = +(3 - 1/y^2 - 1/y + 3y(1-y))*\widehat{S}(-1,0,-1,0) \\
& + (-12 - 2/y^2 + 8/y - 12y/(1-y))*\widehat{S}(-1,0,-1,1) \\
& + (\tfrac{3}{2} + 1/(2y^2) + 3/(2y) + 3y/(2(1-y)))*\widehat{S}(-1,1,-1,0) \\
& + (-3 + 1/y^2 - 3/y - 3y(1-y))*\widehat{S}(-1,1,-1,1) \\
& + (-2 + 6y/(1-y))*\widehat{S}(-1,-1,0,0) \\
& + (24y/(1-y))*\widehat{S}(-1,-1,1,0) \\
& + (\tfrac{3}{2} + 1/(2y^2) - 5/(2y) + 3y/(2(1-y)))*\widehat{S}(-1,-1,-1,0) \\
& + (7 + 1/y^2 - 5/y - 9y/(1-y))*\widehat{S}(-1,-1,-1,1).
\end{aligned}$$

TABLE V. Coefficients B_i of (17) in terms of primal integrals with tadpoles omitted.

$$\begin{aligned}
B_1 = {} & (\tfrac{3}{4} + 3/(4y^3) - 13/(4y^2) + 3/(4y) + 3y/(4(1-y)))*\widehat{S}(-2,0,-1,0) \\
& + (-\tfrac{9}{2} + 3/(2y^3) - 9/(2y^2) + 23/(2y) - 9y/(2(1-y)))*\widehat{S}(-2,0,-1,1) \\
& + (\tfrac{3}{4} - 3/(4y^3) + 3/(4y^2) + 3/(4y) + 3y/(4(1-y)))*\widehat{S}(-2,1,-1,0) \\
& + (-3 - 3/(2y^3) + 3/y^2 - 3/y - 3y/(1-y))*\widehat{S}(-2,1,-1,1) \\
& + (\tfrac{1}{4} + 1/(4y^3) + /(4y^2) + 1/(4y) + y/(4(1-y)))*\widehat{S}(-2,2,-1,0) \\
& + (-\tfrac{1}{2} + 1/(2y^3) - 1/(2y^2) - 1/(2y) - y/(2(1-y)))*\widehat{S}(-2,2,-1,1) \\
& + (\tfrac{1}{4} - 1/(4y^3) + 9(4y^2) - 15/(4y) + y/(4(1-y)))*\widehat{S}(-2,-1,-1,0) \\
& + (14 - 1/(2y^3) + 2/y^2 - 6/y - 2y/(1-y))*\widehat{S}(-2,-1,-1,1) \\
& + (-\tfrac{9}{4} - 13/(4y^3) + 23/(4y^2) - 9/(4y) - 9y/(4(1-y)))*\widehat{S}(-1,0,-1,0) \\
& + (9 - 13/(2y^3) + 13/y^2 - 7/y + 9y/(1-y))*\widehat{S}(-1,0,-1,1) \\
& + (-1 + 1/y^3 - 1/y^2 - 1/y - y/(1-y))*\widehat{S}(-1,1,-1,0) \\
& + (2 + 2/y^3 - 2/y^2 + 2/y + 2y/(1-y))*\widehat{S}(-1,1,-1,1) \\
& + (-6 + 10/y - 6y/(1-y))*\widehat{S}(-1,-2,0,0) \\
& + (-12 + 20y/(1-y))*\widehat{S}(-1,-2,0,1) + (-8 - 24y/(1-y))*\widehat{S}(-1,-2,1,0) \\
& + (16y/(1-y))*\widehat{S}(-1,-2,1,1) \\
& + (-\tfrac{1}{4} - 1/(4y^3) + 7/(4y^2) - 1/(4y) - y/(4(1-y)))*\widehat{S}(-1,-2,-1,0) \\
& + (2 - 1/(2y^3) + 2/y^2 - 10/y + 2y/(1-y))*\widehat{S}(-1,-2,-1,1) \\
& + (18 - 10/y - 6y/(1-y))*\widehat{S}(-1,-1,0,0) + (-8 - 24y/(1-y))*\widehat{S}(-1,-1,1,0) \\
& + (-\tfrac{3}{2} + 5/(2y^3) - 26/(2y^2) + 29/(2y) - 3y/(2(1-y)))*\widehat{S}(-1,-1,-1,0) \\
& + (-23 + 5/y^3 - 15/y^2 + 25/y + 9y/(1-y))*\widehat{S}(-1,-1,-1,1), \\
B_2 = {} & (\tfrac{3}{4} + 3/(4y^2) - 5/(4y) + 3y/(4(1-y)))*\widehat{S}(-2,0,-1,0) \\
& + (-\tfrac{1}{2} - 1/(2y) - 9y/(2(1-y)))*\widehat{S}(2,0,-1,1)
\end{aligned}$$

TABLE V. (*Continued.*)

$$+ (\tfrac{3}{4} - 3/(4y^2) + 3/(4y) + 3y/(4(1-y)))*\hat{S}(-2,1,-1,0)$$
$$+ (-3 - 3y/(1-y))*\hat{S}(-2,1,-1,1)$$
$$+ (\tfrac{1}{4} + 1/(4y^2) + 1/(4y) + y/(4(1-y)))*\hat{S}(-2,2,-1,0)$$
$$+ (-\tfrac{1}{2} - 1/(2y) - y/(2(1-y)))*\hat{S}(-2,2,-1,1)$$
$$+ (\tfrac{1}{4} - 1/(4y^2) + 1/(4y) + y/(4(1-y)))*\hat{S}(-2,-1,-1,0)$$
$$+ (-2 + 1/y - 2y/(1-y))*\hat{S}(-2,-1,-1,1)$$
$$+ (-\tfrac{9}{4} + 1/y^3 - 7/(4y^2) + 7/(4y) - 9y/(4(1-y)))*\hat{S}(-1,0,-1,0)$$
$$+ (9 + 2/y^3 - 4/y^2 + 9y/(1-y))*\hat{S}(-1,0,-1,1)$$
$$+ (-1 - 1/(2y^3) - 1/y - y/(1-y))*\hat{S}(-1,1,-1,0)$$
$$+ (2 - 1/y^3 + 1/y^2 + 2/y + 2y/(1-y))*\hat{S}(-1,1,-1,1)$$
$$+ (-2 - 6y/(1-y))*\hat{S}(-1,-2,0,0) + (4 + 20y/(1-y))*\hat{S}(-1,-2,0,1)$$
$$+ (-24y/(1-y))*\hat{S}(-1,-2,1,0) + (16y/(1-y))*\hat{S}(-1,-2,1,1)$$
$$+ (-\tfrac{1}{4} + 1/(4y^2) - 1/(4y) - y/(4(1-y)))*\hat{S}(-1,-2,-1,0)$$
$$+ (2 - 1/y + 2y/(1-y))*\hat{S}(-1,-2,-1,1)$$
$$+ (-10 + 10/y - 6y/(1-y))*\hat{S}(-1,-1,0,0)$$
$$+ (-8 - 24y/(1-y))*\hat{S}(-1,-1,1,0)$$
$$+ (-\tfrac{3}{2} - 1/(2y^3) + 7/(2y^2) - 3/(2y) - 3y/(2(1-y)))*\hat{S}(-1,-1,-1,0)$$
$$+ (9 - 1/y^3 + 3/y^2 - 11/y + 9y/(1-y))*\hat{S}(-1,-1,-1,1),$$

$$B_3 = (-\tfrac{3}{4} + 1/(4y^2) + 5/(4y) - 3y/(4(1-y)))*\hat{S}(-2,0,-1,0)$$
$$+ (-\tfrac{7}{2} - 1/(2y^2) + 1/(2y) + 9y/(2(1-y)))*\hat{S}(-2,0,-1,1)$$
$$+ (-\tfrac{3}{4} + 1/(4y^2) - 3/(4y) - 3y/(4(1-y)))*\hat{S}(-2,1,-1,0)$$
$$+ (3 + 1/y^2 - 1/y + 3y/(1-y))*\hat{S}(-2,1,-1,1)$$
$$+ (-\tfrac{1}{4} - 1/(4y^2) - 1/(4y) - y/(4(1-y)))*\hat{S}(-2,2,-1,0)$$
$$+ (\tfrac{1}{2} - 1/(2y^2) + 1/(2y) + y/(2(1-y))*\hat{S}(-2,2,-1,1)$$
$$+ (-\tfrac{1}{4} - 1/(4y^2) + 7/(4y) - y/(4(1-y))*\hat{S}(-2,-1,-1,0)$$
$$+ (-6 - 2/y + 2y/(2-y))*\hat{S}(-2,-1,-1,1) + (1/y^3 - 2/y^2)*\hat{S}(-1,0,-2,1)$$
$$+ (\tfrac{9}{4} + 3/y^3 - 23/(4y^2) + 25/(4y) + 9y/(4(1-y)))*\hat{S}(-1,0,-1,0)$$
$$+ (-9 + 6/y^3 - 11/y^2 + 3/y - 9y/(1-y))*\hat{S}(-1,0,-1,1)$$
$$+ (-1/(2y^3))*\hat{S}(-1,1,-2,1)$$
$$+ (1 - 3/(2y^3) + 2/y^2 + 1/y + y/(1-y))*\hat{S}(-1,1,-1,0)$$
$$+ (-2 - 3/y^3 + 4/y^2 - 2/y - 2y/(1-y))*\hat{S}(-1,1,-1,1)$$
$$+ (-10 + 6y/(1-y))*\hat{S}(-1,-2,0,0) + (20 - 20y/(1-y))*\hat{S}(-1,-2,0,1)$$
$$+ (24y/(1-y))*\hat{S}(-1,-2,1,0) + (-16y/(1-y))*\hat{S}(-1,-2,1,1)$$
$$+ (\tfrac{1}{4} + 1/(4y^2) - 7/(4y) + y/(4(1-y)))*\hat{S}(-1,-2,-1,0)$$
$$+ (6 + 2/y - 2y/(1-y))*\hat{S}(-1,-2,-1,1) + (-2 + 6y/(1-y))*\hat{S}(-1,-1,0,0)$$
$$+ (24y(1-y))*\hat{S} - 1,-1,1,0) + (-1/(2y^3) + 4/y^2 - 4/y)*\hat{S}(-1,-1,-2,1)$$
$$+ (\tfrac{19}{2} - 3/(2y^3) + 7/(2y^2) - 21/(2y) + 3y/(2(1-y)))*\hat{S}(-1,-1,-1,0)$$
$$+ (-9 - 3/y^3 + 7/y^2 + 7/y - 9y/(1-y))*\hat{S}(-1,-1,-1,1),$$

$$B_5 = (\tfrac{15}{4} - 3/(4y^3) + 1/(4y^2) + 17/(4y) - 15y/(4(1-y)))*\hat{S}(-2,0,-1,0)$$
$$+ (\tfrac{29}{2} - 3/(2y^3) + 15/(2y^2) - 35/(2y) + 45y/(2(1-y)))*\hat{S}(-2,0,-1,1)$$
$$+ (-\tfrac{15}{4} + 3/(4y^3) + 9/(4y^2) - 15/(4y) - 15y/(4(1-y)))*\hat{S}(-2,1,-1,0)$$
$$+ (15 + 3/(2y^3) - 6/y^2 + 9/y + 15y/(1-y))*\hat{S}(2,1,-1,1)$$
$$+ (-\tfrac{5}{4} - 1/(4y^3) - 5/(4y^2) - 5/(4y) - 5y/(4(1-y)))*\hat{S}(-2,2,-1,0)$$
$$+ (\tfrac{5}{2} - 1/(2y^3) + 3/(2y^2) + 5/(2y) + 5y/(2(1-y)))*\hat{S}(-2,2,-1,1)$$

TABLE V. (*Continued.*)

$$
\begin{aligned}
&+\left(-\tfrac{5}{4}+1/(4y^3)-5/(4y^2)+11/(4y)-5y/(4(1-y))\right)*\hat{S}(-2,-1,-1,0)\\
&+\left(-6+1/(2y^3)-3/y^2+4/y+10y/(1-y)\right)*\hat{S}(-2,-1,-1,1)\\
&+\left(\tfrac{45}{4}+13/(4y^3)-15/(4y^2)-19/(4y)+45y/(4(1-y))\right)*\hat{S}(-1,0,-1,0)\\
&+\left(-45+13/(2y^3)-22/y^2+29/y-45y/(1-y)\right)*\hat{S}(-1,0,-1,1)\\
&+\left(5-1/y^3+1/y^2+5/y+5y/(1-y)\right)*\hat{S}(-1,1,-1,0)\\
&+\left(-10-2/y^3+4/y^2-10/y-10y/(1-y)\right)*\hat{S}(-1,1,-1,1)\\
&+\left(22-2/y+30y/(1-y)\right)*\hat{S}(-1,-2,0,0)\\
&+\left(-28-100y/(1-y)\right)*\hat{S}(-1,-2,0,1)+\left(24+120y/(1-y)\right)*\hat{S}(-1,-2,1,0)\\
&+\left(-80y/(1-y)\right)*\hat{S}(-1,-2,1,1)\\
&+\left(\tfrac{5}{4}+1/(4y^3)-7/(4y^2)+5/(4y)+5y/(4(1-y))\right)*\hat{S}(-1,-2,-1,0)\\
&+\left(-10+1/(2y^3)-3/y^2+8/y-10y/(1-y)\right)*\hat{S}(-1,-2,-1,1)\\
&+\left(-2+2/y+30y/(1-y)\right)*\hat{S}(-1,-1,0,0)+\left(24+120y/(1-y)\right)*\hat{S}(-1,-1,1,0)\\
&+\left(\tfrac{15}{2}-5/(2y^3)+23/(2y^2)-33/(2y)+15y/(2(1-y))\right)*\hat{S}(-1,-1,-1,0)\\
&+\left(19-5/y^3+23/y^2-37/y-45y/(1-y)\right)*\hat{S}(-1,-1,-1,1),
\end{aligned}
$$

$$
\begin{aligned}
B_6=&\left(-\tfrac{15}{4}-5/(4y^2)+17/(4y)-15y/(4(1-y))\right)*\hat{S}(-2,0,-1,0)\\
&+\left(\tfrac{13}{2}-1/(2y^2)-3/(2y)+45y/(2(1-y))\right)*\hat{S}(-2,0,-1,1)\\
&+\left(-\tfrac{15}{4}+7/(4y^2)-7/(4y)-15y/(4(1-y))\right)*\hat{S}(-2,1,-1,0)\\
&+\left(15+1/y^2-1/y+15y/(1-y)\right)*\hat{S}(-2,1,-1,1)\\
&+\left(-\tfrac{5}{4}-3/(4y^2)-5/(4y)-5y/(4(1-y))\right)*\hat{S}(-2,2,-1,0)\\
&+\left(\tfrac{5}{2}-1/(2y^2)+5/(2y)+5y/(2(1-y))\right)*\hat{S}(-2,2,-1,1)\\
&+\left(-\tfrac{5}{4}+1/(4y^2)+3/(4y)-5y/(4(1-y))\right)*\hat{S}(-2,-1,-1,0)\\
&+\left(2-2/y+10y/(1-y)\right)*\hat{S}(-2,-1,-1,1)+\left(1/y^3-4/y\right)*\hat{S}(-1,0,-2,1)\\
&+\left(\tfrac{45}{4}+2/y^3-33/(4y^2)+61/(4y)+45y/(4(1-y))\right)*\hat{S}(-1,0,-1,0)\\
&+\left(-45+4/y^3-9/y^2-1/y-45y/(1-y)\right)*\hat{S}(-1,0,-1,1)\\
&+\left(-1/(2y^3)-1/y^2\right)*\hat{S}(-1,1,-2,1)\\
&+\left(5-1/y^3+4/y^2+5/y+5y/(1-y)\right)*\hat{S}(-1,1,-1,0)\\
&+\left(-10-2/y^3+4/y^2-10/y-10y/(1-y)\right)*\hat{S}(-1,1,-1,1)\\
&+\left(2+30y/(1-y)\right)*\hat{S}(-1,-2,0,0)+\left(-4-100y/(1-y)\right)*\hat{S}(-1,-2,0,1)\\
&+\left(120y/(1-y)\right)*\hat{S}(-1,-2,1,0)+\left(-80y/(1-y)\right)*\hat{S}(-1,-2,1,1)\\
&+\left(\tfrac{5}{4}-1/(4y^2)-3/(4y)+5y/(4(1-y))\right)*\hat{S}(-1,-2,-1,0)\\
&+\left(-2+2/y-10y/(1-y)\right)*\hat{S}(-1,-2,-1,1)+\left(26-2/y+30y/(1-y)\right)*\hat{S}(-1,-1,0,0)\\
&+\left(24+120y/(1-y)\right)*\hat{S}(-1,-1,1,0)+\left(-1/(2y^3)+3/y^2-4/y\right)*\hat{S}(-1,-1,-2,1)\\
&+\left(\tfrac{31}{2}-1/y^3+5/(2y^2)-13/(2y)+15y/(2(1-y))\right)*\hat{S}(-1,-1,-1,0)\\
&+\left(-29-2/y^3+5/y^2-1/y-45y/(1-y)\right)*\hat{S}(-1,-1,-1,1),
\end{aligned}
$$

$$
\begin{aligned}
B_7=&\left(\tfrac{15}{4}-1/(4y^3)+1/(4y^2)-9/(4y)+15y/(4(1-y))\right)*\hat{S}(-2,0,-1,0)\\
&+\left(-\tfrac{29}{2}+1/(2y^3)-1/(2y^2)+11/(2y)-45y/(2(1-y))\right)*\hat{S}(-2,0,-1,1)\\
&+\left(\tfrac{15}{4}-1/(4y^3)-3/(4y^2)+7/(4y)+15y/(4(1-y))\right)*\hat{S}(-2,1,-1,0)\\
&+\left(-15-1/y^3+1/y^2-3/y-15y/(1-y)\right)*\hat{S}(-2,1,-1,1)\\
&+\left(\tfrac{5}{4}+1/(4y^3)+3/(4y^2)+5/(4y)+5y/(4(1-y))\right)*\hat{S}(-2,2,-1,0)\\
&+\left(-\tfrac{5}{2}+1/(2y^3)-1/(2y^2)-5/(2y)-5y/(2(1-y))\right)*\hat{S}(-2,2,-1,1)\\
&+\left(\tfrac{5}{4}+1/(4y^3-9/(4y^2)+5/(4y)+5y/(4(1-y))\right)*\hat{S}(-2,-1,-1,0)\\
&+\left(-10+2/y^2+6/y-10y/(1-y)\right)*\hat{S}(-2,-1,-1,1)\\
&+\left(-1/y^4+2/y^3-2/y^2+4/y\right)*\hat{S}(-1,0,-2,1)
\end{aligned}
$$

TABLE V. (*Continued.*)

$$+(-\tfrac{45}{4}-3/y^4+31/(4y^3)-15/(4y^2)-61/(4y)-45y/(4(1-y)))*\hat{S}(-1,0,-1,0)$$

$$+(45-6/y^4+15/y^3-15/y^2+33/y+45y/(1-y))*\hat{S}(-1,0,-1,1)$$

$$+(1/(2y^4)+1/y^2)*\hat{S}(-1,1,-2,1)$$

$$+(-5+3/(2y^4)-3/y^3-4/y^2-5/y-5y/(1-y))*\hat{S}(-1,1,-1,0)$$

$$+(10+3/y^4-6/y^3+6/y^2+10/y+10y/(1-y))*\hat{S}(-1,1,-1,1)$$

$$+(-18+10/y-30y/(1-y))*\hat{S}(-1,-2,0,0)+(44-20/y+100y/(1-y))*\hat{S}(-1,-2,0,1)$$

$$+(-24-120y/(1-y))*\hat{S}(-1,-2,1,0)+(16+80y/(1-y))*\hat{S}(-1,-2,1,1)$$

$$+(-\tfrac{5}{4}-1/(4y^3)+9/(4y^2)-5/(4y)-5y/(4(1-y)))*\hat{S}(-1,-2,-1,0)$$

$$+(10-2/y^2-6/y+10y/(1-y))*\hat{S}(-1,-2,-1,1)$$

$$+(-42+6/y-30y/(1-y))*\hat{S}(-1,-1,0,0)+(-72-120y/(1-y))*\hat{S}(-1,-1,1,0)$$

$$+(1/(2y^4)-4/y^3+5/y^2)*\hat{S}(-1,-1,-2,1)$$

$$+(-\tfrac{15}{2}+3/(2y^4)-9/(2y^3+29/(2y^2)-35/(2y)-15y/(2(1-y)))*\hat{S}(-1,-1,-1,0)$$

$$+(45+3/y^4-9/y^3+1/y^2+5/y+45y/(1-y))*\hat{S}(-1,-1,-1,1),$$

$$B_8=(\tfrac{15}{4}+3/(2y^3)-9/(4y^2)-9/(4y)+15y/(4(1-y)))*\hat{S}(-2,0,-1,0)$$

$$+(-\tfrac{37}{2}+3/(2y^3)-8/y^2+35/(2y)-45y/(2(1-y)))*\hat{S}(-2,0,-1,1)$$

$$+(\tfrac{15}{4}-3/(4y^3)-3/(4y^2)+15/(4y)+15y/(4(1-y)))*\hat{S}(-2,1,-1,0)$$

$$+(-15-3/(2y^3)+6/y^2-12/y-15y/(1-y))*\hat{S}(-2,1,-1,1)$$

$$+(\tfrac{5}{4}+1/(2y^3)+5/(4y^2)+5/(4y)+5y/(4(1-y)))*\hat{S}(-2,2,-1,0)$$

$$+(-\tfrac{5}{2}+1/(2y^3)-2/y^2-5/(2y)-5y/(2(1-y)))*\hat{S}(-2,2,-1,1)$$

$$+(\tfrac{5}{4}-1/(2y^3)+7/(4y^2)-11/(4y)+5y/(4(1-y)))*\hat{S}(-2,-1,-1,0)$$

$$+(6-1/(2y^3)+4/y^2-7/y-10y/(1-y))*\hat{S}(-2,-1,-1,1)$$

$$+(-\tfrac{45}{4}-2/y^3+21/(4y^2)-13/(4y)-45y/(4(1-y)))*\hat{S}(-1,0,-1,0)$$

$$+(45-5/(2y^3)+2/y^2+8/y+45y/(1-y))*\hat{S}(-1,0,-1,1)$$

$$+(-5+1/(2y^3)-3/y^2-5/y-5y/(1-y))*\hat{S}(-1,1,-1,0)$$

$$+(10+3/y^2+10/y+10y/(1-y))*\hat{S}(-1,1,-1,1)$$

$$+(-26-30y/(1-y))*\hat{S}(-1,-2,0,0)+(44+4/y+100y/(1-y))*\hat{S}(-1,-2,0,1)$$

$$+(-48-120y/(1-y))*\hat{S}(-1,-2,1,0)+(16+80y/(1-y))*\hat{S}(-1,-2,1,1)$$

$$+(-\tfrac{5}{4}+5/(4y^2)-5/(4y)-5y/(4(1-y))*\hat{S}(-1,-2,-1,0)$$

$$+(10-1/(2y^3)+2/y^2-5/y+10y/(1-y))*\hat{S}(-1,-2,-1,1)$$

$$+(-14+2/y-30y/(1-y))*\hat{S}(-1,-1,0,0)+(-72-120y/(1-y))*\hat{S}(-1,-1,1,0)$$

$$+(-\tfrac{15}{2}+3/(2y^3)-11/(2y^2)+17/(2y)-15y/(2(1-y)))*\hat{S}(-1,-1,-1,0)$$

$$+(13+3/y^3+11/y^2+13/y+45y/(1-y))*\hat{S}(-1,-1,-1,1),$$

$$B_9=(-\tfrac{15}{4}-3/(4y^4)+7/(4y^3)+1/(4y^2)+1/(4y)-15y/(4(1-y)))*\hat{S}(-2,0,-1,0)$$

$$+(\tfrac{45}{2}-3/(2y^4)+9/(2y^3)-15/(2y^2)-3/(2y)+45y/(2(1-y)))*\hat{S}(-2,0,-1,1)$$

$$+(-\tfrac{15}{4}+3/(4y^4)+3/(4y^3)-3/(4y^2)-15/(4y)-15y/(4(1-y)))*\hat{S}(-2,1,-1,0)$$

$$+(15+3/(2y^4)-3/y^3+15/y+15y/(1-y))*\hat{S}(-2,1,-1,1)$$

$$+(-\tfrac{5}{4}-1/(4y^4)-3/(4y^3)-5/(4y^2)-5/(4y)-5y/(4/(1-y)))*\hat{S}(-2,2,-1,0)$$

$$+(\tfrac{5}{2}-1/(2y^4)+1/(2y^3)+5/(2y^2)+5/(2y)+5y/(2(1-y)))*\hat{S}(-2,2,-1,1)$$

$$+(-\tfrac{5}{4}+1/(4y^4)-7/(4y^3)+15/(4y^2)-5/(4y)-5y/(4(1-y)))*\hat{S}(-2,-1,-1,0)$$

$$+(10+1/(2y^4)-2/y^3+3/y^2-10/y+10y/(1-y))*\hat{S}(-2,-1,-1,1)$$

$$+(\tfrac{45}{4}+5/(4y^4)-9/(4y^3)-11/(4y^2)+45/(4y)+45y/(4(1-y)))*\hat{S}(-1,0,-1,0)$$

$$+(-45+5/(2y^4)-5/y^3+10/y^2-45/y-45y/(1-y))*\hat{S}(-1,0,-1,1)$$

TABLE V. (*Continued.*)

$$
\begin{aligned}
&+ (5 + 1/y^3 + 5/y^2 + 5/y + 5y/(1-y)) \ast \widehat{S}(-1,1,-1,0)\\
&+ (-10 - 10/y^2 - 10/y - 10y/(1-y)) \ast \widehat{S}(-1,1,-1,1)\\
&+ (30 - 10/y^2 + 18/y + 30y/(1-y)) \ast \widehat{S}(-1,-2,0,0)\\
&+ (-92 + 4/y - 100y/(1-y)) \ast \widehat{S}(-1,-2,0,1) + (88 + 8/y + 120y/(1-y))(-1,-2,1,0)\\
&+ (48 - 80y/(1-y)) \ast \widehat{S}(-1,-2,1,1)\\
&+ (\tfrac{5}{4} + 1/(4y^4) - 9/(4y^3) + 5/(4y^2) + 5/(4y) + 5y/(4(1-y))) \ast \widehat{S}(-1,-2,-1,0)\\
&+ (-10 + 1/(2y^4) - 2/y^3 + 11/y^2 - 10/y - 10y/(1-y)) \ast \widehat{S}(-1,-2,-1,1)\\
&+ (30 - 10/y^2 + 10/y + 30y/(1-y)) \ast \widehat{S}(-1,-1,0,0)\\
&+ (120 + 24/y + 120y/(1-y)) \ast \widehat{S}(-1,-1,1,0)\\
&+ (\tfrac{15}{2} - 3/(2y^4) + 13/(2y^3) - 25/(2y^2 + 15/(2y) + 15y/(2(1-y))) \ast \widehat{S}(-1,-1,-1,0)\\
&+ (-45 - 3/y^4 + 9/y^3 - 1/y^2 - 5/y - 45y/(1-y)) \ast \widehat{S}(-1,-1,-1,1),\\
B_4 = {}& -B_2 + (1/y^3 - 2/y^2) \ast \widehat{S}(-1,0,-1,0)\\
&+ (2/y^3 - 6/y^2 + 8/y) \ast \widehat{S}(-1,0,-1,1) + (-1/(2y^3)) \ast \widehat{S}(-1,1,-1,0)\\
&+ (-1/y^3 + 2/y^2) \ast \widehat{S}(-1,1,-1,1) + (-12 + 10/y) \ast \widehat{S}(-1,-1,0,0)\\
&- 8 \ast \widehat{S}(-1,-1,1,0) + (-1/(2y^3) + 4/y^2 - 4/y) \ast \widehat{S}(-1,-1,-1,0)\\
&+ (16 - 1/y^3 + 4/y^2 - 16/y) \ast \widehat{S}(-1,-1,-1,1),\\
B_{10} = {}& -B_6 + (-1/y^3 + 4/y) \ast \widehat{S}(-1,0,-1,0) + (-2/y^3 + 10/y^2 - 20/y) \ast \widehat{S}(-1,0,-1,1)\\
&+ (1/(2y^3) + 1/y^2) \ast \widehat{S}(-1,1,-1,0) + (1/y^3 - 4/y^2) \ast \widehat{S}(-1,1,-1,1)\\
&+ (8 - 2/y) \ast \widehat{S}(-1,-1,0,0) + 24 \ast \widehat{S}(-1,-1,1,0)\\
&+ (1/(2y^3) - 3/y^2 + 4/y) \ast \widehat{S}(-1,-1,-1,0)\\
&+ (-16 + 1/y^3 - 6/y^2 + 12/y) \ast \widehat{S}(-1,-1,-1,1).
\end{aligned}
$$

[1] H.-C. Lee and M. S. Milgram, Ann. Phys. **157**, 408 (1984); Atomic Energy of Canada Limited report AECL-8261 (1984); H.-C. Lee and M. S. Milgram, Phys. Lett. B **133**, 320 (1983).

[2] M. S. Milgram and H.-C. Lee, CRNL-TP-83-XII-13 (unpublished).

[3] W. Kummer, Acta Phys. Austriaca **41**, 315 (1975).

[4] W. Kummer, Acta Phys. Austriaca **15**, 149 (1961); Y. P. Yao, J. Math. Phys. **5**, 1319 (1964); R. Delbourgo, A. Salam, and J. Strathdee, Nuovo Cimento **23**, 237 (1974); W. Konetschny and W. Kummer, Nucl. Phys. B **100**, 106 A (1975); B. Humpert and W. L. van Neerven, Phys. Lett. B **101**, 101 (1981); D. M. Capper and G. Leibbrandt, Phys. Rev. D **25**, 1002, 1009 (1982).

[5] J. M. Cornwall, Phys. Rev. D **10**, 500 (1974); M. Kaku, Nucl. Phys. B **91**, 99 (1975); J. Scherk and J. H. Schwartz, Gen. Relativ. Gravit. **6**, 537 (1975); D. J. Pritchard and W. J. Stirling, Nucl. Phys. B **165**, 237 (1980); G. Curci, W. Furmanski, and R. Petronzio, Nucl. Phys. B **175**, 27 (1980); J. C. Collins and D. E. Soper, Nucl. Phys. B **194**, 445 (1982); S. Mandelstam, Nucl. Phys. B **213**, 149 (1983); G. Leibbrandt, Phys. Rev. D **29**, 1699 (1984); D. M. Capper, J. J. Dulwich, and M. J. Litvak, Nucl. Phys. B **241**, 463 (1984).

[6] C. G. Bollini, J. J. Giambiagi, and A. Gonzales Dominguez, Nuovo Cimento **31**, 550 (1964); E. R. Speer, J. Math. Phys. **9**, 1404 (1968).

[7] J. C. Taylor, Nucl. Phys. B **33**, 436; A. Slavnov, Theor. Math. Phys. **10**, 99 (1972) (English translation).

[8] H.-C. Lee and M. S. Milgram, preprint CRNL-TP-83-XII-14.

[9] G. 't Hooft and M. Veltman, Nucl. Phys. B **44**, 189 (1972); C. J. Bollini and J. J. Giambiagi, Nuovo Cimento B **12**, 20 (1972); J. F. Ashmore, Nuovo Cimento Lett. **4**, 289 (1972); G. Leibbrandt, Rev. Mod. Phys. **47**, 849 (1975).

[10] A tadpole diagram is a Feynman diagram in which all external lines emanate from one vertex and all internal lines are propagators of massless particles. Conservation of momentum dictates that the total momentum carried by the external lines must be zero. The conventional argument for assigning a zero value to tadpole diagrams is that one could not find a dimensionful scale to express such a diagram were its value finite. It is clear that this argument has at least one loophole: a dimensionless diagram certainly could have a finite (constant) value. By analogy, primal integrals with $K = 0$ may be called tadpole integrals. Since the new analytic method reveals that such integrals are but a special subclass of a much larger class of vanishing (in a certain limit) integrals, we call the whole class of integrals tadpolelike integrals or "tadpoles." They are identified by satisfying either or both of the conditions (11).

[11] See, e.g., E. S. Abers and B. W. Lee, Phys. Rep. C **9**, 1 (1973).

[12] See, e.g., Y. Luke, *The Special Functions and Their Approximations* (Academic, New York, 1969), Chap. 5.

[13] M. S. Milgram and H. C. Lee, preprint CRNL-TP-84-I-31 (to appear in J. Comput. Phys.).

[14] G. Leibbrandt, Ref. 9.

[15] On the other hand, having different sets of integrals on the two sides of (4) does not necessarily imply the two sides are unequal. They would be unequal only if it could be proven that among the different integrals those on one side were independent of those on the other. We have not made an attempt to rigorously assure that the integrals used to express B_i are from an independent set, but have simply reduced them to a simple form using straightforward algebraic identities.

[16] There are actually two types of infrared singularities in the integral (1): those occurring at $q = 0$ and those occurring at $p = q$. The singularities associated with each one of these must not be distinguished.

[17] R. Gastmans and R. Meuldermans, Nucl. Phys. B **63**, 277 (1973); W. J. Marciano, Phys. Rev. D **12**, 3861 (1975).

[18] S. Narison, Phys. Rep. **84**, 263 (1982).

[19] See, e.g., D. M. Capper *et al.*, Ref. 5, and J. M. Cornwall, Ref. 5.

[20] S. Mandelstam, Ref. 5, as discussed in Ref. 19.

[21] J. M. Cornwall, Phys. Rev. D **26**, 1453 (1982).

[22] H. Strubbe, Comput. Phys. Commun. **18**, 1 (1979).

[23] A. C. Hearn, *REDUCE2 User's Manual* (Univ. Utah, Salt Lake City, Utah, 1973).

Convergent expansions for glueball masses in strongly coupled 3 + 1 lattice gauge theories

Michael O'Carroll[a)]
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138
Wilson Dantas Barbosa
Departamento de Matemática do ICEx, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

(Received 18 September 1984; accepted for publication 7 December 1984)

We analyze the mass spectrum of a strongly coupled ($\beta = 2/g^2$ small) Wilson action lattice gauge theory in 3 + 1 dimensions. In the subspace generated by the time zero plaquette functions and their complex conjugates we show that there is at least one and not more than four masses. Each mass admits a representation of the form $m(\beta) = -\ln\beta + r(\beta)$, where $r(\beta)$ is a gauge group representation-dependent function analytic in β or $\beta^{1/2}$ at $\beta = 0$. For a gauge group representation with real character there is at least one and not more than two masses of the above form and $r(\beta)$ is analytic at $\beta = 0$. Furthermore c_n, the nth $\beta = 0$ Taylor series coefficient of $r(\beta)$, can be obtained by a finite algorithm.

I. INTRODUCTION AND MAIN RESULTS

Recently much theoretical and numerical work has been done to determine the low lying glueball mass spectrum of lattice gauge theories (see Refs. 1–4). Here we obtain results on the mass spectrum of strongly coupled ($\beta = 2/g^2$ small) Wilson action gauge theories in 3 + 1 dimensions. We restrict our attention to the spectrum in the subspace generated by the time-zero plaquette functions and work in the Euclidean formulation (see Refs. 5 and 6). We let χ denote the character of the irreducible unitary representation of the gauge group. We state our main result.

Theorem A: The mass spectrum has at least one and not more than four isolated points. Each mass has a representation

$$m(\beta) = -4\ln\beta + r(\beta),$$

where $r(\beta)$ is a gauge group representation-dependent function analytic in β or $\beta^{1/2}$ at $\beta = 0$.

In order to simplify the exposition we restrict our discussion in Secs. II–IV to χ real for which the above theorem takes the form.

Theorem B: For χ real the mass spectrum has at least one and not more than two isolated points. Each mass has a representation

$$m(\beta) = -4\ln\beta + r(\beta),$$

where $r(\beta)$ is a gauge group representation-dependent function analytic in β at $\beta = 0$. The nth $\beta = 0$ Taylor series coefficient of $r(\beta,)$ c_n, can be computed by a finite algorithm. Here, c_n depends on a finite number of $\beta = 0$ Taylor series coefficients of finite lattice correlation functions at a finite number of points.

Both theorems follow from the solution of an implicit equation for the mass or masses which depends on β and momentum space analyticity properties of the Fourier transform of a matrix-valued correlation function (abbreviated cf hereafter) and its matrix-valued convolution inverse. The crucial analyticity properties follow from the space, imaginary-time decay of the cf and a faster imaginary-time decay of its convolution inverse, as in Ref. 1.

The main difference between the 2 + 1 mass spectrum problem treated in Refs. 1 and 2 and the 3 + 1 problem is that in the 3 + 1 case we are dealing with an asymptotically ($\beta \downarrow 0$) degenerate level (rather than a nondegenerate level) so that the spectral analysis is more complicated. We develop techniques of spectral analysis for asymptotically degenerate levels which can be easily generalized to apply to the analysis of the spectrum of other models, i.e., multicomponent spin and gauge-matter models. Generally, the mass spectrum is contained in the set of momentum values (with positive imaginary-time component and zero space component) for which the determinant of the Fourier transform of the convolution inverse matrix of the cf matrix is zero.

We describe the organization of the paper. In Secs. II–IV χ is real. Symmetry, analyticity, and decay properties of the cf and its convolution inverse are given in Sec. II. In Sec. III we develop the spectral analysis needed to relate the momentum space analyticity of the cf to the mass spectrum. In Sec. IV we introduce implicit equations for the masses and prove Theorem B. In Sec. V we treat the case $\chi = \chi_r + i\chi_i$ complex and prove Theorem A. The proofs of lemmas and theorems which do not require major modifications of analogous ones in the 2 + 1 dimension case of Refs. 1 and 2 are not given. Section VI is devoted to some concluding remarks. We take $|\beta|$ to be small throughout.

II. SYMMETRY, DECAY, AND ANALYTICITY OF CORRELATION FUNCTIONS

In this section we introduce a matrix-valued cf and obtain its symmetry, analyticity, and decay properties. The crucial faster imaginary-time decay of the convolution inverse of the cf is also obtained (see Lemma II.3b) as well as its symmetry and analyticity properties.

We let $G(x,\beta)$ denote the 3×3 matrix-valued plaquette–plaquette cf with matrix elements

---

[a)] On leave from Departamento de Fisica do ICEx, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.

 0022-2488/85/071805-05$02.50

$$G_{\alpha\gamma}(x,\beta) = \lim_{\Lambda \uparrow Z^4} G_{\Lambda\alpha\gamma}(y;z,\beta), \quad x = y - z, \quad x,y,z \in \Lambda \subset Z^4,$$

$\alpha,\gamma = 1,2,3$ where we denote points of Z^4 by $x = (x_0, x_1, x_2, x_3) = (x_0,\mathbf{x})$. Here

$$G_{\Lambda\alpha\gamma}(y;z,\beta) = \langle \chi(g_{p_y})\chi(g_{p_z})\rangle_\Lambda - \langle \chi(g_{p_y})\rangle_\Lambda \langle \chi(g_{p_z})\rangle_\Lambda .$$

$P_y\,(P_z)$ is the plaquette located at $y(z)$ perpendicular to the x_0 direction and to the $x_\alpha\,(x_\gamma)$ direction, $\langle\cdot\rangle_\Lambda$ is the Gibbs ensemble average with Boltzmann factor $\exp[\beta\,\Sigma_p\chi(g_p)]$, and χ is a real character of an r-dimensional irreducible unitary representation of the compact gauge group. Existence, β analyticity, and translation invariance of the $\Lambda \uparrow Z^4$ limit follow from the polymer expansion of Ref. 6 for small $|\beta|$.

The symmetry properties of $G_{\alpha\gamma}(x)$ are given below. Let I,M denote the 3×3 matrices with matrix elements $I_{ij} = \delta_{ij}$ and $M_{ij} = (1-\delta_{ij})$ and set $G_{\alpha\gamma}(x_0) = \Sigma_{\mathbf{x}}\, G_{\alpha\gamma}(x_0,\mathbf{x})$. We have the following lemma.

Lemma II.1: $G(x_0) = G_{11}(x_0)I + G_{12}(x_0)M$.

Proof: To show $G_{11}(x_0) = G_{22}(x_0)$ make a rotation of $-\pi/2$ about the x_3 axis in $G_{\Lambda\,11}(0;x_0,\mathbf{x})$, sum over $\mathbf{x}$, and take the $\Lambda \uparrow Z^4$ limit. To show $G_{13}(x_0) = G_{23}(x_0)$ make a rotation of $-\pi/2$ about x_3 in $G_{\Lambda\,13}(0;x_0,\mathbf{x})$, sum over $\mathbf{x}$, and take the $\Lambda \uparrow Z^4$ limit. Similar considerations apply to the other components.

The decay of $G(x)$ and the existence, β analyticity, and decay of the convolution inverse $\Gamma(x,\beta) \equiv \Gamma(y;z,\beta)$, $x = y - z$, where

$$\sum_{z,\rho} G_{\alpha\rho}(x;z,\beta)\Gamma_{\rho\gamma}(z;y,\beta) = \delta_{xy}\delta_{\alpha\gamma}$$

are given in Lemmas II.2 and II.3 below.

Lemma II.2:

(a) $|G_{\alpha\gamma}(x,\beta)| \leqslant c_1|c\beta|^{4|x_0| + |\mathbf{x}|}$,

(b) $G_{\alpha\alpha}(x = (x_0, \mathbf{0}),\beta) > c_3|c_4\beta|^{4|x_0|}, \quad \beta > 0$.

Lemma II.3: Let P have matrix elements $P_{\alpha\gamma}(x;y) = G_{\alpha\alpha}(x;x)\delta_{xy}\delta_{\alpha\gamma}$. Then the following hold.

(a) $\Gamma = G^{-1} = \sum_{n=0}^{\infty} (-1)^n\,[P^{-1}(G-P)]^n P^{-1}$;

the series is convergent in norm and Γ is analytic.

(b) $|\Gamma_{\alpha\gamma}(x,\beta)| \leqslant c_2|c'\beta|^{5|x_0| + |\mathbf{x}|}$,

$x \neq (x_0 = \pm 1, 0), \; \alpha \neq \gamma$;

for $x = (x_0 = \pm 1, 0)$ and $\alpha = \gamma$ replace the 5 by 4.

Let $\sim$ denote the Fourier transform, i.e., $\tilde G(p) = \Sigma_x e^{-ipx}G(x)$, $p = (p_0, \mathbf{p})$, $px = \Sigma_{i=0}^3 p_i x_i$. We set $\tilde G(p_0) = \tilde G(p_0, \mathbf{p} = 0)$ and $\tilde\Gamma(p_0) = \tilde\Gamma(p_0, \mathbf{p} = 0)$. The symmetry, decay, and analyticity properties of Lemmas II.1–II.3 translate into the p space properties of $\tilde G(p_0)$ and $\tilde\Gamma(p_0)$ given in Lemmas II.4 and II.5 below.

Lemma II.4: $\tilde G(p_0,\beta)$ is jointly analytic for $|\beta|$ small, $|\mathrm{Im}\,p_0| < -4\ln|c\beta|$, $|c\beta| < 1$, $|\mathrm{Re}\,p_0| < \pi$, and

(a) $\tilde G(p_0) = \tilde G_{11}(p_0)I + \tilde G_{12}(p_0)M$,

(b) $\det \tilde G(p_0) = (\tilde G_{11}(p_0) + 2\tilde G_{12}(p_0))(\tilde G_{11}(p_0) - \tilde G_{12}(p_0))^2$.

Lemma II.5: $\tilde\Gamma(p_0)$ is jointly analytic for $|\beta|$ small, $|\mathrm{Im}\,p_0| < -5\ln|c'\beta|$, $|c'\beta| < 1$, $|\mathrm{Re}\,p_0| < \pi$, and

(a) $\tilde\Gamma(p_0) \equiv \tilde\Gamma_{11}(p_0)I + \tilde\Gamma_{12}(p_0)M$,

(b) $\det\tilde\Gamma(p_0) = (\tilde\Gamma_{11}(p_0) + 2\tilde\Gamma_{12}(p_0))(\tilde\Gamma_{11}(p_0) - \tilde\Gamma_{12}(p_0))^2$,

(c) $\tilde\Gamma_{11}(p_0) - \tilde\Gamma_{12}(p_0) = (\tilde G_{11}(p_0) - \tilde G_{12}(p_0))^{-1}$,

(d) $\tilde\Gamma_{11}(p_0) + 2\tilde\Gamma_{12}(p_0) = (\tilde G_{11}(p_0) + 2\tilde G_{12}(p_0))^{-1}$.

Proof: By explicit computation.

III. SPECTRAL CONSIDERATIONS

In this section we use the momentum space analyticity properties of $\tilde G(p_0)$ and $\tilde\Gamma(p_0)$ as well as the factorization of their determinants to obtain a spectral representation for $\tilde G(p_0)$ (a multicomponent lattice Kallen–Lehman type representation). Using this representation we give criteria for a point p_0 to belong to the mass spectrum.

Similar to Ref. 1, $\tilde G(p_0)$ has a spectral representation given by the following lemma.

Lemma III.1: There exist finite signed measures $d\rho_{\alpha\gamma}$, positive if $\alpha = \gamma$, such that

$$\tilde G_{\alpha\gamma}(p_0) = \int_{[0,e^{-m}]} \frac{1-\lambda^2}{1 - 2\lambda\cos p_0 + \lambda^2}\, d\rho_{\alpha\gamma}(\lambda),$$

where $m = \lim_{x_0\uparrow\infty} (-1/x_0)\ln(\Sigma_{\mathbf{x}}\, G_{\alpha\alpha}(x_0,\mathbf{x}))$. Furthermore $d\rho_{\alpha\alpha} = d\rho_{\gamma\gamma}$ and $|d\rho_{\alpha\gamma}| \leqslant d\rho_{\alpha\alpha}$.

Let $\sigma(M)$ denote the mass spectrum. We know that $\sigma(M) \subset [m,\infty)$ and that $m' \in \sigma(M)$ if and only if $e^{-m'} \in \cup\,\mathrm{supp}\,d\rho_{\alpha\gamma} = \mathrm{supp}\,d\rho_{11}$. From the above representation we can obtain a direct relation between $\mathrm{supp}\,d\rho_{\alpha\gamma}$ and the points $p_0 \in i[m,\infty)$, where $\tilde G_{\alpha\gamma}(p_0)$ is not analytic. It is convenient to write $\tilde G_{\alpha\gamma}(p_0)$ in terms of the variable a and the measure $d\nu_{\alpha\gamma}$ defined by

$$d\nu_{\alpha\gamma}(a) = \frac{1-\lambda(a)^2}{2\lambda(a)}\, d\rho, \quad a(\lambda) = \frac{(1-\lambda)^2}{2\lambda}.$$

As $a(e^{-m}) = \cosh m - 1$ we have

$$\tilde G_{\alpha\gamma}(p_0) = F_{\alpha\gamma}(\cos p_0 - 1),$$

where

$$F_{\alpha\gamma}(z) = \rho_{\alpha\gamma}(\{0\}) + \int_{\cosh m - 1}^{\infty} \frac{d\nu_{\alpha\gamma}(a)}{a - z}$$

is analytic in $\mathbb{C} - [\cosh m - 1, \infty)$. Using the inversion formula

$$\nu_{\alpha\gamma}(d) - \nu_{\alpha\gamma}(c) = \lim_{\epsilon\downarrow 0} \frac{1}{i\pi}\int_c^d [F_{\alpha\gamma}(\mu + i\epsilon) - F_{\alpha\gamma}(\mu - i\epsilon)]\, d\mu,$$

valid for points c, d, $c < d$, of continuity of $d\nu_{\alpha\gamma}$, it follows that $F_{\alpha\gamma}(z)$ is analytic in $z \in [\cosh m - 1, \infty)$ if and only if $z \notin \mathrm{supp}\,d\nu_{\alpha\gamma}$. Thus we have the following lemma.

Lemma III.2: Let $m' \geqslant m$. Then $e^{-m'} \in \mathrm{supp}\,d\rho_{\alpha\gamma}$ if and only if $\tilde G_{\alpha\gamma}$ is not analytic at $p_0 = im'$.

By Lemma II.3 $\tilde\Gamma(p_0)$ is analytic for $|\mathrm{Re}\,p_0| < \pi$, $0 < \mathrm{Im}\,p_0 < -5\ln|c'\beta|$ so that the functions $\tilde G_{\alpha\gamma}(p_0)$ are meromorphic in this region. Then we have the following.

Lemma III.3: $m' \in \sigma(M) \cap [m, -5\log|c'\beta|)$ if and only if $\tilde\Gamma_{11}(im') - \tilde\Gamma_{12}(im') = 0$ or $\tilde\Gamma_{11}(im') + 2\tilde\Gamma_{12}(im') = 0$.

Proof: From the above observation and Lemma III.2 it follows that $e^{-im'} \in \mathrm{supp}\,d\nu_{\alpha\gamma}$ if and only if im' is a pole of $\tilde G_{\alpha\gamma}$. We conclude that $m' \in \sigma(M)$ if and only if im' is a pole of some $\tilde G_{\alpha\gamma}$, which can occur only if $\det\tilde\Gamma(p_0) = (\tilde\Gamma_{11}(p_0) - \tilde\Gamma_{12}(p_0))^2(\tilde\Gamma_{11}(p_0) + 2\tilde\Gamma_{12}(p_0)) = 0$.

IV. IMPLICIT MASS EQUATION AND PROOF OF THEOREM B

In this section we obtain and solve implicit equations for the masses for χ real. The analytic implicit function theorem does not apply directly to the implicit mass equations. However, by introducing an auxiliary complex variable and function we obtain an implicit equation to which the analytic function theorem applies and yields the masses.

Towards obtaining implicit mass equations we obtain the $\beta = 0$ Taylor expansion of $\tilde{\Gamma}(p_0)$ with the β^m, $0 \leqslant m \leqslant 4$, terms explicited. The explicit terms are obtained from the $\beta = 0$ expansion for $\tilde{G}(p_0)$ and the relation $\tilde{G}(p_0)\tilde{\Gamma}(p_0) = I$. The $\tilde{G}(p_0)$ expansion follows from the $\beta = 0$ expansion of $G(x, \beta)$ for certain x given in the following lemma.

Lemma IV.1:

(a) $x = 0$: $G_{11}(x) = 1 + O(\beta)$, $G_{12}(x) = O(\beta^4)$,

(b) $x = (1,\mathbf{0})$: $G_{11}(x) = \dfrac{\beta^4}{r^4} + O(\beta^5)$, $G_{12}(x) = O(\beta^6)$.

The $\beta = 0$ expansion of $\tilde{G}(p_0)$ is given by the following lemma.

Lemma IV.2:

$$\tilde{G}(p_0) = I\Big(1 + \sum_{k=1}^{4} g_k\beta^k + \frac{\beta^4}{r^4}(e^{-ip_0} + e^{ip_0})\Big) + \alpha\beta^4 M + O(\beta^5),$$

where g_k, $1 \leqslant k \leqslant 4$, and α are group representation dependent constants.

The $\beta = 0$ expansion of $\tilde{\Gamma}(p_0)$ is given by the following.

Lemma IV.3:

$$\tilde{\Gamma}(p_0) = I\Big(1 + \sum_{k=1}^{4} \gamma_k\beta^k - \frac{\beta^4}{r^4}(e^{-ip_0} + e^{ip_0})\Big) - \alpha\beta^4 M + O(\beta^5),$$

where $\gamma_k = \gamma_k(g_1, g_2, g_3, g_4)$, $1 \leqslant k \leqslant 4$, are determined from $\tilde{G}(p_0)\tilde{\Gamma}(p_0) = I$ using Lemma IV.2.

Define

$$\Gamma_s(x) = \Gamma(x) - \sum_{m=0}^{4} \frac{\beta^m}{m!}\frac{\partial^m \Gamma}{\partial\beta^m} \quad (x, \beta = 0)$$

and similarly for $\tilde{\Gamma}_s$. Define, for $n = 0,1,\ldots$,

$$\Gamma_s(n,\beta) = \sum_{\mathbf{x}} \Gamma_s(x_0 = n, \mathbf{x},\beta).$$

Let $R^{\pm}(p_0) = \tilde{\Gamma}_{11}(p_0) + ((1 \pm 3)/2)\tilde{\Gamma}_{12}(p_0)$ so that

$$R^{\pm}(p_0) = 1 + \sum_{k=1}^{4}\gamma_k\beta^k - \frac{\beta^4}{r^4}(e^{-ip_0} + e^{ip_0}) - \frac{(1\pm 3)}{2}\alpha\beta^4 + R_s^{\pm}(n = 0,\beta) + \sum_{n=1}^{\infty} R_s^{\pm}(n,\beta)(e^{-ip_0n} + e^{ip_0n}).$$

Introduce the auxiliary complex variable w and functions $H^{\pm}(w,\beta)$ such that $H^{\pm}(w = 1 - (\beta^4/r^4) \times e^{-ip_0}, \beta) = R^{\pm}(p_0)$, where

$$H^{\pm}(w,\beta) = w + \sum_{k=1}^{4}\gamma_k\beta^k - \frac{\beta^4}{r^4(1-w)} - \frac{(1\pm 3)}{2}\alpha\beta^4 + R_s^{\pm}(n = 0,\beta) + \sum_{n=1}^{\infty} R_s^{\pm}(n,\beta) \times\Big[\Big(\frac{r^4}{\beta^4}(1-w)\Big)^n + \Big(\frac{\beta^4}{r^4(1-w)}\Big)^n\Big].$$

As $H^{\pm}(w,\beta)$ are jointly analytic, $H^{\pm}(0,0) = 0$ and $\partial H^{\pm}/\partial w(0,0) = 1$, we have the following.

Lemma IV.4: For $|\beta|$, $|w|$ small, $\exists$ unique analytic functions $w_{\pm}(\beta)$, not necessarily distinct, $w_{\pm}(0) = 0$, such that $H^{\pm}(w_{\pm}(\beta),\beta) = 0$.

We now give the proof of Theorem B. From the spectral considerations of Sec. III and Lemma IV.4 the mass or masses $m_{\pm}(\beta)$ are given by

$$R^{\pm}(p_0 = im_{\pm}(\beta)) = H^{\pm}(w_{\pm} = 1 - (\beta^4/r^4)e^{m_{\pm}(\beta)},\beta) = 0$$

or

$$m_{\pm}(\beta) = -4\ln\beta + 4\ln r + \ln(1 - w_{\pm}(\beta)).$$

The $\beta = 0$ Taylor coefficients c_n can be determined from those of $w_{\pm}(\beta)$ using $H^{\pm}(w,\beta)$ as in Ref. 2.

V. COMPLEX CHARACTER

We treat the case of a complex character in this section in a manner analogous to the χ real case and give the proof of Theorem A.

For $\chi = \chi_r + i\chi_i$ complex we define the 6×6 matrix valued correlation function $G(x,\beta)$ with matrix elements

$$G_{\alpha\gamma uv}(x,\beta) = G_{\alpha\gamma uv}(y;z,\beta) = \lim_{\Lambda\uparrow Z^4} G_{\Lambda\alpha\gamma uv}(y;z,\beta),$$

$x = y - z$, where

$$G_{\Lambda\alpha\gamma uv}(y;z,\beta) = \langle \chi_u(g_{p_y})\chi_v(g_{p_z})\rangle_\Lambda - \langle\chi_u(g_{p_y})\rangle_\Lambda\langle\chi_v(g_{p_z})\rangle_\Lambda,$$

with u, $v = i,r$; the rest of the notation is as in Sec. II. The convolution inverse $\Gamma(x,\beta) = \Gamma(y;z,\beta)$, $x = y - z$, of G is the 6×6 matrix which satisfies

$$\sum_{w,\delta,t} G_{\alpha\delta ut}(y;w)\Gamma_{\delta\gamma tv}(w;z) = \delta_{yz}\delta_{\alpha\gamma}\delta_{uv},$$

and as in Lemmas II.2.3 we have the following.

Lemma V.1:

(a) $|G_{\alpha\gamma uv}(x,\beta)| \leqslant c_1|c\beta|^{4|x_0| + |\mathbf{x}|}$,

(b) $G_{\alpha\alpha uu}((x_0,\mathbf{0}),\beta) \geqslant c_3(c_4\beta)^{4|x_0|}$, $\beta > 0$,

(c) $|\Gamma_{\alpha\gamma uv}(x,\beta)| \leqslant c_2|c'\beta|^{5|x_0| + |\mathbf{x}|}$,

except for $\alpha = \gamma$, $u = v$, $x = (\pm 1,\mathbf{0})$ where the 5 is replaced by 4.

Recall that $\tilde{G}(p_0) \equiv \tilde{G}(p_0,\mathbf{0})$, $\tilde{\Gamma}(p_0) \equiv \tilde{\Gamma}(p_0,\mathbf{0})$, and $I(M)$ are the 3×3 matrices with elements $I_{ij} = \delta_{ij}(M_{ij} = 1 - \delta_{ij})$. We have the following.

Lemma V.2: $\tilde{G}(p_0,\beta)$ is analytic for small $|\beta|$, $|\mathrm{Im}\, p_0| < -4\ln|c\beta|$, $|\mathrm{Re}\, p_0| < \pi$, and

(a) $\tilde{G}(p_0) = \begin{bmatrix} \tilde{G}_{rr}(p_0) & \tilde{G}_{ir}(p_0) \\ \tilde{G}_{ir}(p_0) & \tilde{G}_{ii}(p_0)\end{bmatrix}$,

where $\tilde{G}_{uv}(p_0) = \tilde{G}_{11uv}(p_0)I + \tilde{G}_{12uv}(p_0)M$ is a 3×3 matrix,

(b) $\det\tilde{G}(p_0) = (\det\tilde{G}^{-}(p_0))^2\det\tilde{G}^{+}(p_0)$, where $\tilde{G}^{\pm}(p_0)$ is a 2×2 matrix with elements $\tilde{G}^{\pm}_{uv}(p_0) = \tilde{G}_{11uv}(p_0) + [(1\pm 3)/2]\tilde{G}_{12uv}(p_0)$.

Proof: (a) Follows from symmetry properties of $\tilde{G}$ in a manner analogous to Lemma II.1.

(b) Since the 3×3 matrices $\tilde{G}_{uv}(p_0)$ commute we have

$\det \tilde G(p_0) = \det(\tilde G_{rr}(p_0)\tilde G_{ii}(p_0) - \tilde G_{ir}(p_0)^2)$; calculating this last determinant gives (b).

Lemma V.3: $\tilde\Gamma(p_0,\beta)$ is analytic for small $|\beta|$, $|\mathrm{Im}\, p_0| < -5\log|c'\beta|$, $|\mathrm{Re}\, p_0| < \pi$, and

(a) $\tilde\Gamma(p_0) = \begin{bmatrix} \tilde\Gamma_{rr}(p_0) & \tilde\Gamma_{ir}(p_0) \\ \tilde\Gamma_{ir}(p_0) & \tilde\Gamma_{ii}(p_0)\end{bmatrix}$,

where $\tilde\Gamma_{uv}(p_0) = \tilde\Gamma_{11uv}(p_0)I + \tilde\Gamma_{12uv}(p_0)M$ is a 3×3 matrix.

(b) $\det\tilde\Gamma(p_0) = (\det\tilde\Gamma^-(p_0))^2 \det\tilde\Gamma^+(p_0)$, where $\tilde\Gamma^\pm(p_0)$ is a 2×2 matrix with matrix elements $\tilde\Gamma^\pm_{uv}(p_0) = \tilde\Gamma_{11uv}(p_0) + [(1\pm 3)/2]\tilde\Gamma_{12uv}(p_0)$.

(c) $(\tilde\Gamma^\pm(p_0))^{-1} = \tilde G^\pm(p_0)$.

Proof: The matrices $\tilde G_{\alpha\gamma}$ commute so that

$$\tilde\Gamma(p_0) = (\tilde G(p_0))^{-1} = \begin{bmatrix} \tilde G_{ii}(\tilde G_{rr}\tilde G_{ii} - \tilde G^2_{ir})^{-1} & -\tilde G_{ir}(\tilde G_{rr}\tilde G_{ii} - \tilde G^2_{ir})^{-1} \\ -\tilde G_{ir}(\tilde G_{rr}\tilde G_{ii} - \tilde G^2_{ir})^{-1} & \tilde G_{rr}(\tilde G_{rr}\tilde G_{ii} - \tilde G^2_{ir})^{-1}\end{bmatrix}.$$

The inverse as well as the product of two matrices of the form $aI + bM$ has the same form so that (a) follows.

(b) Similar to (b) of Lemma V.2.

(c) By direct calculation.

We also have the representation

$$\tilde G_{\alpha\gamma uv}(p_0) = \int_{[0,\bar m]} \frac{1-\lambda^2}{1-2\lambda\cos p_0 + \lambda^2} d\rho_{\alpha\gamma uv}(\lambda),$$

where $\bar m = \inf\{m_i, m_r\}$,

$$m_u = \lim_{x_0\uparrow\infty} -\frac{1}{x_0}\log\Big(\sum_{\mathbf{x}} G_{\alpha\alpha uu}(x_0,\mathbf{x})\Big),$$

and $d\rho_{\alpha\gamma uv}$ is a finite signed measure, positive if $\alpha=\gamma$ and $u=v$.

The mass spectrum $\sigma(M)$ is contained in $[\bar m,\infty)$ and $m\in\sigma(M)$ if and only if $e^{-m}\in\cup\,\mathrm{supp}\, d\rho_{\alpha\gamma uv} = \mathrm{supp}\, d\rho_{11rr}\cup\mathrm{supp}\, d\rho_{11ii}$. By arguments similar to those in Sec. III we see that $m\in\sigma(M)\cap[\bar m, -5\log|c'\beta|)$ if and only if it is a pole of some function $\tilde G_{\alpha\gamma uv}$ which results in the following lemma.

Lemma V.4: $\bar m\in\sigma(M)\cap[m, -5\log|c'\beta|)$ if and only if $\det\tilde\Gamma^-(im)=0$ or $\det\tilde\Gamma^+(im)=0$.

We now obtain implicit equations for the zeroes of $\det\tilde\Gamma^\pm(p_0)$.

We have the $\beta=0$ Taylor expansions, valid for $|\mathrm{Im}\, p_0| < 4\log|c\beta|$,

$$\tilde G^\pm_{uu}(p_0,\beta) = \tilde G_{11uu}(p_0,\beta) + [(1\pm 3)/2]\tilde G_{12uu}(p_0,\beta) = \tfrac12 + \sum_{k=1}^4 g^\pm_{k,u}\beta^k + \frac{\beta^4}{32r^4}(e^{ip_0}+e^{-ip_0}) + O(\beta^5),$$

$$\tilde G^\pm_{ir}(p_0,\beta) = \sum_{k=1}^4 c^\pm_k\beta^k + O(\beta^5)$$

so that

$$\det\tilde G^\pm(p_0,\beta) = \frac14 + \sum_{k=1}^4 d^\pm_k\beta^k + \frac{\beta^4}{32r^4}(e^{ip_0}+e^{-ip_0}) + O(\beta^5).$$

Defining

$$\Gamma_s(x) = \Gamma(x) - \sum_{m=0}^4 \frac{\beta^m}{m!}\frac{\partial^m\Gamma}{\partial\beta^m}(x,\beta=0),$$

$$\Gamma_s(n,\beta) = \sum_{\mathbf{x}}\Gamma_s(x_0=n,\mathbf{x},\beta),\ n=0,1,2...,$$

and using $\tilde\Gamma\tilde G = I$, we have the $\beta=0$ expansions of $\tilde\Gamma$, valid for $|\mathrm{Im}\, p_0| < -5\log|c'\beta|$,

$$\tilde\Gamma^\pm_{uu}(p_0,\beta) = \tilde\Gamma_{11uu}(p_0,\beta) + [(1\pm3)/2]\tilde\Gamma_{12uu}(p_0,\beta) = 2 + \sum_{k=1}^4 a^\pm_{k,u}\beta^k - \frac{\beta^4}{8r^4}(e^{-ip_0}+e^{ip_0}) + \Gamma_{s11uu}(0,\beta) + [1\pm3)/2]\Gamma_{s12uu}(0,\beta) + \sum_{n=1}^\infty\Big(\Gamma_{s11uu}(n,\beta) + \frac{1\pm3}{2}\Gamma_{s12uu}(n,\beta)\Big)\times(e^{ip_0n}+e^{-ip_0n})$$

and

$$\tilde\Gamma^\pm_{ir}(p_0,\beta) = \sum_{k=1}^4 b^\pm_k\beta^k + \Gamma_{s11ir}(0,\beta) + \frac{1\pm3}{2}\Gamma_{s12ir}(0,\beta) + \sum_{n=1}^\infty\Big(\Gamma_{s11ir}(n,\beta) + \frac{1\pm3}{2}\Gamma_{s12ir}(n,\beta)\Big)(e^{ip_0n}+e^{-ip_0n}).$$

Introducing the complex variable w, the 2×2 matrices $H^\pm(w,\beta)$ given by

$$H^\pm_{uv}(w = 2-(\beta^4/8r^4)e^{-ip_0},\beta) = \tilde\Gamma^\pm_{uv}(p_0,\beta)$$

and defining

$$F^\pm(w,\beta) = \det H^\pm(w,\beta),$$

we have as in Theorem III.3 of Ref. 2, using the Weierstrass preparation theorem (see Ref. 7), the following lemma.

Lemma V.5: For $|w|$, $|\beta|$ small

(a) $F^\pm(w,\beta)$ is analytic in w,β,

(b) $F^\pm(0,0)=0,\quad \dfrac{\partial F^\pm}{\partial w}(0,0)=0,\quad \dfrac{\partial^2F^\pm}{\partial w^2}(0,0)=2,$

(c) $F^\pm(w,\beta) = (A^\pm_0(\beta) + A^\pm_1(\beta)w + w^2)M^\pm(w,\beta)$,

where $A^\pm_0(\beta), A^\pm_1(\beta)$ are analytic, $A^\pm_0(0) = A^\pm_1(0)=0$, and $M^\pm(w,\beta)$ is analytic in w,β with $M^\pm(w,\beta)\neq 0$.

We now prove Theorem A. From Lemma V.5 the zeroes of $F^\pm(w,\beta)$ are the zeroes of $A^\pm_0(\beta) + A^\pm_1(\beta)w + w^2$. There are only two possibilities.

(1) There exists one zero $w^\pm(\beta)$, where $w^\pm(\beta)$ is analytic and satisfies $F^\pm(w^\pm(\beta),\beta)=0$ and $w^\pm(0)=0$.

(2) There exist two zeroes $w^\pm_1(\beta), w^\pm_2(\beta)$, where both are analytic functions of β or $\beta^{1/2}$ and satisfy $F^\pm(w^\pm_1(\beta),\beta) = F^\pm(w^\pm_2(\beta),\beta)=0$ and $w^\pm_1(0) = w^\pm_2(0)=0$.

The zeroes of $\det\tilde\Gamma(p_0,\beta)$ are of the form im, $m>0$. As $F^\pm(w=2-(\beta^4/8r^4)e^{-ip_0},\beta) = \det\tilde\Gamma^\pm(p_0,\beta)$ we have $m^\pm(\beta) = -4\log\beta + \log 8r^4 + \log(2-w^\pm(\beta))$, where $w^\pm(\beta)$ satisfies (1) or (2) above.

VI. CONCLUDING REMARKS

It would be nice to have an argument to show that there are not more than four mass points without having to introduce and solve the implicit mass equations. Note that in the nondegenerate case this follows from the resolvent representation of $\tilde{G}$. A preliminary reduction due to symmetries of the space lattice as in Refs. 3 and 8 may simplify the analysis here.

ACKNOWLEDGMENT

One of us (MO) was partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

[1]R. Schor, Nucl. Phys. B **222**, 71 (1983).
[2]M. O'Carroll, G. Braga, and R. Schor, "Glueball mass spectrum and mass splitting in 2 + 1 strongly coupled lattice gauge theories," UFMG Physics Dept. Preprint 1983 (to appear in Commun. Math. Phys.).
[3]G. Munster, Nucl. Phys. B **190**, 439 (1981).
[4]B. Berg and A. Billoire, Nucl. Phys. B **221**, 109 (1983).
[5]K. Osterwalder and E. Seiler, Ann. Phys. **110**, 440 (1978).
[6]E. Seiler, *Gauge Theories as a Problem of Constructive Field Theory and Statistical Mechanics, Lecture Notes in Physics 159* (Springer, New York, 1982).
[7]A. I. Markusevich, *Theory of Functions of a Complex Variable* (Prentice-Hall, New Jersey, 1965), Vol. II.
[8]R. Schor, Phys. Lett. B **132**, 161 (1983); CMP **92**, 369 (1984).

The gauge field propagator for QED and linearized QCD in inhomogeneous media

M. Bickeböller, R. Goldflam,[a] and L. Wilets
Institute for Nuclear Theory, Department of Physics, FM-15, University of Washington, Seattle, Washington 98195

(Received 7 January 1985; accepted for publication 15 March 1985)

The gauge field propagator is calculated in configuration space using an expansion in spherical harmonics for the scalar Green function and in vector spherical harmonics for the tensor Green function. We work in the Coulomb gauge where the scalar Green function is instantaneous and the frequency-dependent tensor Green function is transverse. Explicit expressions are obtained to solve for the propagators in the presence of inhomogeneous dielectric media.

I. INTRODUCTION

Several bag models have been introduced in the last years as phenomenological models of quantum chromodynamics (QCD) (see Refs. 1 and 2). One of these is the soliton bag model of Friedberg and Lee.[2,3] A scalar field σ is introduced to represent the complex structure of the vacuum. Color gluon fields are introduced as in QCD, except that they interact with the soliton field through a color dielectric function $\kappa(\sigma)$. The color magnetic susceptibility is $\mu = \kappa^{-1}$. Color confinement is effected by the requirement that $\kappa = 0$ in the physical vacuum, and $\kappa \approx 1$ in the interior of the "bag."

The bag problem corresponds to a system where the dielectric medium is inhomogeneous. Here we consider the general problem of a dielectric function κ as an arbitrary function of $\mathbf{r}$. The calculation of such a propagator can also be used in a wide range of problems in electrodynamics and electrical engineering.[4-6] The homogeneous κ problem was first solved in polar coordinates by Johnson, Howard, and Dudley[6] in 1979.

We calculate the Green functions by making an expansion in spherical harmonics for the scalar Green function and in vector spherical harmonics for the tensor Green function. In Sec. II we derive the differential equations for the propagators, in Sec. III we solve these for the scalar Green function, and in Sec. IV for the tensor Green function. The numerical implementation is described in Sec. V. A brief summary is given in Sec. VI.

II. DIFFERENTIAL EQUATION FOR THE PROPAGATOR

If we include explicit gluon effects only to lowest order in the strong coupling constant, the non-Abelian term in the QCD gauge field tensor

$$F^c_{\mu\nu} = \partial_\mu A^c_\nu - \partial_\nu A^c_\mu + f_{abc} A^a_\mu A^b_\nu \tag{2.1}$$

can be neglected and the QCD Euler–Lagrange equation for the gauge field reduces to

$$\partial^\mu \kappa F_{\mu\nu} = J_\nu . \tag{2.2}$$

We have suppressed the color index, because the equations decouple for different colors and the gluon propagator is therefore diagonal in color indices. The equations are the same as the inhomogeneous Maxwell equations in quantum electrodynamics (QED).

To solve these equtions, we choose the Coulomb gauge

$$\nabla\cdot(\kappa \mathbf{A}) = 0 \tag{2.3}$$

to decouple the equations into the time component

$$\nabla\cdot\kappa\nabla A_0 = -J_0 \tag{2.4}$$

and the space component

$$-\kappa\ddot{\mathbf{A}} + \nabla^2\kappa\mathbf{A} - \nabla\times(\kappa\mathbf{A}\times(1/\kappa)\nabla\kappa) = -\mathbf{J}_t . \tag{2.5}$$

The subscript "t" stands for the transverse component, which satisfies

$$\nabla\cdot\mathbf{J}_t = 0, \tag{2.6}$$

and can be constructed by

$$\mathbf{J}_t = \frac{1}{4\pi}\nabla\times\nabla\times\int d^3r' \frac{\mathbf{J}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} . \tag{2.7}$$

We should note here that it is $\kappa\mathbf{A}$ that is required to be transverse rather than the vector potential $\mathbf{A}$. Equations (2.4) and (2.5) determine the defining differential equations for the Green functions.

III. SCALAR GREEN FUNCTION

The time-independent equation for the scalar Green function $G(\mathbf{r},\mathbf{r}')\equiv G^{00}(\mathbf{r},\mathbf{r}')$ is

$$\nabla\cdot\kappa(\mathbf{r})\nabla G(\mathbf{r},\mathbf{r}') = \delta^3(\mathbf{r}-\mathbf{r}'). \tag{3.1}$$

To simplify this equation and to avoid infinities for $\kappa \to 0$ we define

$$\overline{G}(\mathbf{r},\mathbf{r}') \equiv \sqrt{\kappa(\mathbf{r})}\, G(\mathbf{r},\mathbf{r}')\sqrt{\kappa(\mathbf{r}')}, \tag{3.2}$$

where $\overline{G}$ satisfies

$$(\nabla^2 - W(\mathbf{r}))\overline{G}(\mathbf{r},\mathbf{r}') = -\delta^3(\mathbf{r}-\mathbf{r}'), \tag{3.3}$$

with the "potential" $W(\mathbf{r})$ given by

$$W(\mathbf{r}) \equiv \tfrac{1}{4}|\nabla \ln \kappa(\mathbf{r})|^2 + \tfrac{1}{2}\nabla^2 \ln \kappa(\mathbf{r}). \tag{3.4}$$

We expand $W(\mathbf{r})$ and $\delta^3(\mathbf{r}-\mathbf{r}')$ into spherical harmonics $Y_{lm}(\Omega)$,

$$W(\mathbf{r}) = W_{LM}(r) Y_{LM}(\Omega), \tag{3.5}$$

$$\delta^3(\mathbf{r}-\mathbf{r}') = (1/r^2)\delta(r-r') Y_{LM}(\Omega) Y^*_{LM}(\Omega'), \tag{3.6}$$

where the repeated index summation convention is used throughout.

[a] Present Address: Mission Research Co., P.O. Drawer 719, 735 State Street, Santa Barbara, California 93102.

 0022-2488/85/071810-08$02.50

For the scalar Green function we make the ansatz

$$\overline{G}(\mathbf{r},\mathbf{r}') = C_{\alpha\alpha'} J^{\alpha}(\mathbf{r}_<) N^{\alpha'}(\mathbf{r}_>)$$
$$= c_{\alpha\alpha'} \frac{1}{r_<} j^{\alpha}_{lm}(r_<) Y_{lm}(\Omega_<)$$
$$\times \frac{1}{r_>} n^{\alpha'}_{l'm'}(r_>) Y^{*}_{l',m'}(\Omega_>), \tag{3.7}$$

where the quantities $(r_<,\Omega_<,i_<)$ refer to (r,Ω,i) if $r<r'$ and to (r',Ω',i') if $r>r'$; $(r_>,\Omega_>,i_>)$ is defined correspondingly. After multiplying with $Y^{*}_{lm}(\Omega)$ and integrating over the angular coordinates Ω, the ansatz (3.7) reduces the homogeneous partial differential equation (3.3) to a differential equation for the radial functions $n^{\alpha'}_{lm}(r)$ and $j^{\alpha}_{lm}(r)$:

$$\left\{\left(-\frac{d^2}{dr^2}+\frac{l(l+1)}{r^2}\right)\delta_{ll'}\delta_{mm'} + W_{LM}(r)\langle lm|Y_{LM}|l'm'\rangle\right\}$$
$$\times \begin{Bmatrix} j^{\alpha}_{l'm'}(r) \\ n^{\alpha'}_{l'm'}(r)\end{Bmatrix} = 0. \tag{3.8}$$

The set $\{\alpha\}$ of solutions, which are regular at the origin, is given by $\{j^{\alpha}_{lm}\}$ and the set $\{\alpha'\}$, which are regular at infinity, by $\{n^{\alpha'}_{l'm'}\}$. If $\kappa(\mathbf{r})$ goes asymptotically to a constant greater than zero the corresponding boundary conditions are

$$j^{\alpha}_{lm}(r) \sim r^{l+1}\delta^{\alpha}_{lm}, \qquad \text{for } r\to 0, \tag{3.9a}$$

$$n^{\alpha'}_{l'm'}(r) \sim \frac{1}{r^{l'}}\delta^{\alpha'}_{l'm'}, \qquad \text{for } r\to\infty. \tag{3.9b}$$

If $\kappa(\mathbf{r})$ has a different asymptotic behavior, another choice for the boundary condition must be made at infinity, e.g., $\kappa(\mathbf{r})$ falling off exponentially, requires also an exponential decline for $n^{\alpha'}_{l'm'}(r)$.

If there are no symmetries in the problem, all (l,m) are coupled. In an axially symmetric problem, the potential W can be expanded in terms with the magnetic quantum number $M=0$ only, and then the integral

$$\langle lm|Y_{L0}|l'm'\rangle$$
$$\equiv \int d\Omega\, Y^{*}_{lm}(\Omega) Y_{L0}(\Omega) Y_{l'm'}(\Omega)$$
$$= (-1)^m \left(\frac{(2l+1)(2L+1)(2l'+1)}{4\pi}\right)^{1/2}$$
$$\times \begin{pmatrix} l & L & l' \\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix} l & L & l' \\ -m & 0 & m'\end{pmatrix} \tag{3.10}$$

shows explicitly the conservation of the quantum number m. The second $3j$ symbol vanishes unless $m=m'$. Therefore, the differential equations (3.8) decouple for different m. Similar considerations hold for reflection symmetry: then the expansion of W has terms with even L only, and the first $3j$ symbol in (3.10) shows explicitly the conservation of parity. Given the solutions of (3.8) we then need to determine the coefficients $C_{\alpha\alpha'}$ in (3.7). The ansatz (3.7) automatically satisfies symmetry under interchange of the variables $\mathbf{r}$ and $\mathbf{r}'$:

$$\overline{G}(\mathbf{r},\mathbf{r}') = \overline{G}(\mathbf{r}',\mathbf{r}). \tag{3.11}$$

Continuity at $r=r'$ for all angular coordinates leads to the first matching condition

$$\lim_{\epsilon\to 0}\overline{G}(r+\epsilon,\Omega,r,\Omega') = \lim_{\epsilon\to 0}\overline{G}(r-\epsilon,\Omega,r,\Omega'). \tag{3.12}$$

A second matching condition arises from the defining differential equation (3.3). Using the partial wave expansion of the δ function (3.6), we obtain

$$\lim_{\epsilon\to 0}\int_{r'-\epsilon}^{r'+\epsilon} dr\, r^2\{\nabla^2 - W(\mathbf{r})\}\overline{G}(\mathbf{r},\mathbf{r}') = -Y_{LM}(\Omega)Y^{*}_{LM}(\Omega'). \tag{3.13}$$

By operating with $\int d\Omega\, Y^{*}_{lm}(\Omega)$ and $\int d\Omega'\, Y_{lm}(\Omega')$ on Eqs. (3.12) and (3.13) we get the following sets of linear equations:

$$\{j^{\alpha}_{lm}(r)n^{\alpha'}_{l'm'}(r) - j^{\alpha}_{l'm'}(r)n^{\alpha'}_{lm}(r)\}C_{\alpha\alpha'} = 0, \tag{3.14a}$$

$$\left\{\left(\frac{d}{dr}j^{\alpha}_{lm}(r)\right)n^{\alpha'}_{l'm'}(r) - j^{\alpha}_{l'm'}(r)\left(\frac{d}{dr}n^{\alpha'}_{lm}(r)\right)\right\}C_{\alpha\alpha'} = \delta_{l,l'}\delta_{m,m'}, \tag{3.14b}$$

at any radius r.

The range of the labels α and α' must be equal to the range of (l,m) and (l',m'). First we note that if we have axial symmetry in the problem the matrices $\{\ \}$ in (3.14) become block diagonal in the quantum number m. Second, in the case of parity conservation, there is no mixing between modes of different parity. This means that (l,l') odd and (l,l') even also give a block diagonal form of the matrix. We have more equations than needed to determine the coefficients. Therefore some of these equations are linearly dependent on the others and redundant. In the set (3.14a) all equations are antisymmetric under interchange of (l,m) and (l',m'), therefore only (3.14a) with (say) $l<l'$ are not redundant. Taking the derivative of (3.14a) with respect to r leads to

$$\left\{\left(\frac{d}{dr}j^{\alpha}_{lm}(r)\right)n^{\alpha'}_{l'm'}(r) - j^{\alpha}_{l'm'}(r)\left(\frac{d}{dr}n^{\alpha'}_{lm}(r)\right)\right\}C_{\alpha\alpha'}$$
$$= \left\{\left(\frac{d}{dr}j^{\alpha}_{l'm'}(r)\right)n^{\alpha'}_{lm}(r) - j^{\alpha}_{lm}(r)\left(\frac{d}{dr}n^{\alpha'}_{l'm'}(r)\right)\right\}C_{\alpha\alpha'}. \tag{3.15}$$

This means that all equations (3.14b) with $l<l'$ are redundant. These considerations reduce the two sets of equations to one linearly independent set of equtions: $l<l'$ for (3.14a) and $l\geqslant l'$ for (3.14b). After solving the linear equations for $C_{\alpha\alpha'}$, the scalar Green function is completely determined.

In the special case $\kappa(\mathbf{r})=1$, $\overline{G}(\mathbf{r},\mathbf{r}') = G(\mathbf{r},\mathbf{r}')$, the differential equation (3.8) reduces to

$$\left(-\frac{d^2}{dr^2}+\frac{l(l+1)}{r^2}\right)\begin{Bmatrix} j^{\alpha}_{l}(r) \\ n^{\alpha'}_{l}(r)\end{Bmatrix} = 0, \tag{3.16}$$

which has the solution $j^{\alpha}_{l}(r) = r^{l+1}\delta^{\alpha}_{l}$ and $n^{\alpha'}_{l}(r) = r^{-l}\delta^{\alpha'}_{l}$. The set of linear equations (3.14) has the solution $C_{ll'} = [1/(2l+1)]\delta_{ll'}$. Therefore

$$G(\mathbf{r},\mathbf{r}') = \sum_{l}\frac{1}{2l+1}\frac{r^{l}_<}{r^{l+1}_>}Y_{lm}(\Omega)Y^{*}_{lm}(\Omega'), \tag{3.17}$$

which is the well-known solution of the Poisson equation.

IV. TENSOR GREEN FUNCTION

The differential equation for the vector potential is time dependent. After a Fourier transformation in the time and defining the transverse Green function as

$$\overline{G}^{\,ll'}(\mathbf{r},\mathbf{r}',\omega) = \kappa(\mathbf{r})G^{\,ll'}(\mathbf{r},\mathbf{r}',\omega), \tag{4.1}$$

we get

$$(\omega^2+\nabla^2)\overline{G}^{ll'}(\mathbf{r},\mathbf{r}',\omega)-\epsilon_{ikl}\,\partial^k(\epsilon_{lmn}\overline{G}^{mi'}(\mathbf{r},\mathbf{r}',\omega)\partial^n\ln\kappa(\mathbf{r}))$$
$$=-\delta_t^{ii'}(\mathbf{r}-\mathbf{r}'),\qquad(4.2)$$

with the restriction

$$\partial^i\overline{G}^{ii'}(\mathbf{r},\mathbf{r}',\omega)=0.\qquad(4.3)$$

The Coulomb gauge condition (2.3) requires the asymmetric definition (4.1) of $\overline{G}^{ii'}(\mathbf{r},\mathbf{r}',\omega)$.

The transverse part of the dyadic δ function is defined by

$$\begin{aligned}\delta_t^{ii'}(\mathbf{r}-\mathbf{r}')&\equiv[\delta^{ii'}\delta^3(\mathbf{r}-\mathbf{r}')]_t\\&=\frac{\delta(r-r')}{r^2}\Bigg\{\mathscr{Y}^i_{llm}(\Omega)\mathscr{Y}^{i'*}_{llm}(\Omega')\\&\quad+\left(\sqrt{\frac{l}{2l+1}}\,\mathscr{Y}^i_{l,l+1,m}(\Omega)+\sqrt{\frac{l+1}{2l+l}}\,\mathscr{Y}^i_{l,l-1,m}(\Omega)\right)\\&\quad\times\left(\sqrt{\frac{l}{2l+1}}\,\mathscr{Y}^{i'*}_{l,l+1,m}(\Omega')+\sqrt{\frac{l+1}{2l+1}}\,\mathscr{Y}^{i'*}_{l,l-1,m}(\Omega')\right)\Bigg\}\\&\quad-\sqrt{l(l+1)}\\&\quad\times\left\{\theta(r-r')\frac{r'^{l-1}}{r^{l+2}}\mathscr{Y}^i_{l,l+1,m}(\Omega)\mathscr{Y}^{i'*}_{l,l-1,m}(\Omega')\right.\\&\quad\left.+\theta(r'-r)\frac{r^{l-1}}{r'^{l+2}}\mathscr{Y}^i_{l,l-1,m}(\Omega)\mathscr{Y}^{i'*}_{l,l+1,m}(\Omega')\right\}.\end{aligned}\qquad(4.4)$$

This form of the dyadic δ function was constructed with the help of (2.7) and the definition of the vector spherical harmonics $\mathscr{Y}_{JLM}$ in Edmonds[7]:

$$\mathscr{Y}_{JLM}(\Omega)\equiv\sum_{m,q}\langle L\;m\;1\;q|J\;M\rangle\mathbf{e}_q\,Y_{Lm}(\Omega),\qquad(4.5)$$

where the unit vectors $\mathbf{e}_q$ are

$$\begin{aligned}\mathbf{e}_{+1}&=-\sqrt{\tfrac{1}{2}}(\mathbf{e}_x+i\mathbf{e}_y),\\\mathbf{e}_0&=\mathbf{e}_z,\\\mathbf{e}_{-1}&=\sqrt{\tfrac{1}{2}}(\mathbf{e}_x-i\mathbf{e}_y),\end{aligned}\qquad(4.6)$$

and $\mathbf{e}_{x,y,z}$ are the unit vectors in the cartesian coordinate system. The set

$$\{f_{Mlm}(r)\,\mathscr{Y}_{llm}(\Omega),\nabla\times f_{Elm}(r)\,\mathscr{Y}_{llm}(\Omega)\}\qquad(4.7)$$

is a complete orthonormal basis for transverse vector functions; $f_{Mlm}(r)$ and $f_{Elm}(r)$ are arbitrary functions of r; E and M denote electric and magnetic modes, respectively.

In calculating the tensor Green function, we pursue the same path as for the scalar Green function. Because the tensor Green function must be transverse, we use an expansion in the vector spherical harmonics out of the set (4.7). From (4.4), we see that because of the θ-function contributions, the transverse δ function is, contrary to Morse and Feshbach,[8] nonlocal. Therefore, we now have to solve a differential equation which is inhomogeneous over all space and not only at $|\mathbf{r}|=|\mathbf{r}'|$. A general solution of the differential equation will be a general solution of the homogeneous equation plus a special solution of the inhomogeneous equation. The ansatz for the solution of the homogeneous differential equation is

$$\begin{aligned}\mathbf{F}(\mathbf{r},\omega)&=(1/r)f_{Mlm}(r,\omega)\,\mathscr{Y}_{llm}(\Omega)\\&\quad-i(\nabla\times(1/r)f_{Elm}(r,\omega)\,\mathscr{Y}_{llm}(\Omega))\end{aligned}\qquad(4.8a)$$

$$\equiv f_{jlm}(r,\omega)\,\mathscr{Y}_{jlm}(\Omega),\qquad(4.8b)$$

where

$$\begin{aligned}f_{j,j+1,m}(r,\omega)&=\left(\frac{j}{2j+1}\right)^{1/2}\left(\frac{\partial}{\partial r}-\frac{j}{r}\right)\frac{1}{r}f_{Ejm}(r,\omega),\\f_{jjm}(r,\omega)&=(1/r)f_{Mjm}(r,\omega),\\f_{j,j-1,m}(r,\omega)&=\left(\frac{j+1}{2j+1}\right)^{1/2}\left(\frac{\partial}{\partial r}+\frac{j+1}{r}\right)\frac{1}{r}f_{Ejm}(r,\omega).\end{aligned}\qquad(4.9)$$

Throughout the paper, every set of functions $\{f_{jlm}(r,\omega)\}$ with three indices is equivalent to the set $\{f_{Mlm},f_{Elm}\}$ and fulfills the relations (4.9). The symbol f is generic here and stands for j,n,z,a,b used below.

After expanding $\ln\kappa(\mathbf{r})=\ln\kappa_{LM}(r)Y_{LM}(\Omega)$ into spherical harmonics the homogeneous differential equation

$$(\omega^2+\nabla^2)\mathbf{F}(\mathbf{r},\omega)-\nabla\times(\mathbf{F}(\mathbf{r},\omega)\times\nabla\ln\kappa(\mathbf{r}))=0\qquad(4.10)$$

can be transformed to a coupled set of equations in r by operating with $\int d\Omega\,\mathscr{Y}^*_{jlm}(\Omega)$:

$$\begin{aligned}0&=\left(-\omega^2-\frac{1}{r}\frac{\partial^2}{\partial r^2}r+\frac{l(l+1)}{r^2}\right)\begin{Bmatrix}j^\alpha_{jlm}(r,\omega)\\n^\alpha_{jlm}(r,\omega)\end{Bmatrix}\\&\quad+(-1)^m\sqrt{(2j'+1)(2j+1)/4\pi}\begin{pmatrix}j&j'&L\\-m&m'&M\end{pmatrix}\left(W^{j'l'}_{jl}(LM;r)+V^{j'l'}_{jl}(LM;r)\frac{\partial}{\partial r}\right)\begin{Bmatrix}j^\alpha_{j'l'm'}(r,\omega)\\n^\alpha_{j'l'm'}(r,\omega)\end{Bmatrix},\end{aligned}\qquad(4.11)$$

where

$$V^{j'l'}_{jl}(LM;r)\equiv-\sqrt{(2L+1)(2l'+1)(2l+1)}\begin{pmatrix}l'&L&l\\0&0&0\end{pmatrix}\begin{Bmatrix}j'&j&L\\l&l'&1\end{Bmatrix}\frac{d}{dr}\ln\kappa_{LM}(r),\qquad(4.12)$$

and

$$\begin{aligned}W^{j'l'}_{jl}(LM;r)&\equiv-\sqrt{(2L+1)(2l'+1)(2l+1)}\begin{pmatrix}l'&L&l\\0&0&0\end{pmatrix}\begin{Bmatrix}j'&j&L\\l&l'&1\end{Bmatrix}\\&\quad\times\left(\frac{d^2}{dr^2}\ln\kappa_{LM}(r)+\frac{2}{r}\frac{d}{dr}\ln\kappa_{LM}(r)-\frac{L(L+1)}{r^2}\ln\kappa_{LM}(r)\right)\\&\quad+(2L+1)(2l'+1)\sqrt{(2l+1)L(L+1)l'(l'+1)}\,\frac{1}{r^2}\ln\kappa_{\mathrm{LM}}(\mathrm{r})\end{aligned}$$

$$\times\begin{pmatrix} l' & L & l \\ 0 & 0 & 0 \end{pmatrix}\begin{Bmatrix} j' & j & L \\ l & l' & 1 \end{Bmatrix}\begin{Bmatrix} L & l' & l \\ l' & L & 1 \end{Bmatrix}$$

$$+(-1)^{l+j}\sqrt{(2l'+1)(2l+1)}$$

$$\times\left\{\begin{pmatrix} l' & L+2 & l \\ 0 & 0 & 0 \end{pmatrix}\begin{Bmatrix} j & j' & L \\ 1 & L+1 & l' \end{Bmatrix}\begin{Bmatrix} L+1 & l' & j \\ l & 1 & L+2 \end{Bmatrix}\right.$$

$$\times\sqrt{(2L+5)(L+2)(L+1)}\left(\frac{d}{dr}-\frac{L+1}{r}\right)\left(\frac{d}{dr}-\frac{L}{r}\right)\ln\kappa_{LM}(r)$$

$$-\begin{pmatrix} l' & L & l \\ 0 & 0 & 0 \end{pmatrix}\begin{Bmatrix} j & j' & L \\ 1 & L+1 & l' \end{Bmatrix}\begin{Bmatrix} L+1 & l' & j \\ l & 1 & L \end{Bmatrix}$$

$$\times(L+1)\sqrt{2L+1}\left(\frac{d}{dr}+\frac{L+2}{r}\right)\left(\frac{d}{dr}-\frac{L}{r}\right)\ln\kappa_{LM}(r)$$

$$-\begin{pmatrix} l' & L & l \\ 0 & 0 & 0 \end{pmatrix}\begin{Bmatrix} j & j' & L \\ 1 & L-1 & l' \end{Bmatrix}\begin{Bmatrix} L-1 & l' & j \\ l & 1 & L \end{Bmatrix}$$

$$\times L\sqrt{2L+1}\left(\frac{d}{dr}-\frac{L-1}{r}\right)\left(\frac{d}{dr}+\frac{L+1}{r}\right)\ln\kappa_{LM}(r)$$

$$+\begin{pmatrix} l' & L-2 & l \\ 0 & 0 & 0 \end{pmatrix}\begin{Bmatrix} j & j' & L \\ 1 & L-1 & l' \end{Bmatrix}\begin{Bmatrix} L-1 & l' & j \\ l & 1 & L-2 \end{Bmatrix}$$

$$\left.\times\sqrt{(2L-3)(L-1)L}\left(\frac{d}{dr}+\frac{L}{r}\right)\left(\frac{d}{dr}+\frac{L+1}{r}\right)\ln\kappa_{LM}(r)\right\}. \tag{4.13}$$

Note that the functions f^{α}_{jlm} and n^{α}_{jlm} contain first derivatives. Therefore Eq. (4.11) is a set of coupled differential equations: one second-order equation for each magnetic function and two third-order for each electric function. By elimination this can be reduced to one second-order equation for each electric function:

$$0=\left(-\omega^2-\frac{\partial^2}{\partial r^2}+\frac{l(l+1)}{r^2}\right)\begin{Bmatrix} f^{\alpha}_{Mlm}(r,\omega) \\ n^{\alpha}_{Mlm}(r,\omega)\end{Bmatrix}+\left(A^{Ml'm'}_{Mlm}(r)+B^{Ml'm'}_{Mlm}(r)\frac{\partial}{\partial r}\right)\begin{Bmatrix} f^{\alpha}_{Ml'm'}(r,\omega) \\ n^{\alpha}_{Ml'm'}(r,\omega)\end{Bmatrix}$$

$$+\left(A^{El'm'}_{Mlm}(r)+B^{El'm'}_{Mlm}(r)\frac{\partial}{\partial r}+C^{El'm'}_{Mlm}(r)\frac{\partial^2}{\partial r^2}\right)\begin{Bmatrix} f^{\alpha}_{El'm'}(r,\omega) \\ n^{\alpha}_{El'm'}(r,\omega)\end{Bmatrix}, \tag{4.14a}$$

$$0=\left(-\omega^2-\frac{\partial^2}{\partial r^2}+\frac{l(l+1)}{r^2}\right)\begin{Bmatrix} f^{\alpha}_{Elm}(r,\omega) \\ n^{\alpha}_{Elm}(r,\omega)\end{Bmatrix}+\left(A^{Ml'm'}_{Elm}(r)+B^{Ml'm'}_{Elm}(r)\frac{\partial}{\partial r}\right)\begin{Bmatrix} f^{\alpha}_{Ml'm'}(r,\omega) \\ n^{\alpha}_{Ml'm'}(r,\omega)\end{Bmatrix}$$

$$+\left(A^{El'm'}_{Elm}(r)+B^{El'm'}_{Elm}(r)\frac{\partial}{\partial r}+C^{El'm'}_{Elm}(r)\frac{\partial^2}{\partial r^2}\right)\begin{Bmatrix} f^{\alpha}_{El'm'}(r,\omega) \\ n^{\alpha}_{El'm'}(r,\omega)\end{Bmatrix}. \tag{4.14b}$$

The matrics A, B, C are defined as follows:

$$A^{Ml'm'}_{Mlm}(r)\equiv(-1)^m\left(\frac{(2l+1)(2l'+1)}{4\pi}\right)^{1/2}\begin{pmatrix} l & l' & L \\ -m & m' & M \end{pmatrix}\left(W^{l'l'}_{ll}(LM;r)-\frac{1}{r}V^{l'l'}_{ll}(LM;r)\right), \tag{4.15a}$$

$$B^{Ml'm'}_{Mlm}(r)\equiv(-1)^m\left(\frac{(2l+1)(2l'+1)}{4\pi}\right)^{1/2}\begin{pmatrix} l & l' & L \\ -m & m' & M \end{pmatrix}V^{l'l'}_{ll}(LM;r), \tag{4.15b}$$

$$A^{El'm'}_{Mlm}(r)\equiv(-1)^m\left(\frac{(2l+1)l'(l'+1)}{4\pi}\right)^{1/2}\begin{pmatrix} l & l' & L \\ -m & m' & M \end{pmatrix}$$

$$\times\left(\sqrt{l'}\left(\frac{1}{r}W^{l',l'-1}_{ll}(LM;r)-\frac{2}{r^2}V^{l',l'-1}_{ll}(LM;r)\right)-\sqrt{l'+1}\left(\frac{1}{r}W^{l',l'+1}_{ll}(LM;r)-\frac{2}{r^2}V^{l',l'+1}_{ll}(LM;r)\right)\right), \tag{4.15c}$$

$$B^{El'm'}_{Mlm}(r)\equiv(-1)^m\left(\frac{2l+1}{4\pi}\right)^{1/2}\begin{pmatrix} l & l' & L \\ -m & m' & M \end{pmatrix}$$

$$\times\left(\sqrt{l'+1}\left(W^{l',l'-1}_{ll}(LM;r)+\frac{l'-1}{r}V^{l',l'-1}_{ll}(LM;r)\right)+\sqrt{l'}\left(W^{l',l'+1}_{ll}(LM;r)-\frac{l'+2}{r}V^{l',l'+1}_{ll}(LM;r)\right)\right), \tag{4.15d}$$

$$C^{El'm'}_{Mlm}(r)\equiv(-1)^m\left(\frac{2l+1}{4\pi}\right)^{1/2}\begin{pmatrix} l & l' & L \\ -m & m' & M \end{pmatrix}\left(\sqrt{l'+1}\,V^{l',l'-1}_{ll}(LM;r)+\sqrt{l'}\,V^{l',l'+1}_{ll}(LM;r)\right). \tag{4.15e}$$

$$A^{Ml'm'}_{Elm}(r)\equiv(-1)^m\left(\frac{2l'+1}{l(l+1)4\pi}\right)^{1/2}\begin{pmatrix} l & l' & L \\ -m & m' & M \end{pmatrix}$$

$$\times\Big(\sqrt{l}\,(rW^{l'l'}_{l,l-1}(LM;r)-V^{l'l'}_{l,l-1}(LM;r))+\sqrt{l+1}\,(rW^{l'l'}_{l,l+1}(LM;r)-V^{l'l'}_{l,l+1}(LM;r))\Big). \tag{4.16a}$$

$$B^{Ml'm'}_{Elm}(r)\equiv(-1)^m\Big(\frac{2l'+1}{l(l+1)4\pi}\Big)^{1/2}\begin{pmatrix} l & l' & L\\ -m & m' & M\end{pmatrix}\Big(\sqrt{l}\,rV^{l'l'}_{l,l-1}(LM;r)+\sqrt{l+1}\,rV^{l'l'}_{l,l+1}(LM;r)\Big), \tag{4.16b}$$

$$A^{El'm'}_{Elm}(r)\equiv(-1)^m\Big(\frac{1}{l(l+1)4\pi}\Big)^{1/2}\begin{pmatrix} l & l' & L\\ -m & m' & M\end{pmatrix}$$

$$\times\Big(l'\sqrt{l(l'+1)}\,\Big(W^{l',l'-1}_{l,l-1}(LM;r)-\frac{2}{r}V^{l',l'-1}_{l,l-1}(LM;r)\Big)$$

$$-(l'+1)\sqrt{ll'}\,\Big(W^{l',l'+1}_{l,l-1}(LM;r)-\frac{2}{r}V^{l',l'+1}_{l,l-1}(LM;r)\Big)$$

$$-l'\sqrt{(l+1)(l'+1)}\,\Big(W^{l',l'-1}_{l,l+1}(LM;r)-\frac{2}{r}V^{l',l'-1}_{l,l+1}(LM;r)\Big)$$

$$+(l'+1)\sqrt{(l+1)l'}\,\Big(W^{l',l'+1}_{l,l+1}(LM;r)-\frac{2}{r}V^{l',l'+1}_{l,l+1}(LM;r)\Big)\Big), \tag{4.16c}$$

$$B^{El'm'}_{Elm}(r)\equiv(-1)^m\Big(\frac{1}{l(l+1)4\pi}\Big)^{1/2}\begin{pmatrix} l & l' & L\\ -m & m' & M\end{pmatrix}$$

$$\times\Big(\sqrt{l(l'+1)}\,\Big(rW^{l',l'-1}_{l,l-1}(LM;r)+(l'-1)V^{l',l'-1}_{l,l-1}(LM;r)\Big)$$

$$+\sqrt{ll'}\,\Big(rW^{l',l'+1}_{l,l-1}(LM;r)-(l'+2)V^{l',l'+1}_{l,l-1}(LM;r)\Big)$$

$$-\sqrt{(l+1)(l'+1)}\,\Big(rW^{l',l'-1}_{l,l+1}(LM;r)+(l'-1)V^{l',l'-1}_{l,l+1}(LM;r)\Big)$$

$$-\sqrt{(l+1)l'}\,\Big(rW^{l',l'+1}_{l,l+1}(LM;r)-(l'+2)V^{l',l'+1}_{l,l+1}(LM;r)\Big)\Big), \tag{4.16d}$$

$$C^{El'm'}_{Elm}(r)\equiv(-1)^m\Big(\frac{1}{l(l+1)4\pi}\Big)^{1/2}\begin{pmatrix} l & l' & L\\ -m & m' & M\end{pmatrix}$$

$$\times\Big[\sqrt{l(l'+1)}\,rV^{l',l'-1}_{l,l-1}(LM;r)+\sqrt{ll'}\,rV^{l',l'+1}_{l,l-1}(LM;r)$$

$$-\sqrt{(l+1)(l'+1)}\,rV^{l',l'-1}_{l,l+1}(LM;r)-\sqrt{(l+1)l'}\,rV^{l',l'+1}_{l,l+1}(LM;r)\Big]. \tag{4.16e}$$

Note that Eq. (4.11) and the definitions (4.15) and (4.16) contain implicit summations over (L,M).

The fundamental solutions of the homogeneous equation are $\{j^\alpha_{Mlm}(r,\omega), j^\alpha_{Elm}(r,\omega)\}$ and $\{n^\alpha_{Mlm}(r,\omega), n^\alpha_{Elm}(r,\omega)\}$. The index α denotes the different sets of solutions; there are as many sets as there are (Mlm), (Elm) values. The $\{j^\alpha_{Mlm}, j^\alpha_{Elm}\}$ are regular at the origin and the $\{n^\alpha_{Mlm}, n^\alpha_{Elm}\}$ are regular at infinity. This implies the same boundary conditions as for the scalar Green function [see discussion relating to (3.9)],

$$\mathbf{T}j^\alpha_{Mlm}(r,\omega)\sim r^{l+1}\delta^\alpha_{Mlm}, \quad \text{for } r\to 0, \tag{4.17a}$$

$$n^\alpha_{Mlm}(r,\omega)\sim(1/r^l)\delta^\alpha_{Mlm}, \quad \text{for } r\to\infty, \tag{4.17b}$$

$$j^\alpha_{Elm}(r,\omega)\sim r^{l+1}\delta^\alpha_{Elm}, \quad \text{for } r\to 0, \tag{4.17c}$$

$$n^\alpha_{Elm}(r,\omega)\sim(1/r^l)\delta^\alpha_{Elm}, \quad \text{for } r\to\infty. \tag{4.17d}$$

Writing the special solution of the inhomogeneous Eq. (4.2) in product form, we get

$$\overline{G}^{ii'}_{\text{spec}}(\mathbf{r},\mathbf{r}',\omega)=F^{i}(\mathbf{r},\omega)F''^{i'}(\mathbf{r}'), \tag{4.18}$$

The inhomogeneous differential equation (4.2) for $r\neq r'$ can be rewritten as

$$\mathbf{F}'(\mathbf{r}')\Big[\Big(-\omega^2-\frac{1}{r}\frac{\partial^2}{\partial r^2}r+\frac{l(l+1)}{r^2}\Big)z^{>;<}_{jlm}(r,\omega)$$

$$+(-1)^m\Big(\frac{(2j'+1)(2j+1)(2l'+1)(2l+1)}{4\pi}\Big)^{1/2}\begin{pmatrix} j & j' & L\\ -m & m' & M\end{pmatrix}\Big(W^{j'l'}_{jl}(LM;r)+V^{j'l'}_{jl}(LM;r)\frac{\partial}{\partial r}\Big)z^{>;<}_{j'l'm'}(r,\omega)\Big]$$

$$=-\sqrt{l(l+1)}\begin{cases}(r'^{j-1}/r^{j+2})\delta_{j+1,l}\mathscr{Y}_{j,j-1,m}(\Omega'), & \text{for } r>r',\\ (r^{j-1}/r'^{j+2})\delta_{j-1,l}\mathscr{Y}_{j,j+1,m}(\Omega'), & \text{for } r<r'.\end{cases} \tag{4.19}$$

Then Eq. (4.19) determines $\mathbf{F}'(\mathbf{r}')$ up to a constant

$$\mathbf{F}'(\mathbf{r}') = -i\nabla'\times\begin{cases} r'^{j}\mathscr{Y}_{jjm}(\Omega'), & \text{for } r>r'; \\ 1/r'^{j+1}\mathscr{Y}_{jjm}(\Omega'), & \text{for } r<r'. \end{cases} \tag{4.20}$$

As before $z^{>;<}_{jlm}(r,\omega)$ contains first derivatives. Therefore using (4.9), (4.15), and (4.16), the inhomogeneous equation in r is transformed the same way as the homogeneous equation (4.11):

$$\begin{aligned}
&\Big(-\omega^2-\frac{\partial^2}{\partial r^2}+\frac{l(l+1)}{r^2}\Big)z^{>;<}_{Mlm}(r,\omega) \\
&+\Big(A^{Ml'm'}_{Mlm}(r)+B^{Ml'm'}_{Mlm}(r)\frac{\partial}{\partial r}\Big)z^{>;<}_{Ml'm'}(r,\omega) \\
&+\Big(A^{El'm'}_{Mlm}(r)+B^{El'm'}_{Mlm}(r)\frac{\partial}{\partial r}+C^{El'm'}_{Mlm}(r)\frac{\partial^2}{\partial r^2}\Big)z^{>;<}_{El'm'}(r,\omega) \\
&=0,
\end{aligned} \tag{4.21a}$$

$$\begin{aligned}
&\Big(-\omega^2-\frac{\partial^2}{\partial r^2}+\frac{l(l+1)}{r^2}\Big)z^{>;<}_{Elm}(r,\omega) \\
&+\Big(A^{Ml'm'}_{Elm}(r)+B^{Ml'm'}_{Elm}(r)\frac{\partial}{\partial r}\Big)z^{>;<}_{Ml'm'}(r,\omega) \\
&+\Big(A^{El'm'}_{Elm}(r)+B^{El'm'}_{Elm}(r)\frac{\partial}{\partial r}+C^{El'm'}_{Elm}(r)\frac{\partial^2}{\partial r^2}\Big)z^{>;<}_{El'm'}(r,\omega) \\
&=\frac{1}{2l+1}\begin{cases} r^{-1}, & \text{for } r>r' \\ r^{l+1}, & \text{for } r<r' \end{cases} \\
&\equiv\begin{cases} b^{>}_{Elm}(r), & \text{for } r>r', \\ b^{<}_{Elm}(r), & \text{for } r<r'. \end{cases}
\end{aligned} \tag{4.21b}$$

Note that the equations (4.21a) and (4.21b) are coupled. Then a special solution of the inhomogeneous equation can be found by a quadrature as follows:

$$\begin{pmatrix} z^{>;<}_{Mlm}(r,\omega) \\ z^{>;<}_{Elm}(r,\omega)\end{pmatrix} = \begin{pmatrix} j^{\alpha}_{Mlm}(r,\omega) \\ j^{\alpha}_{Elm}(r,\omega)\end{pmatrix}\int_{r^{>,<}}^{r} dr''\,\frac{W^{>;<}_{j,\alpha}(r'',\omega)}{W(r'',\omega)} + \begin{pmatrix} n^{\alpha}_{Mlm}(r,\omega) \\ n^{\alpha}_{Elm}(r,\omega)\end{pmatrix}\int_{r^{>,<}}^{r} dr''\,\frac{W^{>;<}_{n,\alpha}(r'',\omega)}{W(r'',\omega)}, \tag{4.22}$$

where

$$r^{>,<}=\begin{cases} \infty, & \text{for } r>r', \\ 0, & \text{for } r<r'. \end{cases} \tag{4.23}$$

The Wronskian of the fundamental solutions is

$$\mathrm{W}(\mathrm{r}'',\omega)=\det\begin{pmatrix} j^{\beta}_{Ml'm'}(r'',\omega) & n^{\beta}_{Ml'm'}(r'',\omega) \\ j^{\beta}_{El'm'}(r'',\omega) & n^{\beta}_{El'm'}(r'',\omega) \\ \frac{\partial}{\partial r}j^{\beta}_{Ml'm'}(r'',\omega) & \frac{\partial}{\partial r}n^{\beta}_{Ml'm'}(r'',\omega) \\ \frac{\partial}{\partial r}j^{\beta}_{El'm'}(r'',\omega) & \frac{\partial}{\partial r}n^{\beta}_{El'm'}(r'',\omega) \end{pmatrix}, \tag{4.24}$$

where β,l',m' are the matrix indices for the determinant. The determinant $W_{j,\alpha}(r'',\omega)$ is defined by replacing the column α of the j functions in $W(r'',\omega)$ by

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ -b^{>}_{Elm}(r'')\end{pmatrix} \text{ for } r>r', \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ -b^{<}_{Elm}(r'')\end{pmatrix} \text{ for } r<r'. \tag{4.25}$$

$W_{n,\alpha}(r'',\omega)$ is defined correspondingly by replacing the column α of the n functions. The determinants are well defined, because the index α runs over all (Mlm,Elm). The functions $z^{<}_{jlm}(r,\omega)$ are regular at the origin and the functions $z^{>}_{jlm}(r,\omega)$ are regular at infinity.

The general solution of (4.2) for $r<r'$ is

$$\begin{aligned}
&\overline{G}^{\,ii'}(\mathbf{r},\mathbf{r}',\omega) \\
&=j^{\alpha}_{jlm}(r,\omega)\mathscr{Y}^{i}_{jlm}(\Omega)a^{\alpha}_{jl'm'}(r',\omega)\mathscr{Y}^{i'}_{jl'm'}(\Omega') \\
&\quad+z^{<}_{jlm}(r,\omega)\mathscr{Y}^{i}_{jlm}(\Omega)\Big(-i\nabla'\times\frac{1}{r'^{l'+1}}\mathscr{Y}_{l'l'm'}(\Omega')\Big)^{i'}
\end{aligned} \tag{4.26a}$$

and for $r>r'$ is

$$\begin{aligned}
\overline{G}^{\,ii'}(\mathbf{r},\mathbf{r}',\omega)&=n^{\alpha}_{jlm}(r,\omega)\mathscr{Y}^{i}_{jlm}(\Omega)b^{\alpha}_{jl'm'}(r',\omega)\mathscr{Y}^{i'}_{jl'm'}(\Omega') \\
&\quad+z^{>}_{jlm}(r,\omega)\mathscr{Y}^{i}_{jlm}(\Omega)\Big[-i\nabla'\times r'^{l'}\mathscr{Y}_{l'l'm'}(\Omega')\Big]^{i'}.
\end{aligned} \tag{4.26b}$$

The functions $\{a^{\alpha}_{jlm}\}$ and $\{b^{\alpha}_{jlm}\}$ are determined by applying similar matching conditions as in the scalar case (3.12) and (3.13). These conditions are

$$\lim_{\epsilon\to0}G^{\,ii'}(r+\epsilon,\Omega,r,\Omega',\omega)=\lim_{\epsilon\to0}G^{\,ii'}(r-\epsilon,\Omega,r,\Omega',\omega) \tag{4.27a}$$

and

$$\begin{aligned}
&\lim_{\epsilon\to0}\int_{r'-\epsilon}^{r'+\epsilon}dr\,r^2\Big\{(\omega^2+\nabla^2)\overline{G}^{\,ii'}(\mathbf{r},\mathbf{r}',\omega) \\
&\quad-\epsilon_{ikl}\partial^k\Big(\epsilon_{lmn}\overline{G}^{\,mi'}(\mathbf{r},\mathbf{r}',\omega)\partial^n\ln\kappa(\mathbf{r})\Big)\Big\} \\
&=-\Big\{\mathscr{Y}^{i}_{llm}(\Omega)\mathscr{Y}^{i'*}_{llm}(\Omega') \\
&\quad-\Big(\Big(\frac{l}{2l+1}\Big)^{1/2}\mathscr{Y}^{i}_{l,l+1,m}(\Omega) \\
&\quad+\Big(\frac{l+1}{2l+1}\Big)^{1/2}\mathscr{Y}^{i}_{l,l-1m}(\Omega)\Big) \\
&\quad\times\Big(\Big(\frac{l}{2l+1}\Big)^{1/2}\mathscr{Y}^{i'*}_{l,l+1,m}(\Omega) \\
&\quad+\Big(\frac{l+1}{2l+1}\Big)^{1/2}\mathscr{Y}^{i'*}_{l,l-1,m}(\Omega')\Big)\Big\}.
\end{aligned} \tag{4.27b}$$

These equations are then transformed into linear equations in r for the unknown functions $\{a^{\alpha}_{Elm},a^{\alpha}_{Mlm}\}$ and $\{b^{\alpha}_{Elm},b^{\alpha}_{Mlm}\}$:

$$j^{\alpha}_{Mlm}(r,\omega)a^{\alpha}_{Ml'm'}(r,\omega)-n^{\alpha}_{Mlm}(r,\omega)b^{\alpha}_{Ml'm'}(r,\omega)=0, \tag{4.28a}$$

$$\begin{aligned}
&j^{\alpha}_{Mlm}(r,\omega)a^{\alpha}_{El'm'}(r,\omega)-n^{\alpha}_{Mlm}(r,\omega)b^{\alpha}_{El'm'}(r,\omega) \\
&\quad+z^{<}_{Mlm}(r,\omega)\frac{1}{r^{l'}}-z^{>}_{Mlm}(r,\omega)r^{l'+1}=0,
\end{aligned} \tag{4.28b}$$

$$j^{\alpha}_{Elm}(r,\omega)a^{\alpha}_{Ml'm'}(r,\omega)-n^{\alpha}_{Elm}(r,\omega)b^{\alpha}_{Ml'm'}(r,\omega)=0, \tag{4.28c}$$

$$\begin{aligned}
&j^{\alpha}_{Elm}(r,\omega)a^{\alpha}_{El'm'}(r,\omega)-n^{\alpha}_{Elm}(r,\omega)b^{\alpha}_{El'm'}(r,\omega) \\
&\quad+z^{<}_{Elm}(r,\omega)\frac{1}{r^{l'}}-z^{>}_{Elm}(r,\omega)r^{l'+1}=0,
\end{aligned} \tag{4.28d}$$

$$\begin{aligned}
&\Big(\frac{\partial}{\partial r}j^{\alpha}_{Mlm}(r,\omega)\Big)a^{\alpha}_{Ml'm'}(r,\omega) \\
&\quad-\Big(\frac{\partial}{\partial r}n^{\alpha}_{Mlm}(r,\omega)\Big)b^{\alpha}_{Ml'm'}(r,\omega)=\delta_{l,l'}\delta_{m,m'},
\end{aligned} \tag{4.29a}$$

$$\left(\frac{\partial}{\partial r} j^{\alpha}_{Mlm}(r,\omega)\right) a^{\alpha}_{El'm'}(r,\omega) - \left(\frac{\partial}{\partial r} n^{\alpha}_{Mlm}(r,\omega)\right) b^{\alpha}_{El'm'}(r,\omega)$$

$$+ \left(\frac{\partial}{\partial r} z^{<}_{Mlm}(r,\omega)\right) \frac{1}{r^{l'}} - \left(\frac{\partial}{\partial r} z^{>}_{Mlm}(r,\omega)\right) r^{l'+1} = 0, \quad (4.29b)$$

$$\left(\frac{\partial}{\partial r} j^{\alpha}_{Elm}(r,\omega)\right) a^{\alpha}_{Ml'm'}(r,\omega)$$

$$- \left(\frac{\partial}{\partial r} n^{\alpha}_{Elm}(r,\omega)\right) b^{\alpha}_{Ml'm'}(r,\omega) = 0, \quad (4.29c)$$

$$\left(\frac{\partial}{\partial r} j^{\alpha}_{Elm}(r,\omega)\right) a^{\alpha}_{El'm'}(r,\omega) - \left(\frac{\partial}{\partial r} n^{\alpha}_{Elm}(r,\omega)\right) b^{\alpha}_{El'm'}(r,\omega)$$

$$+ \left(\frac{\partial}{\partial r} z^{<}_{Elm}(r,\omega)\right) \frac{1}{r^{l'}} - \left(\frac{\partial}{\partial r} z^{>}_{Elm}(r,\omega)\right) r^{l'+1} = 0. \quad (4.29d)$$

These linear equations reduce in the same way as in the scalar case to a set of linearly independent equations.

We now consider spherical symmetrical forms of $\kappa(\mathbf{r}) = \kappa(r)$, then the functional matrices (4.12) and (4.13) become diagonal. The magnetic and the electric modes and different l's decouple. Equation (4.11) has, for the magnetic modes, the form

$$\left(-\omega^2 - \frac{\partial^2}{\partial r^2} + \frac{l(l+1)}{r^2} + \left(\frac{d}{dr}\ln\kappa(r)\right)\frac{\partial}{\partial r} + \left(\frac{d^2}{dr^2}\ln\kappa(r)\right)\right)\begin{Bmatrix} j_{Ml}(r,\omega) \\ n_{Ml}(r,\omega)\end{Bmatrix} = 0, \quad (4.30a)$$

and, for the electric modes,

$$\left(-\omega^2 - \frac{\partial^2}{\partial r^2} + \frac{l(l+1)}{r^2} + \left(\frac{d}{dr}\ln\kappa(r)\right)\frac{\partial}{\partial r}\right)\begin{Bmatrix} j_{El}(r,\omega) \\ n_{El}(r,\omega)\end{Bmatrix} = 0. \quad (4.30b)$$

Because all functions decouple, the index α is not needed; also there is no dependence on the magnetic quantum number m. The inhomogeneity vanishes for the magnetic modes. Therefore the inhomogeneous solutions (4.22) are all transverse magnetic. All other equations remain the same except that there is now no summation over α.

For $\kappa(\mathbf{r}) = 1$ we can give the solutions analytically. The terms involving $\ln\kappa(r)$ in (4.30) vanish and the homogeneous equations (4.30) have the solutions

$$j_{Ml}(r,\omega) = j_{El}(r,\omega) = rj_l(\omega r), \quad (4.31a)$$

$$n_{Ml}(r,\omega) = n_{El}(r,\omega) = rn_l(\omega r), \quad (4.31b)$$

where the spherical Bessel functions $j_l(\omega r)$ and the Neumann function $n_l(\omega r)$ are defined as usual. For $\omega = 0$ the solutions reduce to simple power forms:

$$j_{Ml}(r,0) = j_{El}(r,0) = r^{l+1}, \quad (4.32a)$$

$$n_{Ml}(r,0) = n_{El}(r,0) = r^{-l}. \quad (4.32b)$$

The inhomogeneous equation can be rewritten as

$$\left(-\omega^2 - \frac{\partial^2}{\partial r^2} + \frac{l(l+1)}{r^2}\right)\begin{Bmatrix} z^{>}_{El}(r,\omega) \\ z^{<}_{El}(r,\omega)\end{Bmatrix} = \frac{1}{2l+1}\begin{cases} r^{-l}, & \text{for } r > r' \\ r^{l+1}, & \text{for } r < r'. \end{cases} \quad (4.33)$$

Here we have to distinguish the two cases $\omega = 0$ and $\omega \neq 0$. For $\omega \neq 0$ the solutions are

$$z^{>}_{El}(r,\omega) = -(1/\omega^2)[1/(2l+1)]r^{-l}, \quad (4.34a)$$

$$z^{<}_{El}(r,\omega) = -(1/\omega^2)[1/(2l+1)]r^{l+1}, \quad (4.34b)$$

and for $\omega = 0$

$$z^{>}_{El}(r,0) = -[1/2(2l+1)(1-2l)]r^{2-l}, \quad (4.35a)$$

$$z^{<}_{El}(r,0) = -[1/2(2l+1)(2l+3)]r^{l+3}. \quad (4.35b)$$

The linear equations (4.28) and (4.29) give as solutions for $\omega \neq 0$,

$$a_{Ml}(r,\omega) = -\omega r n_l(\omega r), \quad (4.36a)$$

$$a_{El}(r,\omega) = -(1/\omega) r n_l(\omega r), \quad (4.36b)$$

$$b_{Ml}(r,\omega) = -\omega r j_l(\omega r), \quad (4.36c)$$

$$b_{El}(r,\omega) = -(1/\omega) r j_l(\omega r), \quad (4.36d)$$

and for $\omega = 0$,

$$a_{Ml}(r,0) = -[1/(2l+1)]r^{-l}, \quad (4.37a)$$

$$a_{El}(r,0) = -[1/2(2l+1)(1-2l)]r^{2-l}, \quad (4.37b)$$

$$b_{Ml}(r,0) = -[1/(2l+1)]r^{l+1}, \quad (4.37c)$$

$$b_{El}(r,0) = -[1/2(2l+1)(2l+3)]r^{l+3}. \quad (4.37d)$$

Therefore we can write the solutions for $\omega \neq 0$ in a symmetrical way as

$$G^{ii'}(\mathbf{r},\mathbf{r}',\omega) = -\omega[j_l(\omega r_<)\mathscr{Y}_{llm}(\Omega_<)]^{i_<}[n_l(\omega r_>)\mathscr{Y}^*_{llm}(\Omega_>)]^{i_>}$$
$$-(1/\omega)[\nabla\times j_l(\omega r_<)\mathscr{Y}_{llm}(\Omega_<)]^{i_<} \times[\nabla\times n_l(\omega r_>)\mathscr{Y}^*_{llm}(\Omega_>)]^{i_>}$$
$$-(1/\omega^2)[1/(2l+1)][\nabla\times r^l_< \mathscr{Y}_{llm}(\Omega_<)]^{i_<} \times[\nabla\times(1/r_>^{l+1})\mathscr{Y}^*_{llm}(\Omega_>)]^{i_>}. \quad (4.38)$$

The first term is the magnetic mode and the second and third terms compose the electric mode. The second term is the solution of the homogeneous differential equation (4.11) and the third term is the solution of the inhomogeneous differential equation (4.21).

The behavior for $\omega \to 0$ is of interest. We can construct the Green function by using the solution for $\omega = 0$[(4.32), (4.35), (4.37)] and writing it again in a symmetric way. Another instructive way is to expand the Bessel functions in (4.38), $j_l(\omega r)$ and $n_l(\omega r)$, for small arguments ωr:

$$j_l(\omega r) \to \frac{(\omega r)^l}{(2l+1)!!}\left(1 - \frac{(\omega r)^2}{2(2l+3)} + \cdots\right), \quad (4.39a)$$

$$n_l(\omega r) \to -\frac{(2l-1)!!}{(\omega r)^{l+1}}\left(1 - \frac{(\omega r)^2}{2(1-2l)} + \cdots\right), \quad (4.39b)$$

where $(2l+1)!! = (2l+1)(2l-1)\cdots(3)(1)$. Then (4.38) transforms into

$$G^{ii'}(\mathbf{r},\mathbf{r}',0) = \frac{1}{2l+1}\frac{r_<^l}{r_>^{l+1}}[\mathscr{Y}_{llm}(\Omega_<)]^{i_<}[\mathscr{Y}^*_{llm}(\Omega_>)]^{i_>} - \frac{1}{2(2l+1)(2l+3)}(\nabla\times r_<^{l+2}\mathscr{Y}_{llm}(\Omega_<))^{i_<} \times\left(\nabla\times\frac{1}{r_>^{l+1}}\mathscr{Y}^*_{llm}(\Omega_>)\right)^{i_>} - \frac{1}{2(2l+1)(1-2l)}(\nabla\times r_<^l\,\mathscr{Y}_{llm}(\Omega_<))^{i_<} \times\left(\nabla\times\frac{1}{r_>^{l-1}}\mathscr{Y}^*_{llm}(\Omega_>)\right)^{i_>}. \tag{4.40}$$

The divergent terms in the electric mode in (4.38) cancel each other. This nonuniform behavior for $\omega\to 0$ relates to the nonuniqueness of the special solution of the inhomogeneous equation. We can always add to it a solution of the homogeneous differential equation.

V. NUMERICAL IMPLEMENTATION

We have written a computer program which calculates the propagator for special cases. The differential equations developed above are solved using a Runge–Kutta method with variable step size. The linear equations are solved numerically by matrix inversion. In the above calculation of the propagators we used polar coordinates. To test the program, we fold an arbitrary source J^μ, expanded in spherical harmonics, with the propagators and get the potential A^μ expanded in spherical harmonics. By calculating both sides of the differential equations (2.4) and (2.5) in Cartesian coordinates, we check various factors in the propagator very efficiently.

The program for the scalar and tensor Green functions has been implemented and successfully tested for axially and reflectionally symmetric systems. We will use these Green functions to calculate one gluon exchange energy contributions to various hadronic processes including N–N scattering. Work on the calculation of static properties of spherical bags is completed[9] and will be reported in a subsequent publication.

VI. SUMMARY

We have obtained expressions for the gauge field propagator expanded in spherical harmonics for the scalar Green function and in vector spherical harmonics for the tensor Green function. We require the propagators to be regular at infinity and at the origin. The differential equations for the propagators are solved by first calculating the solutions of the homogeneous equation without the δ function inhomogenity. We work in the Coulomb gauge, which requires transversality for the tensor Green function. Therefore we must use also the transverse dyadic δ function, which is nonlocal and causes a nonlocal inhomogenity in the differential equation for the tensor Green function; a special solution of the inhomogeneous differential equation is found by quadrature.

This leads to two sets of solutions, which have to be matched at $r = r'$, where the δ functions are irregular. The matching conditions are obtained from the requirement that the Green function is continuous at $r = r'$ and the first derivative has a discontinuity given by the δ function. The resulting linear equations determine the Green function completely.

The gauge field propagator is used frequently in electrical engineering. The numerical technique used to calculate the propagator as an expansion in spherical harmonics can be used as a powerful tool in problems involving media, where the dielectric function is not constant, e.g., lenses.[4] The technique also allows different boundary conditions.

In the context of QCD the propagator is very useful in problems relating to bag models.[1–3] In the soliton bag model[2,3] the dielectric function κ is a function of the soliton field, which is calculated in general only numerically. Different bag problems might allow simplifications. For static spherical bags only the M1 mode of the tensor Green function is needed.[9] In scattering of two nucleons the problem is axially symmetric, which allows the decoupling of differential equations and linear equations for different m values.

ACKNOWLEDGMENTS

We wish to thank Dr. M. C. Birse for valuable discussions and for a careful reading of the manuscript, and G. Lübeck for assistance in several problems relating to programming. One of us (M. Bickeböller) submitted part of this work as a Diplom thesis at the Institut für theoretische Kernphysik, Universität Bonn, Germany.

This work was supported in part by the U.S. Department of Energy, and a Fulbright Scholarship from the German and U.S. Governments for M. Bickeböller.

[1] A. Chodos, R. L. Jaffe, K. Johnson, C. B. Thorn, and V. F. Weisskopf, Phys. Rev. D **9**, 3471 (1974); G. E. Brown and M. Rho, Phys. Lett. B **82**, 177 (1979); G. E. Brown, M. Rho, and V. Vento, *ibid.* **84**, 383 (1979); A. Chodos and C. B. Thorn, Phys. Rev. D **1**, 2733 (1975); F. Myhrer, G. E. Brown, and Z. Xu, Nucl. Phys. A **362**, 377 (1981); G. A. Miller, A. W. Thomas, and S. Theberge, Phys. Lett. B **91**, 192 (1980); S. Theberge, A. W. Thomas, and G. A. Miller, Phys. Rev. D **22**, 2838 (1980); A. W. Thomas, S. Theberge, and G. A. Miller, *ibid.* **24**, 216 (1981).

[2] K. Huang and D. R. Stump, Phys. Rev. D **14**, 223 (1976); R. Friedberg and T. D. Lee, Phys. Rev. D **15**, 1694 (1977); **16**, 1096 (1977); **18**, 2623 (1978); T. D. Lee, *Particle Physics and Introduction to Field Theory* (Harwood Academic, New York, 1981).

[3] R. Goldflam and L. Wilets, Phys. Rev. D **25**, 1951 (1982).

[4] C. T. Tai, *Dyadic Green's Functions in Electromagnetic Theory* (International Textbook, Scranton, PA, 1971); C. T. Tai, Proc. IEEE **61**, 480 (1973); C. T. Tai and P. Rozenfeld, IEEE Trans. Microwave Theory Tech. **MTT-24**, 597 (1976).

[5] W. Walter, *Gewöhnliche Differentialgleichungen* (Springer, Berlin, 1976), 2. Auflage.

[6] W. A. Johnson, A. Q. Howard, and D. G. Dudley, Radio Sci. **14**, 961 (1979).

[7] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, 1957), Chap. 5.

[8] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw–Hill, New York, 1953), part 2, p. 1781.

[9] L. Wilets, M. Bickeböller, M. C. Birse, and H. Marschall, "Solitons in Nuclear and Elementary Particle Physics," in *Proceedings of the 1984 Lewes Workshop* (World Scientific, Singapore, 1984); M. Bickeböller, M. C. Birse, H. Marschall, and L. Wilets, Phys. Rev. D (to be published).

A note on the meaning of covariant derivatives in supersymmetry

Víctor Aldaya
Departamento de Física Téorica, Facultad de Ciencias Físicas, Universidad de Valencia, Burjasot, Valencia, Spain

José A. de Azcárraga[a)]
International Center for Theoretical Physics, Trieste, Italy

(Received 3 October 1984; accepted for publication 4 January 1985)

We analyze in this paper the group theoretical meaning of the covariant derivatives, and show that they are horizontal left-invariant vector fields on superspace obtained from a (super)Lie group which at the same time exhibits the structure of a principal bundle with a canonical connection. The geometrical construction is general and not restricted to the super-Poincaré group.

I. INTRODUCTION

As is well known, covariant derivatives constitute an essential ingredient for the formulation of supersymmetric theories on Salam–Strathdee superspace.[1] Covariant derivatives were originally introduced[1,2] (see also, e.g., Refs. 3 and 4) as differential operators which are invariant under supertranslations and which are covariant in the sense that they transform as Dirac (Weyl) spinors, and also through the different transformation laws which can be defined for superfields. This is essentially the way they are still presented today in the physical literature while, as far as we know, they have not been discussed in more mathematical analyses (see, e.g., Refs. 5–7) but for a few exceptions (see, e.g., Refs. 8–10). In this paper we wish to present a simple derivation of the covariant derivative which is valid for any graded Lie group having a semidirect structure whose soluble part is the Z_2-graded translation group grTr (i.e., translations plus supertranslations) and which is based on treating superspace as a principal bundle endowed with an invariant connection.

II. THE GRADED TRANSLATION GROUP AND THE SUPER-POINCARÉ GROUP AS GROUP EXTENSIONS

Let us recall first a few notions from group extension theory by an Abelian kernel.[11] An extension of a group B by an Abelian group A on which B acts as a group of operators (i.e., there exists a homomorphism on the group of automorphisms of A, $\sigma:B\to\text{Aut }A$) is given by a group G such that A is normal in G and $G/A\sim B$. Such an extension is characterized by its factor system. Because not all factor systems lead to different extensions, the inequivalent extensions are in one-to-one correspondence with the cohomology group obtained by taking the quotient by the equivalent system of factors (the second cohomology group, i.e., two cocycles/two coboundaries; factor systems which give the same extension differ in a coboundary). When B does not act on A (σ is trivial), A belongs to the center of G and the extension is called central.

According to the above definitions, it is simple to see that the graded translations group grTr is the result of centrally extending the supertranslations by the translations. The fact that the starting point is the supertranslations group exhibits once more the fundamental character of the Fermi variables. Using Majorana spinors in the Weyl realization $\theta=(\theta_\alpha,\overline{\theta}^{\dot\alpha})$ $(\alpha,\dot\alpha=1,2)$, the supertranslations sup Tr group law ($b''=b'*b$) is given by

$$\theta''=\theta'+\theta \quad [\text{representation } D^{0,1/2}\oplus D^{1/2,0} \text{ of } SL(2,C)] \tag{2.1a}$$

or

$$\theta''_\alpha=\theta'_\alpha+\theta_\alpha \quad (D^{0,1/2}),\quad \overline{\theta}''^{\dot\alpha}=\overline{\theta}'^{\dot\alpha}+\overline{\theta}^{\dot\alpha} \quad (D^{1/2,0}). \tag{2.1b}$$

The extension of (2.1) by the ordinary translations requires the definition of a two-cocycle (the "factor system") with values in $\mathbb{R}^4$ (four-translations group $\sim$ Minkowski space), i.e., an application $\xi^\mu:B\times B\to\mathbb{R}^4$ such that $\xi^\mu(b,e)=\xi^\mu(e,b)=0\in\mathbb{R}^4$ and

$$\xi^\mu(b',b)+\xi^\mu(b'*b,b'')=\xi^\mu(b,b'')+\xi^\mu(b',b*b''), \tag{2.2}$$

which is just a consequence of the associativity of the group law. It is easy to see that

$$\begin{aligned}\xi^\mu(\theta',\theta)&=-i\overline{\theta}'\gamma^\mu\theta\\&=-i[\theta'^\alpha(\sigma^\mu)_{\alpha\dot\beta}\,\overline{\theta}^{\dot\beta}+\overline{\theta}'_{\dot\rho}(\sigma^\mu)^{\dot\rho\sigma}\theta_\sigma],\end{aligned} \tag{2.3}$$

where $(\sigma^\mu)_{\alpha\dot\beta}\equiv(1,\boldsymbol{\sigma})$ and $(\sigma^\mu)^{\dot\alpha\beta}\equiv(1,-\boldsymbol{\sigma})$, fulfills (2.2) and thus the finite composition law of the centrally extended group, which in general is given by

$$(a',b')*(a,b)=(a'',b'')=(a'+a+\xi(b',b),b'b), \tag{2.4}$$

is in this case

$$(x''^\mu,\theta'')=(x'^\mu+x^\mu-i\overline{\theta}'\gamma^\mu\theta,\ \ \theta'+\theta). \tag{2.5}$$

There are two ways of obtaining the super-Poincaré group as an extension (now obviously noncentral, $\sigma\neq0$). The first one is by extending the Lorentz group by grTr [(2.5)]. In this case there is a homomorphism $\sigma:L^\uparrow_+\to\text{Aut (grTr)}$ (the action of B on grTr) given by

$$\begin{aligned}&\sigma:\Lambda\in L^\uparrow_+\ [SL(2,C)]\to\sigma(\Lambda),\\&\sigma(\Lambda):(x^\mu,\theta)\to(\Lambda^\mu{}_\nu x^\nu,S(\Lambda)\theta),\end{aligned} \tag{2.6}$$

[a)] On leave of absence from the Departamento de Física Téorica, Facultad de Ciencias Físicas, Universidad de Valencia, Burjasot (Valencia), Spain.

 0022-2488/85/071818-04$02.50

where $S(\Lambda)$ belongs to the spinorial (Dirac) representation of SL (2, C) [(2.1a)]. In this way, one gets the super-Poincaré group as the semidirect extension of L by grTr, with group law ($\Lambda \in L^{\uparrow}_{+}$, $(x, \theta) \in$ grTr),

$$[(x'^{\mu}, \theta'), \Lambda'] * [(x^{\mu}, \theta), \Lambda]$$
$$= [(x'^{\mu} + (\Lambda' x)^{\mu} - i\overline{\theta}'\gamma^{\mu} S(\Lambda')\theta, \quad \theta' + S(\Lambda')\theta), \Lambda'\Lambda]; \tag{2.7a}$$

in terms of dotted and undotted spinors

$$x''^{\mu} = x'^{\mu} + \Lambda'^{\mu}_{\ \nu} x^{\nu} - i[\theta'^{\alpha}(\sigma^{\mu})_{\alpha\dot{\beta}} S^{1/2,0}(\Lambda')^{\dot{\beta}}_{\ \dot{\gamma}} \overline{\theta}^{\dot{\gamma}} + \overline{\theta}'_{\dot{\alpha}}(\sigma^{\mu})^{\dot{\alpha}\beta} S^{0,1/2}(\Lambda')_{\beta}^{\ \gamma}\theta_{\gamma}], \tag{2.7b}$$
$$\theta''_{\alpha} = \theta'_{\alpha} + S^{0,1/2}(\Lambda')_{\alpha}^{\ \beta}\theta_{\beta},$$
$$\overline{\theta}''^{\dot{\alpha}} = \overline{\theta}'^{\dot{\alpha}} + S^{1/2,0}(\Lambda')^{\dot{\alpha}}_{\ \dot{\beta}}\overline{\theta}^{\dot{\beta}}.$$

The other way of obtaining the super-Poincaré group is to extend the group composed of the Lorentz group and supertranslations $(\theta', \Lambda')*(\theta, \Lambda) = (\theta' + S(\Lambda')\theta, \Lambda'\Lambda)$ by ordinary translations; in this case $A = \mathrm{Tr}$, $B = \sup \mathrm{Tr} \circ L$. The group law for an extension of B by A in the general case is

$$(a', b')*(a, b) = (a' + \sigma(b')a + \xi(b', b), b'b), \tag{2.8a}$$

where now $\xi(b'b)$ satisfies

$$\xi(b', b) + \xi(b'b, b'') = \sigma(b')\xi(b, b'') + \xi(b', bb''). \tag{2.8b}$$

Both expressions reduce to (2.4) and (2.2), respectively, for central ($\sigma = 0$) extensions.[12] For the super-Poincaré group, the extension is achieved by means of the following cocycle [cf. (2.3)]

$$\xi^{\mu}(b', b) = -i\overline{\theta}'\gamma^{\mu}S(\Lambda')\theta, \tag{2.9}$$

which is easily seen to satisfy (2.8b) with $\sigma(b') = \Lambda'$, and again the group law (2.7) is obtained.

The above reasonings show that the extension nature of the super-Poincaré group allows us to give different expressions for the group law if the cocycle which defines the extension is modified with the addition of a coboundary. For our purposes it will be sufficient to consider the group grTr [(2.5)]; the modification introduced here by a coboundary is easily transported to the complete super-Poincaré group or it can be directly introduced in (2.8a) by means of different ξ's, instead of simply using (2.9).

In general, a two-cocycle is said to be a coboundary ξ_c if there is an A-valued function on B, $\zeta(b)$, such that $\zeta(e) = 0 \in A$ and

$$\xi_c(b', b) = \zeta(b'b) - \sigma(b')\zeta(b) - \zeta(b). \tag{2.10}$$

For the case of the graded translations group, the extension is central and $\sigma(b')$ is trivial. It is very simple to check that, for instance, the following functions ($\lambda \in \mathbb{R}$):

$$\zeta^{\mu}: \sup \mathrm{Tr} \to \mathbb{R}^4,$$
$$\zeta^{\mu}(\theta) \equiv \zeta^{\mu}(\theta_{\alpha}, \overline{\theta}^{\dot{\alpha}}) = -i(\lambda/2)\overline{\theta}\gamma^5\gamma^{\mu}\theta = \lambda i\theta^{\alpha}(\sigma^{\mu})_{\alpha\dot{\beta}}\overline{\theta}^{\dot{\beta}}, \tag{2.11}$$

generate the family of coboundaries

$$\xi_c(\theta', \theta) = \lambda i\theta'^{\alpha}(\sigma^{\mu})_{\alpha\dot{\beta}}\overline{\theta}^{\dot{\beta}} + \lambda i\theta^{\alpha}(\sigma^{\mu})_{\alpha\dot{\beta}}\overline{\theta}'^{\dot{\beta}}. \tag{2.12}$$

Adding (2.12) to (2.3) we obtain new cocycles defining the same extension but through different group laws. Here we give a couple of new cocycles for two specific values of λ:

$$\lambda = +1, \quad \xi^{\mu}(\theta', \theta) = 2i\theta^{\alpha}(\sigma^{\mu})_{\alpha\dot{\beta}}\overline{\theta}'^{\dot{\beta}}, \tag{2.13a}$$
$$\lambda = -1, \quad \xi^{\mu}(\theta', \theta) = -2i\theta'^{\alpha}(\sigma^{\mu})_{\alpha\dot{\beta}}\overline{\theta}^{\dot{\beta}}. \tag{2.13b}$$

III. SUPERSPACE AS A PRINCIPAL FIBER BUNDLE WITH INVARIANT CONNECTION AND COVARIANT DERIVATIVES

Given a (graded) Lie group law, we may derive from it the expression for the left-invariant and right-invariant vector fields. Both sets of vector fields generate the (graded) Lie algebra by commutation, their commutators being equal except for a minus sign. The left (X^{L}) [right (X^{R})] invariant vector fields of the *graded translations group* corresponding to the cocycles (2.3), (2.13a), and (2.13b) are given, respectively, using Van der Waerden–Weyl notation for the spinors, by

$$X^{\mathrm{L}}_{\alpha} = \frac{\partial}{\partial\theta^{\alpha}} + i(\sigma^{\mu})_{\alpha\dot{\beta}}\overline{\theta}^{\dot{\beta}}\frac{\partial}{\partial x^{\mu}}, \quad X^{\mathrm{R}}_{\alpha} = \frac{\partial}{\partial\theta^{\alpha}} - i(\sigma^{\mu})_{\alpha\dot{\beta}}\overline{\theta}^{\dot{\beta}}\frac{\partial}{\partial x^{\mu}},$$
$$\overline{X}^{\mathrm{L}}_{\dot{\alpha}} = \frac{\partial}{\partial\overline{\theta}^{\dot{\alpha}}} + i\theta^{\beta}(\sigma^{\mu})_{\beta\dot{\alpha}}\frac{\partial}{\partial x^{\mu}}, \quad \overline{X}^{\mathrm{R}}_{\dot{\alpha}} = \frac{\partial}{\partial\overline{\theta}^{\dot{\alpha}}} - i\theta^{\beta}(\sigma^{\mu})_{\beta\dot{\alpha}}\frac{\partial}{\partial x^{\mu}}, \tag{3.1a}$$

$$X^{\mathrm{L}}_{\alpha} = \frac{\partial}{\partial\theta^{\alpha}} + 2i(\sigma^{\mu})_{\alpha\dot{\beta}}\overline{\theta}^{\dot{\beta}}\frac{\partial}{\partial x^{\mu}}, \quad X^{\mathrm{R}}_{\alpha} = \frac{\partial}{\partial\theta^{\alpha}},$$
$$\overline{X}^{\mathrm{L}}_{\dot{\alpha}} = \frac{\partial}{\partial\overline{\theta}^{\dot{\alpha}}}, \quad \overline{X}^{\mathrm{R}}_{\dot{\alpha}} = \frac{\partial}{\partial\overline{\theta}^{\dot{\alpha}}} - 2i\theta^{\beta}(\sigma^{\mu})_{\beta\dot{\alpha}}\frac{\partial}{\partial x^{\mu}}, \tag{3.1b}$$

$$X^{\mathrm{L}}_{\alpha} = \frac{\partial}{\partial\theta^{\alpha}}, \quad X^{\mathrm{R}}_{\alpha} = \frac{\partial}{\partial\theta^{\alpha}} - 2i(\sigma^{\mu})_{\alpha\dot{\beta}}\overline{\theta}^{\dot{\beta}}\frac{\partial}{\partial x^{\mu}},$$
$$\overline{X}^{\mathrm{L}}_{\dot{\alpha}} = \frac{\partial}{\partial\overline{\theta}^{\dot{\alpha}}} + 2i\theta^{\beta}(\sigma^{\mu})_{\beta\dot{\alpha}}\frac{\partial}{\partial x^{\mu}}, \quad \overline{X}^{\mathrm{R}}_{\dot{\alpha}} = \frac{\partial}{\partial\overline{\theta}^{\dot{\alpha}}}, \tag{3.1c}$$

plus $X^{\mathrm{L}}_{\mu} = X^{\mathrm{R}}_{\mu} = \partial/\partial x^{\mu}$. As already mentioned, the X^{L} and X^{R} of (3.1) satisfy commutation (anticommutation) relations which differ in a relative minus sign. Specifically, we have

$$\langle X^{\mathrm{L}}_{\alpha}, \overline{X}^{\mathrm{L}}_{\dot{\beta}}\rangle = 2i(\sigma^{\mu})_{\alpha\dot{\beta}}\,\partial_{\mu}, \quad \langle X^{\mathrm{R}}_{\alpha}, \overline{X}^{\mathrm{R}}_{\dot{\beta}}\rangle = -2i(\sigma^{\mu})_{\alpha\dot{\beta}}\,\partial_{\mu}, \tag{3.2a}$$

$$\langle X^{\mathrm{L}}_{\mu}, X^{\mathrm{L}}_{\alpha}\rangle = 0 = \langle X^{\mathrm{L}}_{\mu}, \overline{X}^{\mathrm{L}}_{\dot{\alpha}}\rangle,$$
$$\langle X^{\mathrm{R}}_{\mu}, X^{\mathrm{R}}_{\alpha}\rangle = 0 = \langle X^{\mathrm{R}}_{\mu}, \overline{X}^{\mathrm{R}}_{\dot{\alpha}}\rangle, \tag{3.2b}$$

and

$$\langle X^{\mathrm{L}}, X^{\mathrm{R}}\rangle = 0. \tag{3.3}$$

It is clear from their expression that we may identify the various vector fields X^{L}_{α} with the covariant derivatives D^{L}_{α} in their different forms; in particular, those of (3.1b) and (3.1c) are associated with the chiral and antichiral representations, respectively. In the same way, the vector fields X^{R} may be identified with the operators Q defining the group action. The compatibility between the restrictions imposed by the covariant derivatives and the group action is guaranteed by the general property (3.3) of commutativity (anticommutativity) of the left- and right-invariant vector fields. Note, nevertheless, that the above are only convenient particular choices of the group law, and that they are not the only possibilities because of the arbitrariness in the election of the coboundary.

On any Lie group there is a left-invariant canonical one-form (see, e.g., Ref. 13, p. 41). It is simple to generalize it for graded Lie groups. Writing

$$\Lambda = \Lambda^{(\alpha)} X^{\mathrm{L}}_{\alpha} + \Lambda^{(\dot\alpha)} X^{\mathrm{L}}_{\dot\alpha} + \Lambda^{(\mu)} X^{\mathrm{L}}_{(\mu)}, \tag{3.4}$$

where $\Lambda^{(\alpha)}$, $\Lambda^{(\dot\alpha)}$ have Grassmann character, we have

$$\Lambda^{(\alpha)} = \Lambda^{(\alpha)}_{\mu}\, dx^{\mu} + \Lambda^{(\alpha)}_{\beta}\, d\theta^{\beta} + \Lambda^{(\alpha)}_{\dot\beta}\, d\bar\theta^{\dot\beta} \tag{3.5}$$

and similar expressions for the other two components $\Lambda^{(\dot\alpha)}$ and $\Lambda^{(\mu)}$. Now, from the definition conditions $\Lambda(X^{\mathrm{L}}_{\mu}) = X^{\mathrm{L}}_{\mu}$, $\Lambda(X^{\mathrm{L}}_{\alpha}) = X^{\mathrm{L}}_{\alpha}$, $\Lambda(X^{\mathrm{L}}_{\dot\alpha}) = X^{\mathrm{L}}_{\dot\alpha}$, we get[14]

$$\Lambda^{(\mu)} = dx^{\mu} + i[d\theta^{\alpha}(\sigma^{\mu})_{\alpha\dot\beta}\bar\theta^{\dot\beta} - \theta^{\beta}(\sigma^{\mu})_{\beta\dot\alpha}\, d\bar\theta^{\dot\alpha}],$$

$$\Lambda^{(\alpha)} = d\theta^{\alpha}, \quad \Lambda^{(\dot\alpha)} = d\bar\theta^{\dot\alpha}; \tag{3.6a}$$

$$\Lambda^{(\mu)} = dx^{\mu} + 2i\, d\theta^{\alpha}(\sigma^{\mu})_{\alpha\dot\beta}\bar\theta^{\dot\beta},$$

$$\Lambda^{(\alpha)} = d\theta^{\alpha}, \quad \Lambda^{(\dot\alpha)} = d\bar\theta^{\dot\alpha}; \tag{3.6b}$$

$$\Lambda^{(\mu)} = dx^{\mu} - 2i\, \theta^{\alpha}(\sigma^{\mu})_{\alpha\dot\beta}\, d\bar\theta^{\dot\beta},$$

$$\Lambda^{(\alpha)} = d\theta^{\alpha}, \quad \Lambda^{(\dot\alpha)} = d\bar\theta^{\dot\alpha}, \tag{3.6c}$$

expressions which correspond to the cocycles given by (2.3), (2.13a), and (2.13b), respectively. Note that the difference between the various $\Lambda^{(\mu)}$ is an exact one-form.

Let us now turn to the fibered structure of Salam–Strathdee superspace. Because (graded translation group)/(translations in Minkowski space) $\sim$ supertranslations, it is clear that superspace has the structure of a principal bundle whose base is the supertranslation group and whose structure group is the ordinary translation group. Furthermore, it is a principal bundle endowed with an invariant connection Θ, which is given by the "vertical" component of Λ, i.e., by the translation Lie algebra (vector) valued one-form $\Lambda^{(\mu)}$. Because the structure group is Abelian, the curvature is simply given by

$$\Omega^{(\mu)} = d\Lambda^{(\mu)} \equiv d\Theta^{(\mu)} = -2i(\sigma^{\mu})_{\alpha\dot\beta}\, d\theta^{\alpha} \wedge d\bar\theta^{\dot\beta}, \tag{3.7}$$

which is obtained from any of the (3.6) because although $\Lambda^{(\mu)}$ depends on the election of the coboundary, the curvature is insensitive to it.[15] Moreover, one may immediately check that the curvature Ω is also given by

$$\Omega(X^{\mathrm{L}}, Y^{\mathrm{L}}) = \langle X^{\mathrm{L}}, Y^{\mathrm{L}}\rangle|^{\mathrm{Tr}} = -\Theta(\langle X^{\mathrm{L}}, Y^{\mathrm{L}}\rangle), \tag{3.8}$$

where $|^{\mathrm{Tr}}$ means the component on the translation Lie algebra and X^{L}, Y^{L} are any left-invariant vector fields generating supertranslations ("horizontal" vector fields); (3.8) is a consequence of extending a general theorem on invariant connections (Ref. 13, p. 103) to graded Lie groups.

We are now in a position to understand why the left-invariant vector fields $X^{\mathrm{L}} \equiv D$ of (3.1) in their various forms are indeed covariant derivatives, even before their behavior under Lorentz transformations is considered. Once a principal bundle is endowed with a connection Θ, the lifting of vector fields X defined on the base (super)manifold is given by the vector field $\hat X$,

$$\hat X / \Theta(\hat X) = 0, \quad \pi^{T} \circ \hat X = X, \tag{3.9}$$

where π is the projection gr Tr$\xrightarrow{\pi}$sup Tr. Here, Θ is the vertical component Λ^{v} of the canonical one-form. In our case, the basic vector fields tangent to the supertranslations manifold (the base space of the bundle) are

$$\frac{\partial}{\partial\theta^{\alpha}}, \quad \frac{\partial}{\partial\bar\theta^{\dot\alpha}}, \tag{3.10}$$

and accordingly their horizontal liftings—by definition, *the covariant derivatives in superspace*—are given by

$$D_{\alpha} = \frac{\partial}{\partial\theta^{\alpha}} - \left[\Lambda^{(\mu)}\left(\frac{\partial}{\partial\theta^{\alpha}}\right)\right]\frac{\partial}{\partial x^{\mu}},$$

$$\bar D_{\dot\alpha} = \frac{\partial}{\partial\bar\theta^{\dot\alpha}} - \left[\Lambda^{(\mu)}\left(\frac{\partial}{\partial\bar\theta^{\dot\alpha}}\right)\right]\frac{\partial}{\partial x^{\mu}}. \tag{3.11}$$

It is clear that (3.11) reproduces the customary expressions for the covariant derivatives [(3.1)] by using the appropriated one-forms $\Lambda^{(\mu)}$ [(3.6)].

As is well known, the covariant derivatives may be used to define chiral *or* antichiral superfields. There is no field which satisfies both $D_{\alpha}\phi = 0$, $\bar D_{\dot\alpha}\phi = 0$; this is due, in this language, to the fact that the (anti)commutator of D_{α} and $\bar D_{\dot\alpha}$ is no longer horizontal. In general, the maximum set of compatible conditions is given by the maximal horizontal left-invariant subalgebra. This observation is especially relevant for the N-extended super-Poincaré group, with or without central charges.

IV. FINAL COMMENTS

The foregoing analysis may be generalized to the complete super-Poincaré group, whose group law is given by (2.7). Looking at it as an extension of (sup Tr) $\circ$ SL(2, C) by the space-time translations, it is not difficult to evaluate the vertical one-form dual of the ordinary translations algebra. The left-invariant vector fields are now

$$X^{\mathrm{L}}_{(x^{\mu})} = \Lambda^{\nu}_{\cdot\mu}\frac{\partial}{\partial x^{\nu}}, \quad X^{\mathrm{L}}_{(\theta_{\alpha})} = D^{\beta} S^{0,1/2}(\Lambda)_{\beta\cdot}^{\ \alpha},$$

$$X^{\mathrm{L}}_{(\bar\theta^{\dot\alpha})} = \bar D_{\dot\beta} S^{1/2,0}(\Lambda)^{\dot\beta}_{\cdot\dot\alpha} \tag{4.1}$$

(note the explicit appearance of the Lorentz transformations because of the noncentral character of the extension) and (3.6) is replaced by

$$\tilde\Lambda^{(\mu)} = (\Lambda^{-1})^{\mu}_{\cdot\nu}\Lambda^{(\nu)}. \tag{4.2}$$

The left-invariant vector fields associated with supertranslations in the base manifold (sup Tr $\circ$ SL(2, C)) are given by

$$S^{0,1/2}(\Lambda)_{\beta\cdot}^{\ \alpha}\frac{\partial}{\partial\theta_{\beta}}, \quad S^{1/2,0}(\Lambda)^{\dot\beta}_{\cdot\dot\alpha}\frac{\partial}{\partial\bar\theta^{\dot\beta}} \tag{4.3}$$

and the horizontal lifting of these fields now leads to

$$\mathbb{D}^{\alpha} = D^{\beta} S^{0,1/2}(\Lambda)_{\beta\cdot}^{\ \alpha}, \quad \mathbb{D}_{\dot\alpha} = D_{\dot\beta} S^{1/2,0}(\Lambda)^{\dot\beta}_{\cdot\dot\alpha}, \tag{4.4}$$

as expected. (The left vector fields associated with rotations and boosts, which we will not write here, are already horizontal and so left unaltered by the lifting.)

The differential operators $\mathbb{D}$, $\overline{\mathbb{D}}$ differ from the earlier covariant derivatives in the presence of the SL(2, C) matrices. But these are regular matrices and may be ignored when using these operators for imposing constraints on the superfields. However, when this is done, the remaining $D, \overline{D}$ are no longer left-invariant vector fields commuting with the generators of the group action (the right-invariant ones). As a result, the differential operators $D, \overline{D}$ transform as Weyl spinors under the Lorentz group, a property customarily used in their definition.

Obviously, this construction also applies to the N-extended super-Poincaré group. Also, the presence of central charges does not alter the fibered structures that we have previously defined because of their central character. For instance, for the $N=2$ super-Poincaré group the principal bundle structure discussed in Sec. III now includes the U(1) parameter $\eta \equiv e^{i\varphi}$ in the base manifold and the covariant derivatives are the left-invariant vector fields

$$\frac{\partial}{\partial \theta_{\beta i}} + \overline{\theta}_{\dot{\gamma} i}(\sigma^\mu)^{\dot{\gamma}\beta} \frac{\partial}{\partial x^\mu} + m\epsilon_{ij}\theta_j^\beta \Xi, \tag{4.5a}$$

$$\frac{\partial}{\partial \overline{\theta}_i^{\dot{\beta}}} + \theta_i^\gamma (\sigma^\mu)_{\gamma\dot{\beta}} \frac{\partial}{\partial x^\mu} - m\epsilon_{ij}\overline{\theta}_{j\dot{\beta}}\, \Xi \quad \left(\Xi \equiv i\eta \frac{\partial}{\partial \eta}\right), \tag{4.5b}$$

where we have interpreted the central charge associated with the U(1) generator Ξ as the mass.[16] (The role of the central charge as the mass is discussed in Refs. 17 and 18, and in Ref. 19 in the context of a pseudoclassical particle model.) Other fibrations are of course possible. In particular, for the $N=2$ super-Poincaré extended by a central charge it is especially interesting to take the group U(1) as the structure group; in this case the above left-invariant vector fields (4.5) are also horizontal.[20] To conclude let us mention that the same procedure can be applied to the super-Galilei group,[21] and that there the covariant derivatives also play the same role.

Note added in proof: After this paper had been completed, the authors became aware of a paper [A. Jadczyk and K. Pilch, Commun. Math. Phys. **78**, 373 (1981)] in which a fiber structure of the superspaces underlying rigid supersymmetry, also with the supertranslations in the base and the translations in the fiber, is discussed. A fibered structure of superspace is also discussed in J. Hruby and J. Souček, Rep. Math. Phys. **14**, 15 (1978). Finally, one of the authors (J.A.) wishes to thank Kurt Sundermayer for an interesting discussion.

ACKNOWLEDGMENTS

One of the authors (J. A.) would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and the United Nations Educational, Scientific, and Cultural Organization (UNESCO) for hospitality at the International Center for Theoretical Physics, Trieste.

This paper has been partially supported by the CAICYT (Spain) under Contract No. 1261-82.

[1]A. Salam and J. Strathdee, Phys. Lett. B **51**, 353 (1974); Phys. Rev. D **11**, 1521 (1975).

[2]S. Ferrara, B. Zumino, and J. Wess, Phys. Lett. B **51**, 239 (1974).

[3]J. Wess and J. Bagger, *Supersymmetry and Supergravity, Princeton Series in Physics* (Princeton U. P., Princeton, 1983).

[4]S. J. Gates, Jr., M. T. Grisaru, M. Roček, and W. Siegel, *Superspace or One Thousand and One Lessons in Supersymmetry* (Benjamin–Cummings, Reading, MA, 1983).

[5]F. A. Berezin, Sov. J. Nucl. Phys. **29**, 857 (1979).

[6]D. A. Leites, Russ. Math. Surv. **35**, 1 (1980) [Usp. Mat. Nauk. **35**, 3 (1980)].

[7]A. Rogers, J. Math. Phys. **22**, 443 (1981); **22**, 939 (1981).

[8]B. Zumino, in *Proceedings of the Conference on Gauge Theories and Modern Field Theory*, edited by R. Arnowitt and P. Nath (MIT, Cambridge, MA, 1975).

[9]J. Wess, in *Topics in Quantum Field Theory and Gauge Theories*, edited by J. A. de Azcárraga (Springer, New York, 1977).

[10]Y. Ne'eman and T. Regge, Riv. Nuovo Cimento **1**, 1 (1978).

[11]See, e.g., A. G. Kurosh, *The Theory of Groups* (Chelsea, New York, 1960), Vol. II, Chap. XII.

[12]One observes that if an extension is both central and semidirect, it is a direct product.

[13]The canonical form of a Lie group is defined, e.g., in S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I. The generalization to a graded Lie group is made along similar lines.

[14]By construction, all the above forms have zero Lie derivative with respect to the right-invariant vector fields (the Q's); thus, one may use (3.6a) say, to define the invariant line element in superspace. Equation (3.6a) is also given in Ref. 8.

[15]We note in passing that $d\theta^\alpha \wedge d\overline{\theta}^{\dot{\beta}} = d\overline{\theta}^{\dot{\beta}} \wedge d\theta^\alpha$ because of the Grassmann character of the θ's; the "symplectic" form is symmetric.

[16]The last terms in (4.5a) and (4.5b) are usually written as $Z_{ij}\theta^{\beta j}$, $\overline{Z}_{ij}\overline{\theta}^{\dot{\beta} j}$.

[17]V. Aldaya and J. A. de Azcárraga, Phys. Lett. B **121**, 331 (1983).

[18]V. Aldaya and J. A. de Azcárraga, in *Supersymmetry and Supergravity '83*, edited by B. Milewski (World Scientific, Singapore, 1983), pp. 466–484.

[19]J. A. de Azcárraga and J. Lukierski, Phys. Lett. B **113**, 170 (1982).

[20]An extensive use of this fibration has been made in the context of a quantization formalism on a group manifold (see also Refs. 17 and 18).

[21]V. Aldaya and J. A. de Azcárraga, J. Phys. A **16**, 2633 (1983).

The Fock wave function as classified by the supergroup chain $U(N/M) \supset OSp(N/M) \supset O(N) \times Sp(M)$

Qi-Zhi Han
Department of Physics, Beijing University, Beijing, People's Republic of China

Hong-Zhou Sun
Department of Physics, Qing-Hua University, Beijing, People's Republic of China

Mei Zhang[a)]
Department of Physics and Atmospheric Science, Drexel University, Philadelphia, Pennsylvania 19104

Da Hsuan Feng[b)]
Theoretical Physics Program, National Science Foundation, Washington, D.C. 20550

(Received 23 July 1984; accepted for publication 14 December 1984)

The wave functions for bosons and fermions, classified according to the supergroup chain $U(N/M) \supset OSp(N/M) \supset O(N) \times Sp(M) \supset \cdots$, are given in this paper. The method presented here adheres closely to the standard raising and lowering method commonly used in quantum mechanics. When the so-called star representation $[n\}$ is decomposed into the irreducible representation of the OSp(N/M), one obtains a reducible, but not completely reducible, representation of the OSp(N/M). This decomposition is quite different from a regular Lie group chain and has very interesting physical implications.

I. INTRODUCTION

As an outgrowth of the boson models for collective nuclear states,[1] dynamical supersymmetry (SUSY)[2] was recently proposed as the unifying symmetry of the even–even[1] and even–odd[3] nuclei. The normal dynamical supersymmetry (or group chains) used in this context is

$$U(6/M) \supset U^B(6) \times U^F(M) \supset \cdots \supset \text{Spin}(3). \tag{1.1}$$

In Eq. (1.1), the superscript B denotes boson groups (fully symmetric representations) and the superscript F denotes fermion groups (antisymmetric representations). In the context of applying it to nuclear physics, a salient feature of such a group chain is that it leads to a dynamical symmetry Hamiltonian which conserves baryonic number. Such chain decompositions have been discussed at length by various authors[2] and will not be repeated here.

On the other hand, there is another (potentially important) group chain, proposed previously[4] (in the present context) as a possibility for the analysis of nuclear spectra, especially when the "standard" decomposition of Eq. (1.1) is not possible. This group chain differs from Eq. (1.1) in a fundamental way in that the symmetry of the supergroup U(N/M) is *not* broken by a regular Lie (product) group but by another supergroup: an orthosymplectic supergroup OSp(N/M). Such a chain is as follows:

$$U(N/M) \supset OSp(N/M) \supset O^B(N) \times Sp^F(M) \supset \cdots \supset \text{Spin}(3). \tag{1.2}$$

In this paper, we shall discuss the mathematical structure of this chain. In Sec. II, a review of the decompositions of the Lie groups $U^B(N)$ and $U^F(M)$ will be discussed. This serves the purpose of defining the necessary mathematical language for the following sections. In Sec. III, the so-called Bose–Fermi realization as well as the Casimir operator for the OSp(N/M) are given, followed by, in the next section (IV), a construction of the irreducible representations for the OSp(N/M) within the same realization. The Fock wave functions are then calculated according to the subgroup chain (1.2) in Sec. V. Finally, a discussion of the mathematical results obtained in this paper together with their physical implications is given in the last section (VI).

II. DECOMPOSITION OF LIE GROUPS $U^B(N)$ AND $U^F(M)$ IN THE BOSE AND FERMI REALIZATIONS

In constructing the group chain based on Eq. (1.1), one needs to futher decompose the unitary algebras $U^B(N)$ and $U^F(M)$. Clearly, the most relevant (for physics, at least) decompositions here are those in the Bose and Fermi realizations. The discussions in this section, which are about such decompositions, will serve as the basis for the subsequent sections. Since the material presented in this section is well known, we shall only quote the relevant results here.

A. Bose realization of $U^B(N)$

In the Bose realization, the generators for the $U^B(N)$ are defined as

$$E_{lm,l'm'} = b^\dagger_{lm} b_{l'm'}, \tag{2.1}$$

where $b^\dagger_\alpha (b_\alpha)[\alpha \equiv (l,m)]$ is the creation (annihilation) boson operator with integer angular momentum quantum number l and z component m. Clearly $N = (2l+1)$. The first- and second-order Casimir operators are

$$C_{1\,U(N)} = \hat{N}_b, \tag{2.2a}$$

$$C_{2\,U(N)} = \hat{N}_b(\hat{N}_b + N - 1), \tag{2.2b}$$

where $\hat{N}_b$ is the boson number operator. In this realization, the irreducible representation (IR) is the fully symmetric representation, denoted as $[n_b]$, where n_b is the total number of bosons.

[a)] On leave of absence from the Physics Department, Qing-Hua University, Beijing, People's Republic of China.

[b)] Permanent address: Drexel University, Philadelphia, Pennsylvania 19104.

Likewise, the generators for the O(N) are defined as

$$\Xi_{\alpha\alpha'} = (-)^{l+m} b^{\dagger}_{l\,-m} b_{l'm'} - (-)^{l'+m'} b^{\dagger}_{l'\,-m'} b_{lm} \,. \tag{2.3}$$

The second-order Casimir operator for the O(N) is

$$C_{2\,\mathrm{O}(N)} = \widehat{N}_b(\widehat{N}_b + N - 2) - 2P^{\dagger}_b \widetilde{P}_b \,, \tag{2.4}$$

where the operators $P^{\dagger}_b$ and $\widetilde{P}_b$ are the two invariants of O(N), defined as

$$P^{\dagger}_b = \sum_{lm} (-)^{l-m} b^{\dagger}_{lm} b^{\dagger}_{l\,-m} (2)^{-1/2}, \tag{2.5a}$$

$$\widetilde{P}_b = P_b = \sum_{lm} (-)^{l-m} \tilde{b}_{lm} \tilde{b}_{l\,-m} (2)^{-1/2},$$

$$\tilde{b}_{lm} = (-)^{l+m} b_{l\,-m}. \tag{2.5b}$$

According to the group chain U(N)⊃O(N)⊃···, one obtains the following orthonormal basis vectors:

$$|n_b\, \sigma_b\, \gamma_b\rangle = \left(\frac{(N + 2\sigma_b - 2)!!}{\delta_b!(N + 2\sigma_b + 2\delta_b - 2)!!}\right)^{1/2} \times P^{\dagger\delta_b}_b |\sigma_b\, \sigma_b\, \gamma_b\rangle \,,$$

$$\delta_b = \tfrac{1}{2}(n_b - \sigma_b), \tag{2.6}$$

where it is assumed that

$$\widetilde{P}_b |\sigma_b\, \sigma_b\, \gamma_b\rangle = 0. \tag{2.7}$$

The quantum number σ_b is the O(N) IR's label while γ_b labels the basis vectors of the IR of O(N). We shall refer to σ_b as the generalized seniority number of O(N), which is in accordance with its definition in Eq. (2.6).

B. Fermi realization of U$^{\mathrm{F}}$(M)

In the Fermi realization, the generators for the U$^{\mathrm{F}}$(M) are

$$E_{j\mu j'\mu'} = a^{\dagger}_{j\mu} a_{j'\mu'} \,, \tag{2.8}$$

where $a^{\dagger}_{\lambda}$ $(a_{\lambda})[\lambda \equiv (j,\mu)]$ is the fermionic creation (annihilation) operator with half-integer angular momentum j and z component μ. Clearly $M = \Sigma(2j+1)$. The first- and second-order Casimir operations are

$$C_{1\,\mathrm{U}(M)} = \widehat{N}_f, \tag{2.9a}$$

$$C_{2\,\mathrm{U}(M)} = \widehat{N}_f(M + 1 - \widehat{N}_f). \tag{2.9b}$$

In the Fermi realization, the IR is fully antisymmetric, denoted as $[1^{n_f}]$ where n_f is the total number of fermions.

Likewise, the generators for Sp(M) are defined as

$$\Xi_{j\mu j'\mu'} = (-)^{j+\mu} a^{\dagger}_{j\,-\mu} a_{j'\mu'} + (-)^{j'+\mu'} a^{\dagger}_{j'\,-\mu'} a_{j\mu}. \tag{2.10}$$

The second-order Casimir operator of Sp(M) is

$$C_{2\,\mathrm{Sp}(M)} = \widehat{N}_f(M + 2 - \widehat{N}_f) + 2P^{\dagger}_f \widetilde{P}_f, \tag{2.11}$$

where $P^{\dagger}_f$ and $\widetilde{P}_f$ are the two invariants of Sp(M), defined as

$$P^{\dagger}_f = \sum_{j\mu} (-)^{j-\mu} a^{\dagger}_{j\mu} a^{\dagger}_{j\,-\mu} (2)^{-1/2}, \tag{2.12a}$$

$$\widetilde{P}_f = \sum_{j\mu} (-)^{j-\mu} \tilde{a}_{j\mu} \tilde{a}_{j\mu} (2)^{-1/2}, \tag{2.12b}$$

where $\tilde{a}_{j\mu} = (-)^{j+\mu} a_{j\,-\mu}$.

According to the group chain U(M)⊃Sp(M)⊃···, the orthonormal basis vectors in this realization can be obtained as

$$|n_f\, \sigma_f\, \nu_f\rangle = \left(\frac{(M - 2\sigma_f - 2\delta_f)!!}{\delta_f!(M - 2\delta_f)!!}\right)^{-1/2} P^{\dagger\delta_f}_f |\sigma_f\, \sigma_f\, \gamma_f\rangle,$$

$$\delta_f = \tfrac{1}{2}(n_f - \sigma_f), \tag{2.13}$$

where it is assumed, in analogy to Eq. (2.6), for the Bose realization, that

$$\widetilde{P}_f |\sigma_f\, \sigma_f\, \gamma_f\rangle = 0. \tag{2.14}$$

The quantum number σ_f is the Sp(M) IR's label while γ_f labels the basis vectors of the IR of Sp(M). We refer to σ_f as the generalized seniority number of Sp(M), which is in accordance with its definition in Eq. (2.13).

III. BOSE–FERMI REALIZATIONS OF U(N/M) AND OSp(N/M)

The generators of the so-called Lie supergroup U(N/M) are

$$E_{lm\,l'm'} = b^{\dagger}_{lm}\, b_{l'm'} \,, \tag{3.1a}$$

$$E_{j\mu\, j'\mu'} = a^{\dagger}_{j\mu}\, a_{j'\mu'} \,, \tag{3.1b}$$

$$F_{lmj\mu} = b^{\dagger}_{lm}\, a_{j\mu} \,, \tag{3.1c}$$

$$F_{j\mu lm} = a^{\dagger}_{j\mu}\, b_{lm} \,. \tag{3.1d}$$

We have assumed that the operators of bosons and fermions are mutually commutative. Calculated according to the usual definition, the first- and second-order Casimir operators are

$$C_{1\,\mathrm{U}(N/M)} = \widehat{N}, \tag{3.2a}$$

$$C_{2\,\mathrm{U}(N/M)} = \widehat{N}(\widehat{N} + N - M - 1), \tag{3.2b}$$

where $\widehat{N}$ is

$$\widehat{N}_b = \widehat{N}_f + \widehat{N}, \tag{3.3}$$

$$\widehat{N}_b = \sum b^{\dagger}_{lm}\, b_{lm}, \tag{3.4a}$$

$$\widehat{N}_f = \sum a^{\dagger}_{j\mu}\, a_{j\mu}. \tag{3.4b}$$

In (3.3), $\widehat{N}_b$ and $\widehat{N}_f$ are, obviously, the respective number operators for bosons and fermions.

The generators of the OSp(N/M) can be written as

$$\Xi_{lml'm'} = (-)^{l+m} b^{\dagger}_{l\,-m}\, b_{l'm'} - (-)^{l'+m'}\, b^{\dagger}_{l'\,-m'}\, b_{lm}, \tag{3.5a}$$

$$\Xi_{j\mu j'\mu'} = (-)^{j+\mu} a^{\dagger}_{j\,-\mu}\, a_{j'\mu'} + (-)^{j'+\mu'}\, a^{\dagger}_{j'\,-\mu'}\, a_{j\mu}, \tag{3.5b}$$

$$\Xi_{lmj\mu} = (-)^{l+m} b^{\dagger}_{l\,-m}\, a_{j\mu} + (-)^{j+\mu}\, a^{\dagger}_{j\,-\mu}\, b_{lm}, \tag{3.5c}$$

where the operators $\Xi_{lml'm'}$ and $\Xi_{j\mu j'\mu'}$ are the so-called even operators while the operators $\Xi_{lmj\mu}$ are the so-called odd generators.

Using the creation and annihilation operators of bosons and fermions, one finds that there are two invariants of OSp(N/M), namely $P^{\dagger}$ and $\widetilde{P}$, which will commute with all the generators of OSp(N/M). The definitions of $P^{\dagger}$ and $\widetilde{P}$ are

$$P^{\dagger} = P^{\dagger}_f + P^{\dagger}_b, \tag{3.6a}$$

$$\widetilde{P} = \widetilde{P}_b + \widetilde{P}_f, \tag{3.6b}$$

where $P^{\dagger}_b$ and $\widetilde{P}_b$ are defined by Eqs. (2.5a) and (2.5b), respectively, while $P^{\dagger}_f$ and $\widetilde{P}_f$ are defined by Eqs. (2.12a) and (2.12b), respectively.

It is straightforward to verify the following commutation relations:

$$[P^\dagger, \Xi_{lml'm'}] = [\widetilde{P}, \Xi_{lm,l'm'}] = 0, \tag{3.7a}$$

$$[P^\dagger, \Xi_{j\mu j'\mu'}] = [\widetilde{P}, \Xi_{j\mu j'\mu'}] = 0, \tag{3.7b}$$

$$[P^\dagger, \Xi_{lmj\mu}] = [\widetilde{P}, \Xi_{lmj\mu}] = 0. \tag{3.7c}$$

It is natural to regard $P^\dagger$ and $\widetilde{P}$ as the generalized pair creation and annihiliation operators. Although it is obvious, it is nevertheless worth stressing that $P = P_b + P_f$ is not an invariant of the OSp(N/M).

The second-order Casimir operator can, in terms of $\widehat{N}$, $P^\dagger$, and $\widetilde{P}$, be written as

$$C_{2\,\mathrm{OSp}(N/M)} = \widehat{N}(\widehat{N} + N - M - 2) - 2P^\dagger \widetilde{P}. \tag{3.8}$$

The results for the Casimir operators in this section are identical to that given in Ref. (5).

IV. IRREDUCIBLE REPRESENTATIONS OF THE OSp(N/M) IN THE FOCK SPACE

As a natural extension of the results of Sec. II, the wave functions, as classified according to the group chain OSp(N/M)$\supset$O(N)$\times$Sp(M), can be written as

$$|\sigma\, \sigma_b\, \gamma_b\, \sigma_f\, \gamma_f\rangle, \tag{4.1}$$

where σ labels the IR of OSp(N/M). The other quantum numbers σ_b, γ_b, σ_f, and γ_f, have previously been defined in Sec. III. From Sec. III, we can easily obtain the basis vectors for the group chain U(N)$\times$U(M)$\supset$O(N)$\times$Sp(M)$\supset\cdots$ as

$$|n_b\, \sigma_b\, \gamma_b\rangle\, |n_f\, \sigma_f\, \gamma_f\rangle. \tag{4.2}$$

We shall now establish a relationship between the basis vectors of (4.1) and (4.2).

It is easy to see that

$$\widetilde{P}\,|\sigma_b\, \sigma_b\, \gamma_b\rangle\, |\sigma_f\, \sigma_f\, \gamma_f\rangle = 0, \tag{4.3}$$

where $|\sigma_b\, \sigma_b\, \gamma_b\rangle$ and $|\sigma_f\, \sigma_f\, \gamma_f\rangle$ were defined previously in Sec. II. If we let

$$|\sigma = \sigma_f + \sigma_b, \sigma_b\, \gamma_b\, \sigma_f\, \gamma_f\rangle = |\sigma_b\, \sigma_b\, \gamma_b\rangle |\sigma_f\, \sigma_f\, \gamma_f\rangle, \tag{4.4}$$

then using Eqs. (3.8) and (4.3), we get

$$\begin{aligned} &C_{2\mathrm{OSp}}(N/M)|\sigma = \sigma_b + \sigma_f, \sigma_b\, \gamma_b\, \sigma_f\, \gamma_f\rangle \\ &\quad = \sigma(\sigma + N - M - 2)|\sigma = \sigma_b + \sigma_f, \sigma_b\, \gamma_b\, \sigma_f\, \gamma_f\rangle. \end{aligned} \tag{4.5}$$

Next, we define a series of operators $Q(2)$, $Q(4)$,..., $Q(2\Delta)$, where

$$Q(2) = (M - 2\sigma_f)P_b^\dagger + (N + 2\sigma_b)P_f^\dagger, \tag{4.6a}$$

$$\begin{aligned} Q(4) &= (M - 2\sigma_f)(M - 2\sigma_f - 2)P_b^{\dagger 2} \\ &\quad + 2(M - 2\sigma_f - 2)(N + 2\sigma_b + 2)P_b^\dagger P_f^\dagger + (N + 2\sigma_b)(N + 2\sigma_b + 2)P_f^{\dagger 2}, \end{aligned} \tag{4.6b}$$

$$\vdots$$

$$Q(2\Delta) = \sum_{\delta=0}^{\Delta} \binom{\Delta}{\delta} \left(\frac{(M - 2\sigma_f - 2\delta)!!(N + 2\sigma_b + 2\Delta - 2)!!}{(M - 2\sigma_f - 2\Delta)!!(N + 2\sigma_b + 2\Delta - 2\delta - 2)!!} \right)^{1/2} P_b^{\dagger\Delta - \delta} P_f^{\dagger\delta}, \tag{4.6c}$$

whose action on $|\sigma, \sigma_b\, \gamma_b\, \sigma_f\, \gamma_f\rangle$ will not alter the value σ_b or σ_f, but only σ. From the commutation relation between $\widetilde{P}$ and $Q(2\Delta)$, it is easy to show that

$$\widetilde{P}Q(2\Delta)|\sigma_b + \sigma_f\, \sigma_b\, \gamma_b\, \sigma_f\, \gamma_f\rangle = 0. \tag{4.7}$$

In this way, the relation between (4.1) and (4.2) can be established. Clearly, $Q(2\Delta)|\sigma_b + \sigma_f\, \sigma_b\, \gamma_b\, \sigma_f\, \gamma_f\rangle$ must be an eigenvector of $C_{2\,\mathrm{OSp}(N/M)}$, just as $|\sigma_b + \sigma_f, \sigma_b\, \gamma_b\, \sigma_f\, \gamma_f\rangle$ is but with different eigenvalue, that is,

$$\begin{aligned} &C_{2\,\mathrm{OSp}(N/M)}Q(2\Delta)|\sigma_b + \sigma_f, \sigma_b\, \gamma_b\, \sigma_f\, \gamma_f\rangle \\ &\quad = (\sigma_b + \sigma_f + 2\Delta)(\sigma_b + \sigma_f + 2\Delta + N - M - 2) \\ &\quad\quad \times Q(2\Delta)|\sigma_b + \sigma_f, \sigma_b\, \gamma_b\, \sigma_f\, \gamma_f\rangle. \end{aligned} \tag{4.8}$$

Using Eqs. (2.6) and (2.12), we obtain

$$\begin{aligned} &|\sigma = \sigma_b + \sigma_f + 2\Delta\, \sigma_b\, \gamma_b\, \sigma_f\, \gamma_f\rangle \\ &\quad = C'Q(2\Delta)|\sigma_b + \sigma_f\, \sigma_b\, \gamma_b\, \sigma_f\, \gamma_f\rangle \\ &\quad = C \sum_{\delta_b + \delta_f = \Delta} \binom{\Delta_f}{\delta_b} \\ &\quad\quad \times \left(\frac{(M - 2\sigma_f - 2\delta)!!(N + 2\sigma_b + 2\Delta - 2)!!}{(M - 2\sigma_f - 2\Delta)!!(N + 2\sigma_b + 2\delta_b - 2)!!} \right)^{1/2} \\ &\quad\quad \times |n_b = \sigma_b + 2\delta_b, \sigma_b\, \gamma_b\rangle |n_f = \sigma_f + 2\delta_f, \sigma_f\, \gamma_f\rangle. \end{aligned} \tag{4.9}$$

The C and C' in (4.9) are normalization constants, the indices δ_f and Δ satisfy the inequalities

$$\delta_f \leqslant M/2, \quad \Delta \leqslant M/2, \tag{4.10}$$

which result from the inherent Pauli exclusion principle. The quantum numbers σ, σ_b, γ_b, σ_f, γ_f will determine the wave functions entirely. Thus σ, and only σ, labels the IR of OSp(N/M) in a Bose–Fermi realization. Thus, because of Eq. (4.7), we shall refer to σ here as the generalized seniority number of OSp(N/M).

The wave function $|\sigma\, \sigma_b\, \gamma_b\, \sigma_f\, \gamma_f\rangle$ is also an eigenvector of the total particle number operator $\widehat{N}$ with eigenvalue σ, i.e.,

$$\widehat{N}\,|\sigma\, \sigma_b\, \gamma_b\, \sigma_f\, q_f\rangle = \sigma|\sigma\, \sigma_b\, \gamma_b\, \sigma_f\, \gamma_f\rangle. \tag{4.11}$$

For a given IR$\langle\sigma\rangle$ of OSp(N/M), decomposed according to the group chain OSp(N/M)$\supset$O(N)$\times$Sp(M)$\supset\cdots$, we have

$$\begin{aligned} &\sigma_f = 0, 1, \ldots, M/2, \\ &\sigma_b + \sigma_f = \sigma, \sigma - 2, \ldots, \sigma - M. \end{aligned} \tag{4.12}$$

All the wave functions (basis vectors of IR $\langle\sigma\rangle$) can be derived from (4.9).

We notice that the wave functions here with different quantum numbers are orthogonal. For illustration, consider the IR$\langle\sigma\rangle$ of OSp(3/2)[4], where we have

$$|\sigma\, \sigma\, \gamma_b\, 0\, \gamma_f\rangle = |\sigma\, \sigma\, \gamma_b\rangle |0\, 0\, \gamma_f\rangle,$$

$$|\sigma\, \sigma - 1\;\, \gamma_b\, 1\, \gamma_f\rangle = |\sigma - 1\; \sigma - 1\, \gamma_b\rangle |1\, 1\, \gamma_f\rangle,$$

$$|\sigma\,\sigma-2\,\gamma_b\,0\,\gamma_f\rangle=\sqrt{2}|\sigma\,\sigma-2\,\gamma_b\rangle|0\,0\,\gamma_f\rangle + \sqrt{2\sigma-1}|\sigma-2\,\sigma-2\,\gamma_b\rangle|2\,0\,\gamma_f\rangle. \tag{4.13}$$

V. DECOMPOSITION ACCORDING TO THE SUPERGROUP CHAIN U(N/M)$\supset$OSp(N/M) $\supset$O(N)$\times$Sp(M)$\supset\cdots$

In the Bose–Fermi realization, the eigenvalue n of total particle number $\widehat{N}$ labels the IR of U(N/M). From (4.11), the wave function $|\sigma\,\sigma_b\,\gamma_b\,\sigma_f\,\gamma_f\rangle$ in this supergroup chain is

$$\begin{array}{llllll} |n(=\sigma) & \sigma & \sigma_b & \gamma_b & \sigma_f & \gamma_f\rangle \\ \mathrm{U}(N/M) & \mathrm{OSp}(N/M) & \mathrm{O}(N) & & \mathrm{Sp}(M) & \end{array}$$
$$=|\sigma\,\sigma_b\,\gamma_b\,\sigma_f\,\gamma_f\rangle. \tag{5.1}$$

Operating repeatedly on the state given in Eq. (5.1) by the invariant $P^\dagger$ of OSp(N/M), we can obtain the general wave functions, classified by the supergroup chain U(N/M) $\supset$OSp(N/M)$\supset$O(N)$\times$Sp(M)$\supset\cdots$, as

$$\begin{aligned}|n=\sigma+2\rho\,\sigma\,\sigma_b\,\gamma_b\,\sigma_f\,\gamma_f\rangle &= CP^{\dagger\rho}|n=\sigma\,\sigma\,\sigma_b\,\gamma_b\,\sigma_f\,\gamma_f\rangle \\ &= C\sum_{\delta_b+\delta_f=\rho}\binom{\rho}{\delta_b}P_b^{\dagger\delta_b}P_f^{\dagger\delta_f}|n=\sigma\,\sigma\,\sigma_b\,\gamma_b\,\sigma_f\,\gamma_f\rangle.\end{aligned} \tag{5.2}$$

For a definite IR $[n\}$ of U(N/M), we have

$$\begin{aligned}&\sigma=n,n-2,\ldots,1\text{ or }0,\\ &\sigma_b+\sigma_f=\sigma,\sigma-2,\ldots,\sigma-M,\\ &\sigma_f=0,1,\ldots,(M/2).\end{aligned} \tag{5.3}$$

We see that, in particular, when $\sigma=\sigma_b+\sigma_f$, (5.2) can be written explicitly in the following form:

$$\begin{aligned}&|n=\sigma+2\rho\,\sigma=\sigma_b+\sigma_f\,\sigma_b\,\gamma_b\,\sigma_f\,\gamma_f\rangle \\ &=C\sum_{\delta_b+\delta_f=\rho}\binom{\rho}{\delta_b}\times\left(\frac{(N+2\sigma_b+2\delta_b-2)!!(M-2\delta_f)!!}{(N+2\sigma_b-2)!!((M-2\sigma_f-2\delta_f)!!}\right)^{1/2} \\ &\quad\times|\sigma_b+2\,\delta_b\,\sigma_b\,\gamma_b\rangle|\sigma_f+2\,\delta_f\,\sigma_f\,\gamma_f\rangle.\end{aligned} \tag{5.4}$$

As an example, the wave functions of the IR $[n\}$ of U(3/2) are given in Table I.

From the table we see that the wave functions $|2\,2\,0\,\gamma_b\,0\,\gamma_f\rangle$ and $|2\,0\,0\,\gamma_b\,0\,\gamma_f\rangle$, $|3\,3\,1\,\gamma_b\,0\,\gamma_f\rangle$ and $|3\,1\,1\,\gamma_b\,0\,\gamma_f\rangle$, $|4\,4\,2\,\gamma_b\,0\,\gamma_f\rangle$ and $|4\,2\,2\,\gamma_b\,0\,\gamma_f\rangle,\ldots$, etc., are not orthogonal according to the usual meaning of orthogonality. This means the wave functions of the supergroup chain with the same quantum number n, σ_b, γ_b, σ_f, γ_f but different σ are not orthogonal. This nonorthogonality property is general for any supergroup chain of the type U(N/M)$\supset$OSp(N/M)$\supset$O(N)$\times$Sp(M).

VI. EXAMPLE: U(3/2)$\supset$OSp(3/2)$\supset$O(3)$\times$Sp(2)

Here we give an example of such a decomposition. This decomposition is different from that of a regular Lie algebra chain. It is well known that a Hermitian (unitary) representation of a Lie algebra is completely reducible when it is decomposed according to a compact Lie algebra chain. In order to study the properties of superalgebras, Scheunert, Nahm, and Rittenberg[5] have generalized the concept of Hermitian representations to Lie superalgebras. They defined the adjoint and grade adjoint operations instead of the usual adjoint operations in Lie algebras. For completeness and for the purpose of the present discussion, we will briefly summarize their results here.

An adjoint operation in a Lie superalgebra L is a mapping $A\to A^\dagger$ of L into itself which satisfies the four conditions of Eq. (6.1):

the adjoint of an even (odd) element is even (odd), (6.1a)

$$(aA+bB)^\dagger=a^*A^\dagger+b^*B^\dagger, \tag{6.1b}$$

$$\langle A,B\rangle^\dagger=\langle B^\dagger,A^\dagger\rangle, \tag{6.1c}$$

$$(A^\dagger)^\dagger=A, \tag{6.1d}$$

TABLE I. The wave functions of U(3/2) classified according to U(3/2)$\supset$OSp(3/2)$\supset$O(3)$\times$Sp(2).

n	σ	σ_b	σ_f	$\|n\,\sigma\,\sigma_b\,\gamma_b\,\sigma_f\,\gamma_f\rangle$
0	0	0	0	$\|0\,0\,\gamma_b\rangle\|0\,0\,\gamma_f\rangle$
1	1	1	0	$\|1\,1\,\gamma_b\rangle\|0\,0\,\gamma_f\rangle$
		0	1	$\|0\,0\,\gamma_b\rangle\|1\,1\,\gamma_f\rangle$
2	2	2	0	$\|2\,2\,\gamma_b\rangle\|0\,0\,\gamma_f\rangle$
		1	1	$\|1\,1\,\gamma_b\rangle\|1\,1\,\gamma_f\rangle$
		0	0	$\sqrt{\frac{2}{5}}\|2\,0\,\gamma_b\rangle\|0\,0\,\gamma_f\rangle+\sqrt{\frac{3}{5}}\|0\,0\,\gamma_b\rangle\|2\,0\,\gamma_f\rangle$
	0	0	0	$\sqrt{\frac{3}{5}}\|2\,0\,\gamma_b\rangle\|0\,0\,\gamma_f\rangle+\sqrt{\frac{2}{5}}\|0\,0\,\gamma_b\rangle\|2\,0\,\gamma_f\rangle$
3	3	3	0	$\|3\,3\,\gamma_b\rangle\|0\,0\,\gamma_f\rangle$
		2	1	$\|2\,2\,\gamma_b\rangle\|1\,1\,\gamma_f\rangle$
		1	0	$\sqrt{\frac{2}{7}}\|3\,1\,\gamma_b\rangle\|0\,0\,\gamma_f\rangle+\sqrt{\frac{5}{7}}\|1\,1\,\gamma_b\rangle\|2\,0\,\gamma_f\rangle$
3	1	1	0	$\sqrt{\frac{5}{7}}\|3\,1\,\gamma_b\rangle\|0\,0\,\gamma_f\rangle+\sqrt{\frac{2}{7}}\|1\,1\,\gamma_b\rangle\|2\,0\,\gamma_f\rangle$
		0	1	$\|2\,0\,\gamma_b\rangle\|1\,1\,\gamma_f\rangle$
4	4	4	0	$\|4\,4\,\gamma_b\rangle\|0\,0\,\gamma_f\rangle$
		3	1	$\|3\,3\,\gamma_b\rangle\|1\,1\,\gamma_f\rangle$
		2	0	$\sqrt{\frac{2}{9}}\|4\,2\,\gamma_b\rangle\|0\,0\,\gamma_f\rangle+\sqrt{\frac{7}{9}}\|2\,2\,\gamma_b\rangle\|2\,0\,\gamma_f\rangle$
4	2	2	0	$\sqrt{\frac{7}{9}}\|4\,2\,\gamma_b\rangle\|0\,0\,\gamma_f\rangle+\sqrt{\frac{2}{9}}\|2\,2\,\gamma_b\rangle\|2\,0\,\gamma_f\rangle$
		1	1	$\|3\,1\,\gamma_b\rangle\|1\,1\,\gamma_f\rangle$
		0	0	$\sqrt{\frac{4}{9}}\|4\,0\,\gamma_b\rangle\|0\,0\,\gamma_f\rangle+\sqrt{\frac{5}{9}}\|2\,0\,\gamma_b\rangle\|2\,0\,\gamma_f\rangle$
4	0	0	0	$\sqrt{\frac{5}{9}}\|4\,0\,\gamma_b\rangle\|0\,0\,\gamma_f\rangle+\sqrt{\frac{4}{9}}\|2\,0\,\gamma_b\rangle\|2\,0\,\gamma_f\rangle$

where A, B L, a,b, are complex numbers and

$$\langle A,B\rangle = \begin{cases} [A,B], & \text{when either } A \text{ or } B, \text{ or both are even,} \\ \{A,B\}, & \text{when } A \text{ and } B \text{ are both odd.} \end{cases}$$

A grade adjoint operation in a Lie superalgebra L is a mapping $A \to A^{\dagger}$ of L into itself which satisfies the conditions in (6.2):

the grade adjoint of an even (odd) element is even (odd), (6.2a)

$$(aA + bB) = a^*A + b^*B, \tag{6.2b}$$

$$\langle A,B\rangle = (-)^{\gamma\cdot\beta}\langle B,A\rangle, \tag{6.2c}$$

$$(A^{\dagger})^{\dagger} = (-)^{\gamma}A, \tag{6.2d}$$

where γ and β are the degrees of the elements A and B. The generalized Hermitian representations in Lie superalgebras are the star and grade star representations. If Lie superalgebra L is equipped with an adjoint (grade adjoint) operation, then the star (grade star) representation of L in a graded Hilbert space V is a representation ρ which satisfies

$$\rho(A^{\dagger}) = \rho(A)^{\dagger} \tag{6.3a}$$

In the realization of U(N/M) and OSp(N/M) as we have discussed previously, the adjoint of bosonic and fermionic operators have already been defined in the Fock space. From (3.1) we see that adjoint operation in the usual sense maps U(N/M) into itself, because it satisfies the conditions given by (6.1). So the finite-dimensional representations $[n\}$ of U(N/M) are, by definition, the star representations. However, from (3.5c) we see the usual adjoints of the odd elements of OSp(N/M)

$$\Xi^{\dagger}_{lmju} = (-)^{j+u+l-m} \times [(-)^{l-m} b^{\dagger}_{lm}\, a_{j\,-u} - (-)^{j-u} a^{\dagger}_{ju} b_{l\,-m}] \tag{6.4}$$

do not belong to OSp(N/M). The usual adjoint does not map OSp(N/M) into itself. Thus, OSp(N/M) neither has the adjoint operation nor the star representation in this realization. Therefore, the irreducible representations $\langle\sigma\rangle$ are not star representations. This is consistent with the conclusion that[6] "There is no finite-dimensional star representation of OSp(N/M) when $N \neq 2$."

Many authors have previously decomposed[6,7,8] the star representation $[n\}$ according to a compact Lie algebra chain; it is thus completely reducible. Yet, when one decomposes $[n\}$ according to the superalgebra chain U(N/M) $\supset$OSp(N/M)$\supset\cdots$, then one is actually decomposing a star representation according to the nonstar representation $\langle\sigma\rangle$. Therefore, it is reasonable in Sec. V to expect that OSp(N/M) is a reducible but not completely reducible representation, even though we may still use the quantum number σ for state classification.

In order to understand the above points easily, we shall demonstrate it in detail via the example U(3/2)$\supset$OSp(3/2)$\supset$O(3)$\times$Sp(2). The generators of OSp(3/2) can be written as

$$\begin{aligned} &L_1 = -\Xi_{1\,-110}, \quad L_0 = \Xi_{1\,-111}, \quad L_{-1} = -\Xi_{1011}, \\ &J_1 = -\Xi_{1/2\,-1/2\,1/2\,-1/2}/2\sqrt{2}, \\ &J_0 = \Xi_{1/2\,1/2\,1/2\,-1/2}/2, \\ &J_{-1} = -\Xi_{1/2\,1/2\,1/2\,1/2}/2\sqrt{2}, \\ &T_{1p(1/2)q} = (-)^{p+q+1/2}\Xi_{1\,-p(1/2)\,-q}, \\ &p = 0, \pm 1, \quad q = \pm \tfrac{1}{2}. \end{aligned} \tag{6.5}$$

In the decomposition OSp(3/2)$\supset$O(3)$\times$Sp(2), we can choose the simultaneous eigenvectors $|lmju\rangle$ of L^2, L_0, J^2, and J_0 as the basis vectors. The reduced matrix element is defined as

$$\begin{aligned} &\langle l'\,m'\,j'\,\mu'|T_{1p(1/2)q}|l\,m\,j\,\mu\rangle \\ &\quad = \langle l\,m 1 p|l'\,m'\rangle\langle j\,\mu\,\tfrac{1}{2}\,q|j'\,\mu'\rangle\langle l'\,j'\,|\,|T|\,|l\,j\rangle, \end{aligned} \tag{6.6}$$

where $\langle l\,m\,1\,p|l'\,m'\rangle$, $\langle j\,\mu\,\frac{1}{2}\,q|j'\,\mu'\rangle$ are the Clebsch–Gordan coefficients with $l = \sigma_b$, $m = \gamma_b$, $j = \sigma_f/2$, and $\mu = \gamma_f$. The multiplets $|l\,m\,j\,\mu\rangle$ contained in a representation of OSp(3/2) are decided entirely by the reduced matric elements $\langle l'\,j'|\,|T|\,|l\,j\rangle$.

From (5.2) we obtained the three multiplets with the same quantum number n and σ, when $\sigma \geqslant 2$.

$$\begin{aligned} &|n\,\sigma\,\sigma-1\,m\,1\,\mu\rangle = |n-1\,\sigma-1\,m\rangle|1\,1\,\mu\rangle = |\chi(\sigma)\rangle, \\ &|n\,\sigma\,\sigma\,m\,0\,0\rangle = [(n+\sigma+1)/(2n+1)]^{1/2}|n\,\sigma\,m\rangle|0\,0\,0\rangle \\ &\qquad + [(n-\sigma)/(2n+1)]^{1/2}|n-2\,\sigma\,m\rangle|2\,0\,0\rangle \\ &\qquad = |\psi_1(\sigma)\rangle, \\ &|n\,\sigma\,\sigma-2\,m\,0\,0\rangle \\ &\quad = [(n-\sigma+2)/(2n+1)]^{1/2}|n\,\sigma-2\,m\rangle|0\,0\,0\rangle \\ &\qquad + [(n+\sigma-1)/(2n+1)]^{1/2}|n-2\,\sigma-2\,m\rangle|2\,0\,0\rangle \\ &\quad = |\phi_1(\sigma)\rangle. \end{aligned} \tag{6.7}$$

The wave function orthogonal to $|\phi_1(\sigma)\rangle(|\psi_1(\sigma)\rangle)$ is $|\phi_2(\sigma)\rangle$ $(|\psi_2(\sigma)\rangle)$, where

$$\begin{aligned} &|\phi_2(\sigma)\rangle = [(n+\sigma-1)/(2n+1)]^{1/2}|n\,\sigma-2\,m\rangle|0\,0\,0\rangle \\ &\qquad - [(n-\sigma+2)/(2n+1)]^{1/2} \\ &\qquad \times|n-2\,\sigma-2\,m\rangle|2\,0\,0\rangle, \\ &|\psi_2(\sigma)\rangle = [(n-\sigma)/(2n+1)]^{1/2}|n\,\sigma\,m\rangle|0\,0\,0\rangle \\ &\qquad - [(n+\sigma+1)/(2n+1)]^{1/2}|n-2\,\sigma\,m\rangle|2\,0\,0\rangle. \end{aligned} \tag{6.8}$$

The corresponding reduced matrix elements are

$$\begin{aligned} &\langle\psi_1(\sigma)|\,|T|\,|\chi(\sigma)\rangle = [2\sigma(2n+1)/(2\sigma+1)]^{1/2}, \\ &\langle\chi(\sigma)|\,|T|\,|\psi_1(\sigma)\rangle \\ &\quad = -(2\sigma+1)[\sigma/(2n+1)(2\sigma+1)]^{1/2}, \\ &\langle\phi_1(\sigma)|\,|T|\,|\chi(\sigma)\rangle \\ &\quad = -[(2\sigma-2)(2n+1)/(2\sigma-3)]^{1/2}, \\ &\langle\chi(\sigma)|\,|T|\,|\phi_1(\sigma)\rangle \\ &\quad = (2\sigma-3)[(\sigma-1)/(2n+1)(2\sigma-1)]^{1/2}, \\ &\langle\chi(\sigma)|\,|T|\,|\psi_2(\sigma)\rangle \\ &\quad = -2[\sigma(n-\sigma)(n+\sigma+1)/(2n+1)(2\sigma-1)]^{1/2}, \\ &\langle\chi(\sigma)|\,|T|\,|\phi_2(\sigma)\rangle \\ &\quad = -2[(\sigma-1)(n-\sigma+2)(n+\sigma-1)/ \end{aligned} \tag{6.9}$$

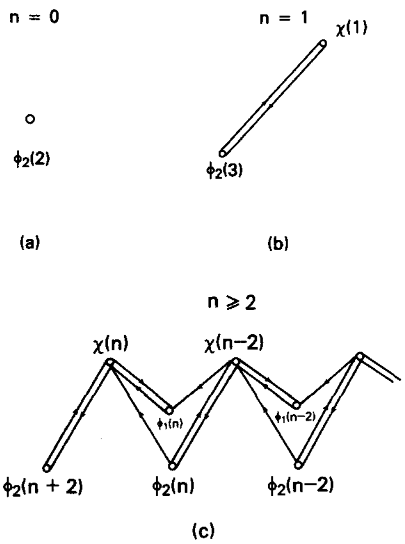


FIG. 1. Diagrammatic representations of the multiplets and the reduced matrix elements. The circle O represents a state. The line, with an arrow sign, represents a nonzero reduced matrix element from one state to another.

$$(2n+1)(2\sigma-1)]^{1/2},$$

$$\langle\psi_2(\sigma)|\,|T|\,|\chi(\sigma)\rangle = \langle\phi_2(\sigma)|\,|T|\,|\chi(\sigma)\rangle = 0.$$

The reduced matrix elements between $|\phi_1(\sigma)\rangle$, $|\phi_2(\sigma)\rangle$, and $|n\,\sigma-2\,\sigma-3\,m\,1\,\mu\rangle = |\chi(\sigma-2)\rangle$ are

$$\begin{aligned}
&\langle\phi_1(\sigma)|\,|T|\,|\chi(\sigma-2)\rangle \\
&\quad = 2[(2\sigma-4)(n-\sigma+2)(n+\sigma-1)/ \\
&\qquad (2\sigma-3)(2n+1)]^{1/2}, \\
&\langle\chi(\sigma-2)|\,|T|\,|\phi(\sigma)\rangle = 0, \\
&\langle\phi_2(\sigma)|\,|T|\,|\chi(\sigma-2)\rangle \\
&\quad = [(2\sigma-3)(2\sigma-4)/(2n+1)]^{1/2}, \\
&\langle\chi(\sigma-2)|\,|T|\,|\phi_2(\sigma)\rangle \\
&\quad = -[(\sigma-2)(2n+1)/(2\sigma-5)]^{1/2}.
\end{aligned} \tag{6.10}$$

From (6.9) and (6.10) we obtain Fig. 1.

From Fig. 1 we see the multiplets $\phi_2(n+2)$, $\chi(n)$, $\phi_1(n)$ form an invariant subspace; $\phi_2(n+2)$, $\chi(n)$, $\phi_1(n)$, $\phi_2(n)$, $\chi(n-2)$, $\phi_1(n-2)$ form another invariant subspace; and so on. Thus the $[n\}$ representation of U(3/2) is a reducible but not completely reducible representation of OSp(3/2).

VII. CONCLUSIONS

The use of dynamical supersymmetry as a means to study the spectroscopy of nuclear and atomic[9] structure is now widespread. In this paper we have studied the Fock wave functions of one such chain [Eq. (1.2)]. There are still some rather serious difficulties of applying such a symmetry chain to nuclear structure physics; one such difficulty is that it will lead to baryonic nonconservation. This is clearly evident in the wave functions of Table I which show clearly that they contain an admixture of states with different baryon number (i.e., nucleus with different mass number). How detrimental this will be for the study of nuclear structure remains to be seen. One should be reminded, however, that baryonic nonconservation is by no means a nontolerated problem in nuclear physics (i.e., BCS). The question here is, of course, how serious are the baryonic nonconserving components in this theory. Such studies are currently being pursued.

ACKNOWLEDGMENTS

We are grateful to Dr. A. B. Balantekin for several useful and illuminating discussions. One of us (MZ) would like to thank the Nuclear Theory Group of Drexel University for their warm hospitality.

One of us (MZ) gratefully acknowledges the financial support of the United States National Science Foundation for the provision of a postdoctoral research fellowship. One of us (HZS) was supported in part by the National Science Foundation under Grant No. INT83-05938. This work was partly supported by the National Science Foundation under Grant No. PHY83-04368 and the Chinese Science Foundation.

[1]A. Arima and F. Iachello, Phys. Rev. Lett. **25**, 1069 (1974); Ann. Rev. Nucl. Sci. **31**, 75 (1981).

[2]F. Iachello, Phys. Rev. Lett. **44**, 772 (1980); A. B. Balantekin, I. Bars, R. Bijker, and F. Iachello, Phys. Rev. C **27**, 1761 (1983); H-Z. Sun, A. Frank, and P. van Isacker, Phys. Lett. B **124**, 275 (1983); H-Z. Sun, M. Vallieres, D. H. Feng, R. Gilmore, and R. Casten, Phys. Rev. C **29**, 352 (1984); Phys. Lett. B **135**, 339 (1984); R. Bijker, Ph. D. thesis, University of Groningen, 1984; D. H. Feng, Nucl. Phys. A **421** (1984).

[3]F. Iachello and O. Scholten, Phys. Rev. Lett. **44**, 679 (1979); O. Scholten, Ph. D. thesis, University of Groningen, 1980.

[4]H-Z. Sun and Q-Z. Han, Phys. Energ. For. Phys. Nucl. **6**, 317 (1982) (in Chinese).

[5]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146, 155 (1977).

[6]A. B. Balkantekin and I. Bars, J. Math. Phys. **23**, 4 (1982).

[7]P. D. Jarvis and H. S. Green, J. Math. Phys. **20**, 2115 (1979).

[8]Q-Z. Han and H-Z. Sun, Commun. Theor. Phys. **2**, 1137 (1983).

[9]A. Kostelecky and M. Nieto, Phys. Rev. Lett. **53**, 2285 (1984).

Supertwistor realization of SU(2,2/1)

C. P. Luehr and M. Rosenbaum
Centro de Estudios Nucleares, Universidad Nacional Autónoma de México, Circuito Exterior, C. U. 04510 México, D. F., Mexico

(Received 13 April 1984; accepted for publication 15 March 1985)

A space of supertwistors is introduced as a graded vector space by starting with ordinary twistor space multiplied with even Grassmann coefficients and combining it by an othogonal direct sum with a one-dimensional space multiplied with odd Grassmann coefficients. A treatment of the algebraic properties of supertwistor space and of Grassmann-linear transformations on this space leads to a systematic procedure for deriving the structure equations for the graded Lie algebra su(2,2/1).

I. INTRODUCTION

Shortly after supersymmetry was proposed independently by Volkov and Akulov[1] and Wess and Zumino[2] as a means of constructing action functionals which are invariant under the interchange of bosons and fermions, it was recognized that the gauge groups for supersymmetric theories were the graded Lie groups.

The mathematical structures of graded Lie algebras and graded Lie groups have been extensively discussed in the literature,[3] and Salam and Strathdee[4] made use of SU(2,2/1) for the purpose of introducing the concept of superspace and superfields.

Although both Penrose[5] and the authors of the paper cited above suggested that supersymmetry and twistors might be closely related, it was in the works of Daniel[6] and Ferber[7] that the precise connection between these two theories was further investigated.

The approach of Daniel towards establishing a connection between the spinorial charges of supersymmetry and twistors is based on the identification of the superconformal algebra as the graded Lie algebra of the generators of the inhomogeneous group of rotations in the eight-dimensional spinor space associated with the pseudo-Euclidean space $E(4,2)$. In this procedure, twistors appear as linear combinations of the four-component semispinors into which the spinorial charges split off. The spinorial charges, themselves, transform according to the eight-dimensional representation of SU(2,2).

Ferber's treatment, on the other hand, is based on an enlargement of the space of twistors via the introduction of additional Grassmann components. The resulting supertwistors transform, then, under a representation of SU(2,2/1) [or SU(2,2/N), in general, if internal symmetries are included].

Thus, while twistor variables in Penrose's theory[8] appear as basic ingredients from which composite space-time is constructed, the supertwistors of Ferber are formed by the space of all triples $(\lambda^A, \mu_{A'}, \xi^\alpha)$ from which a point (X, θ) in superspace is defined as the incidence of trajectories given by the equations for the conserved generators [cf. Eqs. (6), (9), and (10) in Ref. 7] for particular values of such a triple. In this formalism, the spinor components λ^A are taken to be ordinary complex numbers, and the components μ_A in conjugate spinor space are assumed to be even elements of the Grassmann algebra generated by the anticommuting components $\theta^{\alpha A}$ of N Weyl spinors ($\alpha = 1,2,...,N$). The anticommuting Lorentz scalars ξ^α are defined by $\xi^\alpha = \theta^\alpha_A \lambda^A$.

Moreover, by requiring invariance of a bilinear form[9] constructed on supertwistor space, Ferber exhibits the general form which superconformal infinitesimal transformations [generalized to SU(2,2/N)] must take. Note, however (as was also pointed out by Ferber), that if the λ^A are assumed to be ordinary complex numbers, then the space of triples $(\lambda^A, \mu_{A'}, \xi^\alpha)$ will not be invariant under action of the superconformal infinitesimal transformations. In the derivation below of explicit forms of the generators of the graded algebra, we will use a larger supertwistor space (in essence, both λ^A and μ_A, will be allowed to contain even elements of the Grassmann algebra) to remedy this problem.

The extension of the concept of supertwistors to curved fermionic twistor space has been studied by Lukierski,[10] who also derived differential realizations of the generators of the SU(2,2/1) algebra on supertwistor space.

Given the above summary of some of the published material pertinent to the subject, we now set out to describe our own objectives. In some previous work[11–13] we initiated a program with the purpose of combining the twistor formalism and fiber bundle techniques in order to be able to construct in a unified manner gauge theories incorporating both compact and noncompact group symmetries.

Our approach differs from the authors' investigations to which we referred before in the essential fact that we do not use twistors and supertwistors as basic ingredients to build up space-time and superspace, respectively. We assume, instead, the formal point of view of regarding twistor space in its strictly mathematical conception as the fundamental complex four-dimensional linear space representation of the group SU(2,2), which is (4–1) homomorphic to the conformal group via SU(2,2) $\rightarrow$ O(2,4) $\rightarrow$ C(3,1).

In analogy, supertwistors will be regarded as the fundamental complex five-dimensional [$(4+N)$-dimensional if internal symmetries are included] linear space representation of the graded group SU(2,2/1) [or SU(2,2/N) for extended supersymmetries].

The present two consecutive papers (see Ref. 14) are intended as a continuation of our program which will allow supergravity theories to be included in the formalism. Spe-

 0022-2488/85/071828-06$02.50

cifically, this paper has a threefold purpose: (i) to derive a formalism of supertwistors as a graded vector space which acts as a natural representation space of SU(2,2/1); (ii) to derive explicit linear realizations of the generators of the superconformal algebra expressed as anti-Hermitian supertwistors which, to our knowledge, have not been given in this form in the literature, and (iii) to provide the necessary material on the characteristic graded Lie group for the development, in the accompanying paper, of supergravity gauge theories based on supertwistor fiber bundles.

Thus, our second paper relies to a large extent on the expressions derived here.

The presentation is organized as follows: In Sec. II we review some of the basic twistor structures which have been extensively discussed in Refs. 5, 8, and 11. Supertwistors are then introduced by enlarging ordinary twistor space to a graded vector space given as an orthogonal direct sum of subspaces with even and odd Grassmann coefficients.[15] The structure of the subspace of odd degree in the Grassmann coefficients contains features which make our formalism somewhat different from other approaches appearing in the literature.

The essential aspects of supertwistor algebra are treated here following, as closely as possible, the formalism originally introduced by Penrose[5,8] for ordinary twistors, since that notation seems to be more widely accepted and understood than the abstract, index-free, formalism proposed in Ref. 11. We feel, however, that the axiomatic approach introduced in our above cited paper has some advantages, at least within the context of the theories for which we use twistors. Those readers interested in further exploring this possibility will find in the notational equivalences provided in Ref. 11 a useful tool for readily going over from one type of notation to the other.

In Sec. III we present a systematic procedure for deriving explicit supertwistor expressions for the generators of the graded Lie algebra for SU(2,2/1), in a form most directly applicable to the construction of the supertwistor fiber bundles on which our supergravity theories in the following paper are based. Finally, in Sec. IV we display the relations for the resulting graded algebra, and point out the procedure by means of which our results can be generalized in a straightforward manner to SU(2,2/N), in order to include the internal symmetry group SU(N).

II. SUPERTWISTORS

Twistor space $\mathscr{U} \equiv \mathscr{U}_{2,2}$ is a four-dimensional complex vector space with a Hermitian type inner product $\langle l|m\rangle$, antilinear in $l^\alpha \in \mathscr{U}$ and linear in $m^\alpha \in \mathscr{U}$, having the signature $(+\,+\,-\,-)$ (lowercase Greek indices run from 0 to 3). The dual twistor space $\mathscr{U}' \equiv \mathscr{U}'_{2,2}$ is the set of linear functionals on $\mathscr{U}$, such that each element $k_\alpha \in \mathscr{U}'$ acts on each element $l^\alpha \in \mathscr{U}$ to produce a complex number $k_\alpha l^\alpha$.

Twistors in $\mathscr{U}$ can be mapped onto twistors in $\mathscr{U}'$ by means of the conjugation operation $l^\alpha \in \mathscr{U} \rightarrow \bar{l}_\alpha \in \mathscr{U}'$, which is an antilinear map defined by

$$\bar{l}_\alpha m^\alpha = m^\alpha \bar{l}_\alpha = \langle l|m\rangle, \tag{2.1}$$

for all $m^\alpha \in \mathscr{U}$. From the Hermiticity of the inner product, it follows that

$$(\bar{l}_\alpha m^\alpha)^* = \langle l|m\rangle^* = \langle m|l\rangle = \bar{m}_\alpha l^\alpha, \tag{2.2}$$

where * denotes ordinary complex conjugation.

Define the conjugation operation on $\mathscr{U}'$, $k_\alpha \in \mathscr{U}' \rightarrow \bar{k}^\alpha \in \mathscr{U}$ to be the inverse of the conjugation operation on $\mathscr{U}$, i.e., $\bar{\bar{l}}^\alpha = l^\alpha$. Note that with this definition (2.2) implies

$$\bar{m}_\alpha \bar{k}^\alpha = (k_\alpha m^\alpha)^*. \tag{2.2'}$$

With these axioms, and the requirement on transformations that the inner product be invariant, $\mathscr{U}$ constitutes a fundamental representation space for the group SU(2,2).

In the following, we shall use the notation.

$$l^\alpha = (\omega^A, \pi_{A'})$$

to denote a twistor which corresponds to the pair of spinors $(\omega^A, \pi_{A'})$.

Similarly

$$\bar{l}_\alpha = (\bar{\pi}_A, \bar{\omega}^{A'})$$

will be used to denote the conjugate twistor.

We complete this brief summary of twistor space with some material that will be required later in our discussion. First note that from $\mathscr{U}_{2,2}$ we can construct various kinds of tensor spaces such as $\mathscr{U}^{\wedge 2}_{2,2}$, $\mathscr{U}^{\wedge 4}_{2,2}$, $\mathscr{U}'^{\wedge 2}_{2,2}$, $\mathscr{U}'^{\wedge 4}_{2,2}$, $\mathscr{U}_{2,2} \otimes \mathscr{U}'_{2,2}$, etc. In the one-dimensional subspace $\mathscr{U}^{\wedge 4}_{2,2}$, assume that a privileged element $\eta^{\alpha\beta\gamma\delta}$ is given which satisfies the normalization requirement

$$\bar{\eta}_{\alpha\beta\gamma\delta}\eta^{\alpha\beta\gamma\delta} = 4!. \tag{2.3}$$

With $\bar{\eta}_{\alpha\beta\gamma\delta}$ and $\eta^{\alpha\beta\gamma\delta}$ we can form duals of antisymmetic twistors by means of the following operations:

$$B^{[\gamma\delta]} \in \mathscr{U}^{\wedge 2} \rightarrow B_{[\alpha\beta]} \in \mathscr{U}'^{\wedge 2} : B_{[\alpha\beta]} = \tfrac{1}{2}\bar{\eta}_{\alpha\beta\gamma\delta}B^{[\gamma\delta]}, \tag{2.4}$$

$$C_{[\gamma\delta]} \in \mathscr{U}'^{\wedge 2} \rightarrow C^{[\alpha\beta]} \in \mathscr{U}^{\wedge 2} : C^{[\alpha\beta]} = \tfrac{1}{2}\eta^{\alpha\beta\gamma\delta}C_{[\gamma\delta]}. \tag{2.5}$$

Also, with the aid of $\bar{\eta}_{\alpha\beta\gamma\delta}$, we can define inner products in $\mathscr{U}^{\wedge 2}$ and $\mathscr{U}'^{\wedge 2}$ as

$$\tfrac{1}{2}A^{[\alpha\beta]}\bar{\eta}_{\alpha\beta\gamma\delta}B^{[\gamma\delta]} = A^{[\alpha\beta]}B_{[\alpha\beta]} = A_{[\alpha\beta]}B^{[\alpha\beta]} \tag{2.6}$$

and

$$\tfrac{1}{2}\bar{A}_{[\alpha\beta]}\eta^{\alpha\beta\gamma\delta}\bar{B}_{[\gamma\delta]} = \bar{A}^{[\alpha\beta]}\bar{B}_{[\alpha\beta]} = \bar{A}_{[\alpha\beta]}\bar{B}^{[\alpha\beta]}. \tag{2.7}$$

Skew-symmetric twistors $B^{[\alpha\beta]} = B^{\alpha\beta} \in \mathscr{U}^{\wedge 2}$, for which

$$B_{\alpha\beta}B^{\alpha\beta} = 0, \tag{2.8}$$

are called null, and skew-symmetric twistors of the form $B^{\alpha\beta} = l^\alpha m^\beta - m^\alpha l^\beta$ are called simple and satisfy the property

$$B^{\alpha\beta}B_{\beta\gamma} = 0. \tag{2.9}$$

(Note also that $B^{\alpha\beta}$ being simple implies $B^{\alpha\beta}$ is null, and vice versa.) Of special importance is the space $\mathscr{E}$ of real skew-symmetric twistors, which is a real subspace of $\mathscr{U}^{\wedge 2}$ with the inner product $A_{\alpha\beta}B^{\alpha\beta}$ as part of its structure, and is defined by

$$\mathscr{E} \equiv \mathscr{E}_{2,4} = \{\, P^{\alpha\beta} \,|\, P^{\alpha\beta} \in \mathscr{U}^{\wedge 2},\ \ \bar{P}^{\alpha\beta} = P^{\alpha\beta} \}. \tag{2.10}$$

The inner product in $\mathscr{E}$ has signature $(+\,+\,-\,-\,-\,-)$.

Null, real twistors form a null cone $\mathscr{N}$ in $\mathscr{E}$. Among these, we distinguish the infinity twistor $I^{\alpha\beta}$ which is simple, and the origin twistor $O^{\alpha\beta}$ which is simple and satisfies $I_{\alpha\beta}O^{\alpha\beta}=2$. The infinity twistor allows the definition of a five-dimensional hyperplane $\mathscr{K}=\{P^{\alpha\beta}|P^{\alpha\beta}\in\mathscr{E}, I_{\alpha\beta}P^{\alpha\beta}=2\}$ which intersects the null cone $\mathscr{N}$ to form a four-dimensional hypersurface $\mathscr{W}=\{P^{\alpha\beta}|P^{\alpha\beta}\in\mathscr{E}, P_{\alpha\beta}P^{\alpha\beta}=0;\ I_{\alpha\beta}P^{\alpha\beta}=2\}$, which corresponds to Minkowski space-time. Acting as a tangent vector space to this surface at $O^{\alpha\beta}$ we have the four-dimensional subspace, $\mathscr{F}=\{T^{\alpha\beta}|T^{\alpha\beta}\in\mathscr{E}, I_{\alpha\beta}T^{\alpha\beta}=0,\ \ O_{\alpha\beta}T^{\alpha\beta}=0\}$. The inner product in $\mathscr{F}$ is a Minkowski inner product with signature $(+---)$.

Recall now that the twistor space $\mathscr{U}_{2,2}$ can be decomposed into the direct sum[8,11]

$$\mathscr{U}_{2,2}=\mathscr{S}_I\oplus\mathscr{S}_O \tag{2.11}$$

(where $\mathscr{S}_I$ is the space of two-component Weyl spinors and $\mathscr{S}_O$ is the corresponding conjugate spinor space) by making use of the idempotents

$$(S_I)^\alpha{}_\beta\equiv I^{\alpha\gamma}O_{\beta\gamma}\leftrightarrow\epsilon^{AG}\epsilon_{BG}=\delta^A{}_B, \tag{2.12}$$

$$(S_O)^\alpha{}_\beta\equiv O^{\alpha\gamma}I_{\beta\gamma}\leftrightarrow\epsilon_{A'G'}\epsilon^{B'G'}=\delta^{B'}_{A'}. \tag{2.13}$$

In fact, since $(S_I)^\alpha{}_\beta$ and $(S_O)^\alpha{}_\beta$ are mutually orthogonal projection operators from $\mathscr{U}_{2,2}$ onto $\mathscr{S}_I$ and $\mathscr{S}_O$, respectively, we can define $\mathscr{S}_I$ and $\mathscr{S}_O$ by

$$S_I=\{(S_I)^\alpha{}_\beta l^B|l^\beta\in\mathscr{U}_{2,2}\}=\{\lambda^A\}, \tag{2.14}$$

$$\mathscr{S}_O=\{(S_O)^\alpha{}_\beta l^\beta|l^\beta\in\mathscr{U}_{2,2}\}=\{\pi_{A'}\}, \tag{2.15}$$

or, equivalently, by

$$\mathscr{S}_I=\{l^\alpha\in\mathscr{U}_{2,2}|(S_I)^\alpha{}_\beta l^\beta=l^\alpha\} \tag{2.14'}$$
$$=\{l^\alpha\in\mathscr{U}_{2,2}|(S_O)^\alpha{}_\beta l^\beta=0\},$$

$$\mathscr{S}_O=\{l^\alpha\in\mathscr{U}_{2,2}|(S_O)^\alpha{}_\beta l^\beta=l^\alpha\}$$
$$=\{l^\alpha\in\mathscr{U}_{2,2}|(S_I)^\alpha{}_\beta l^\beta=0\}. \tag{2.15'}$$

We now enlarge our twistor space $\mathscr{U}_{2,2}$ and introduce the concept of supertwistors by following an analogous procedure to that used by Ferber.[7] We define a supertwistor as an element of the graded vector space given by the orthogonal direct sum

$$\mathscr{V}_{3,2}=\mathscr{V}_{2,2}\oplus\mathscr{V}_1,$$

where

$$\mathscr{V}_{2,2}=\mathscr{G}_e\otimes\mathscr{U}_{2,2},\quad \mathscr{V}_1=\mathscr{G}_o\otimes\mathscr{W}_1, \tag{2.16}$$

and $\mathscr{G}_e$ and $\mathscr{G}_o$ are the even and odd subsets, respectively, of a real Grassmann algebra $\mathscr{G}$ of dimension 2^d generated by a d-dimensional real vector space K_d, where $d>1$ can be finite or infinite.

The space $\mathscr{W}_1$ is a one-dimensional complex vector space with its inner product being i times a Hermitian type inner product. So the inner product $\langle\eta|\xi\rangle$ for $\xi^1,\eta^1\in\mathscr{W}_1$ is antilinear in η^1, linear in ξ^1, with

$$\langle\xi|\eta\rangle^*=-\langle\eta|\xi\rangle, \tag{2.17}$$

and $\langle\eta|\eta\rangle$ is pure imaginary.

This type of inner product is an essential ingredient, which leads to the correct structure for $\mathscr{V}_{3,2}$ which will be invariant under SU(2,2/1).

We also have a conjugation operation $\xi^1\in\mathscr{W}_1\rightarrow\bar{\xi}_1\in\mathscr{W}'_1$ (where $\mathscr{W}'_1$ is the space dual to $\mathscr{W}_1$) as an antilinear map defined by

$$\bar{\xi}_1\eta^1=\langle\xi|\eta\rangle. \tag{2.18}$$

It follows from (2.17) and (2.18) that

$$(\bar{\xi}_1\eta^1)^*=\langle\xi|\eta\rangle^*=-\langle\eta|\xi\rangle=-\bar{\eta}_1\xi^1. \tag{2.19}$$

Now define the map $\lambda_1\in\mathscr{W}'_1\rightarrow\bar{\lambda}^1\in\mathscr{W}_1$ to be minus the inverse of the conjugation map from $\mathscr{W}_1$ to $\mathscr{W}'_1$. Thus

$$\bar{\bar{\xi}}^1=-\xi^1, \tag{2.20}$$

and, by Eq. (2.19), we have

$$\bar{\eta}_1\bar{\lambda}^1=(\lambda_1\eta^1)^*. \tag{2.21}$$

Furthermore, by virtue of (2.17) and (2.18), we can take as a normalized basis for $\mathscr{W}_1$ any element $\tau^1\in\mathscr{W}_1$ satisfying

$$\langle\tau|\tau\rangle=\bar{\tau}_1\tau^1=i. \tag{2.22}$$

The spaces $\mathscr{U}_{2,2}$ and $\mathscr{W}_1$ discussed above can be viewed as subspaces of a larger space $\mathscr{U}_{3,2}$ given by $\mathscr{U}_{3,2}=\mathscr{U}_{2,2}\oplus\mathscr{W}_1$. In this sense we make the identifications

$$\mathscr{U}_{2,2}\equiv\mathscr{U}_{2,2}\oplus\{0\},\quad \mathscr{W}_1\equiv\{0\}\oplus\mathscr{W}_1. \tag{2.23}$$

Prior to introducing Grassmann coefficients in order to arrive at the gradation given in (2.16), it will prove convenient to consider some operations in $\mathscr{U}_{3,2}$, since generalization of the results to $\mathscr{V}_{3,2}$ can be achieved readily by extending linearly in the real Grassmann coefficients.

Thus we first represent an element $L^\Sigma\in\mathscr{U}_{3,2}$ by $L^\Sigma=(l^\sigma,\xi^1)$, where $l^\sigma\in\mathscr{U}_{2,2}$ and $\xi^1\in\mathscr{W}_1$ (capital Greek indices will range from 0 to 4). In particular, twistors in the subspace $\mathscr{U}_{2,2}$ will be written as $l^\Sigma\equiv(l^\sigma,0)$, and elements in the subspace $\mathscr{W}_1$ will be written as $\xi^\Sigma\equiv(0,\xi^1)$. We can also write $L^\Sigma=l^\Sigma+\xi^\Sigma$.

An element $K_\Sigma\in\mathscr{U}'_{3,2}=\mathscr{U}'_{2,2}\oplus\mathscr{W}'_1$, where $\mathscr{U}'_{3,2}$ is the space dual to $\mathscr{U}_{3,2}$, will be represented by $K_\Sigma=(k_\sigma,\lambda_1)$ where $k_\sigma\in\mathscr{U}'_{2,2}$ and $\lambda_1\in\mathscr{W}'_1$. By definition, the action of K_Σ on L^Σ is given by

$$K_\Sigma L^\Sigma=k_\sigma l^\sigma+\lambda_1\xi^1. \tag{2.24}$$

We define the inner product of $L^\Sigma=(l^\sigma,\xi^1)$ and $M^\Sigma=(m^\sigma,\eta^1)$ in $\mathscr{U}_{3,2}$ by

$$\langle L|M\rangle=\langle l|m\rangle+\langle\xi|\eta\rangle. \tag{2.25}$$

Note that by (2.2) and (2.17)

$$\langle L|M\rangle^*=\langle m|l\rangle-\langle\eta|\xi\rangle\neq\langle M|L\rangle. \tag{2.26}$$

Making use of this inner product, we can now define a conjugation operation

$$L^\Sigma\in\mathscr{U}_{3,2}\rightarrow\bar{L}_\Sigma\in\mathscr{U}'_{3,2}$$

by $\qquad$ (2.27)

$$\bar{L}_\Sigma M^\Sigma=\langle L|M\rangle,$$

which together with (2.25) implies

$$\bar{L}_\Sigma=(\bar{l}_\sigma,\bar{\xi}_1). \tag{2.28}$$

It follows from (2.26) that

$$(\bar{L}_\Sigma M^\Sigma)^*=\bar{M}_\Sigma G^\Sigma{}_\Gamma L^\Gamma, \tag{2.29}$$

where the operation of

$$G^\Sigma{}_\Gamma=\delta^\Sigma{}_\Gamma-2\delta^\Sigma_4\delta^4_\Gamma \tag{2.30}$$

on $L^\Sigma = (l^\sigma, \xi^1)$ just reverses the sign of the last component of L^Σ, i.e.,

$$G^\Sigma{}_\Gamma L^\Gamma = (l^\sigma, -\xi^1). \tag{2.31}$$

For $\mathscr{U}'_{3,2}$ we define the conjugation operation $K_\Sigma \in \mathscr{U}'_{3,2} \to \bar{K}^\Sigma \in \mathscr{U}_{3,2}$ by the equation

$$\bar{K}^\Sigma \bar{M}_\Sigma = (K_\Sigma M^\Sigma)^*, \tag{2.32}$$

for all $M^\Sigma \in \mathscr{U}_{3,2}$. It follows from (2.32) that for $K_\Sigma = (k_\sigma, \lambda_1) \in \mathscr{U}'_{3,2}$ and for $L^\Sigma \in \mathscr{U}_{3,2}$ we have

$$\bar{K}^\Sigma = (\bar{k}^\sigma, \bar{\lambda}^1) \tag{2.33}$$

and

$$G^\Sigma{}_\Gamma \bar{\bar{L}}^\Gamma = L^\Sigma. \tag{2.34}$$

It is important to note from (2.34) that the conjugation operation is not an involution.

Multiplying elements of the subspaces $\mathscr{U}_{2,2}$ and $\mathscr{W}_1$ of $\mathscr{U}_{3,2}$ as identified in (2.23) by even and odd Grassmann numbers, respectively, leads directly to the space $\mathscr{V}_{3,2}$. Thus for a supertwistor $L^\Sigma = l^\Sigma + \xi^\Sigma = (l^\sigma, \xi^1) \in \mathscr{V}_{3,2}$, l^σ is a twistor with complexified even Grassmann components while ξ^1 will be a complexified odd Grassmann component. (From hereon we shall use capital Latin letters with capital Greek indices to denote supertwistors.)

The inner product $\langle L | M \rangle$ for $L^\Sigma, M^\Sigma \in \mathscr{V}_{3,2}$ follows directly by extending (2.25) bilinearly in real Grassmann coefficients. In particular, because of the anticommutivity of the odd Grassmann numbers, we will now have

$$\langle L | M \rangle^* = \langle M | L \rangle, \tag{2.35}$$

instead of (2.26).

Note that in our formalism the complex conjugate of a complexified Grassmann quantity is obtained by replacing i by $-i$ wherever it appears. Thus, in the case of products of complex Grassmann numbers the order of the factors is preserved under the operation of complex conjugation.

A similar grading procedure applied to the dual spaces leads to $\mathscr{V}'_{2,2} = \mathscr{G}_e \otimes \mathscr{U}'_{2,2}$, $\mathscr{V}'_1 = \mathscr{G}_o \otimes \mathscr{W}'_1$ and $\mathscr{V}'_{3,2} = \mathscr{V}'_{2,2} \oplus \mathscr{V}'_1$. Thus for an element $K_\Sigma = k_\Sigma + \lambda_\Sigma = (k_\sigma, \lambda_1) \in \mathscr{V}'_{3,2}$, k_σ is a dual twistor with complexified even Grassmann components, while λ_1 will be a complexified odd Grassmann component. The action of K_Σ on a supertwistor L^Σ is given by

$$K_\Sigma L^\Sigma = k_\sigma l^\sigma + \lambda_1 \xi^1. \tag{2.36}$$

It is important to note that

$$L^\Sigma K_\Sigma = K_\Sigma G^\Sigma{}_\Gamma L^\Gamma, \tag{2.37}$$

where $G^\Sigma{}_\Gamma$ was defined in (2.30).

We can now directly define the conjugation operations $L^\Sigma \in \mathscr{V}_{3,2} \to \bar{L}_\Sigma \in \mathscr{V}'_{3,2}$ and $K_\Sigma \in \mathscr{V}'_{3,2} \to \bar{K}^\Sigma \in \mathscr{V}_{3,2}$, from the previously introduced conjugation operations in $\mathscr{U}_{3,2}$ and $\mathscr{U}'_{3,2}$, respectively [cf. Eqs. (2.27) and (2.32)], as an extension linear in real Grassmann coefficients. Thus we obtain

$$\bar{L}_\Sigma M^\Sigma = \langle L | M \rangle, \tag{2.38}$$

$$\bar{K}^\Sigma \bar{M}_\Sigma = (K_\Sigma M^\Sigma)^*, \tag{2.39}$$

and (2.35) together with (2.38) implies

$$(\bar{L}_\Sigma M^\Sigma)^* = \bar{M}_\Sigma L^\Sigma. \tag{2.40}$$

Making use of (2.40), (2.39), and (2.37) yields

$$\bar{M}_\Sigma L^\Sigma = (\bar{L}_\Sigma M^\Sigma)^* = \bar{\bar{L}}^\Sigma \bar{M}_\Sigma = \bar{M}_\Sigma G^\Sigma{}_\Gamma \bar{\bar{L}}^\Gamma$$

for arbitrary M^Σ. Therefore

$$L^\Sigma = G^\Sigma{}_\Gamma \bar{\bar{L}}^\Gamma. \tag{2.41}$$

Note that (2.41) is the same as the result in (2.34) for the nongraded space $\mathscr{U}_{3,2}$. In fact (2.41) is just the linear extension by real Grassmann coefficients of (2.34).

In the next section we shall be considering the space $\mathscr{A}$ of transformations $\mathscr{V}_{3,2} \to \mathscr{V}_{3,2}$ which are linear in complex even Grassmann coefficients. An arbitrary transformation $U^\Sigma{}_\Gamma \in \mathscr{A}$ can be expressed in general as complex linear combinations of elementary maps of the following forms:

$$A^\Sigma{}_\Gamma = \rho_e l^\Sigma \bar{m}_\Gamma \ :\mathscr{V}_{2,2} \to \mathscr{V}_{2,2}, \tag{2.42a}$$

$$B^\Sigma{}_\Gamma = \rho_e \tau^\Sigma \bar{\tau}_\Gamma \ :\mathscr{V}_1 \to \mathscr{V}_1, \tag{2.42b}$$

$$C^\Sigma{}_\Gamma = \rho_o l^\Sigma \bar{\tau}_\Gamma \ :\mathscr{V}_1 \to \mathscr{V}_{2,2}, \tag{2.42c}$$

$$D^\Sigma{}_\Gamma = \rho_o \tau^\Sigma \bar{m}_\Gamma \ :\mathscr{V}_{2,2} \to \mathscr{V}_1, \tag{2.42d}$$

where $l^\Sigma, m^\Sigma \in \mathscr{U}_{2,2}; \rho_e, \rho_o$ are any real even and odd Grassmann numbers, respectively, and τ^Σ is our chosen normalized basis in $\mathscr{W}_1$ $(\bar{\tau}_\Sigma \tau^\Sigma = i)$. The Hermitian conjugate $(U^\dagger)^\Sigma{}_\Gamma$ of $U^\Sigma{}_\Gamma$ is defined by

$$Y^\Sigma = U^\Sigma{}_\Gamma Z^\Gamma, \quad \bar{Y}_\Sigma = \bar{Z}_\Gamma (U^\dagger)^\Gamma{}_\Sigma, \tag{2.43}$$

for all $Z^\Gamma \in \mathscr{V}_{3,2}$. Since in the discussion below we will make extensive use of this operation, we write down the Hermitian conjugate of the elementary forms given by Eqs. (2.42) as

$$(A^\dagger)^\Sigma{}_\Gamma = \rho_e m^\Sigma \bar{l}_\Gamma \ :\mathscr{V}_{2,2} \to \mathscr{V}_{2,2}, \tag{2.44a}$$

$$(B^\dagger)^\Sigma{}_\Gamma = -\rho_e \tau^\Sigma \bar{\tau}_\Gamma \ :\mathscr{V}_1 \to \mathscr{V}_1, \tag{2.44b}$$

$$(C^\dagger)^\Sigma{}_\Gamma = \rho_o \tau^\Sigma \bar{l}_\Gamma \ :\mathscr{V}_{2,2} \to \mathscr{V}_1, \tag{2.44c}$$

$$(D^\dagger)^\Sigma{}_\Gamma = \rho_o m^\Sigma \bar{\tau}_\Gamma \ :\mathscr{V}_1 \to \mathscr{V}_{2,2}. \tag{2.44d}$$

These expressions follow readily from (2.43) and the properties of the various operations on twistors and supertwistors defined previously. For example, (2.44c) is derived as follows:

$$Y^\Sigma = C^\Sigma{}_\Gamma Z^\Gamma = \rho_o l^\Sigma \bar{\tau}_\Gamma Z^\Gamma,$$
$$\bar{Y}_\Sigma = \rho_o \bar{l}_\Sigma (\bar{\tau}_\Gamma Z^\Gamma)^* = -\rho_o \bar{l}_\Sigma \bar{Z}_\Gamma \tau^\Gamma$$
$$= \bar{Z}_\Gamma \rho_o \tau^\Gamma \bar{l}_\Sigma = \bar{Z}_\Gamma (C^\dagger)^\Gamma{}_\Sigma,$$

where use has been made of (2.19), which also applies when η^1 in that equation is an odd complex Grassmann number, and the anticommutativity of odd Grassmann numbers was also used.

Finally, it can be verified from these results that for $U^\Sigma{}_\Gamma \in \mathscr{A}$ we have

$$(U^{\dagger\dagger})^\Sigma{}_\Gamma = U^\Sigma{}_\Gamma.$$

III. THE GENERATORS OF SU(2,2/1)

The set of transformations which leave the inner product $\langle Z | W \rangle = \bar{Z}_\Sigma W^\Sigma$ in $\mathscr{V}_{3,2}$ invariant, and have graded determinant equal to one constitute a realization of the graded group SU(2,2/1). To obtain the generators of this group, first let

$$U^\Sigma{}_\Gamma = \mathbf{1}^\Sigma{}_\Gamma + \epsilon S^\Sigma{}_\Gamma \tag{3.1}$$

denote an infinitesimal transformation. Here ϵ is a real infinitesimal parameter, the element $1^{\Sigma}{}_{\Gamma} \in (\mathscr{U}_{2,2} \otimes \mathscr{U}'_{2,2}) \oplus (\mathscr{W}_1 \otimes \mathscr{W}'_1)$ acts as the identity operator on $\mathscr{V}_{3,2}$, i.e., $1^{\Sigma}{}_{\Gamma} Z^{\Gamma} = Z^{\Sigma}$ for all $Z^{\Sigma} \in \mathscr{V}_{3,2}$, and $S^{\Sigma}{}_{\Gamma}$ is an infinitesimal generator of SU(2,2/1).

The most general form of $U^{\Sigma}{}_{\Gamma}$ (and therefore also $S^{\Sigma}{}_{\Gamma}$) has to be in the space $(\mathscr{V}_{3,2} \otimes \mathscr{V}'_{3,2}) \oplus (\mathscr{W}_1 \otimes \mathscr{W}'_1)$. In order to further restrict $S^{\Sigma}{}_{\Gamma}$, note that from the invariance of the inner product in $\mathscr{V}_{3,2}$ and from (3.1),

$$(S^{\dagger})^{\Sigma}{}_{\Gamma} = -S^{\Sigma}{}_{\Gamma}. \tag{3.2}$$

This expression is the generalization of the anti-Hermitian property to supertwistors.

Moreover, since we are considering transformations with graded determinant equal to 1 we have, from the property

$$\ln[\operatorname{gdet}(U^{\Sigma}{}_{\Gamma})] = \operatorname{gtr}(\ln U^{\Sigma}{}_{\Gamma}), \tag{3.3}$$

that

$$\operatorname{gtr}(S^{\Sigma}{}_{\Gamma}) = 0. \tag{3.4}$$

The definition of the graded trace operation is equivalent to the following:

$$\operatorname{gtr}(S^{\Sigma}{}_{\Gamma}) = S^{\Sigma}{}_{\Gamma} G^{\Gamma}{}_{\Sigma}, \tag{3.5}$$

where $G^{\Gamma}{}_{\Sigma}$ was defined in (2.30).

Note that this definition leads directly to the usual cyclicity property

$$\operatorname{gtr}(M^{\Sigma}{}_{\Gamma} N^{\Gamma}{}_{\Omega}) = \operatorname{gtr}(N^{\Sigma}{}_{\Gamma} M^{\Gamma}{}_{\Omega}), \tag{3.6}$$

for $M^{\Sigma}{}_{\Gamma}$, $N^{\Gamma}{}_{\Omega} \in \mathscr{A}$ as required for graded transformations.

Conditions (3.2) and (3.4) are sufficient to determine all possible forms of the infinitesimal generators $S^{\Sigma}{}_{\Gamma}$. In fact, as we pointed out previously, $S^{\Sigma}{}_{\Gamma}$ will be given as linear combinations of the elementary forms (2.42) such that

$$S^{\Sigma}{}_{\Gamma} \in \mathscr{A}_{\text{aH}} \subset \mathscr{A},$$

where

$$\mathscr{A}_{\text{aH}} = (\mathscr{V}_{3,2} \otimes_{\text{aH}} \mathscr{V}'_{3,2}) \oplus (\mathscr{W}_1 \otimes_{\text{aH}} \mathscr{W}'_1), \tag{3.7}$$

and the subscript "aH" indicates the restriction to the subspace of anti-Hermitian elements, using (3.2) as the definition of anti-Hermiticity.

First consider anti-Hermitian transformations $S^{\Sigma}{}_{\Gamma}$ constructed from (2.42c) and (2.44c). We have

$$S^{\Sigma}{}_{\Gamma} = C^{\Sigma}{}_{\Gamma} - (C^{\dagger})^{\Sigma}{}_{\Gamma} = \rho_o(l^{\Sigma}\bar{\tau}_{\Gamma} - \tau^{\Sigma}\bar{l}_{\Gamma}). \tag{3.8}$$

The condition (3.4) is automatically satisfied in (3.8) because of the orthogonality of $\mathscr{W}_1$ and $\mathscr{U}_{2,2}$. Thus, $S^{\Sigma}{}_{\Gamma}$ will assume the following two possible forms:

$$\text{(i)}\quad (S_{ST})^{\Sigma}{}_{\Gamma} = \rho_o(h^{\Sigma}\bar{\tau}_{\Gamma} - \tau^{\Sigma}\bar{h}_{\Gamma}), \tag{3.9}$$

where $h^{\Sigma} = (h^{\sigma},0) = (\lambda^{A},0,0) \in \mathscr{S}_I \subset \mathscr{U}_{2,2}$; and

$$\text{(ii)}\quad (S_{SK})^{\Sigma}{}_{\Gamma} = \rho_o(p^{\Sigma}\bar{\tau}_{\Gamma} - \tau^{\Sigma}\bar{p}_{\Gamma}),$$

where $p^{\Sigma} = (p^{\sigma},0) = (0,\pi_{A'},0) \in \mathscr{S}_o \subset \mathscr{U}_{2,2}$. (3.10)

The infinitesimal generators $(S_{ST})^{\Sigma}{}_{\Gamma}$ and $(S_{SK})^{\Sigma}{}_{\Gamma}$ are the generators of supertranslations and superconformal boosts, respectively.

Note that the same results will follow from (2.42d) and (2.44d).

Next we analyze the anti-Hermitian transformation $S^{\Sigma}{}_{\Gamma}$ constructed from (2.42a) and (2.44a). We obtain

$$S^{\Sigma}{}_{\Gamma} = \rho_e(l^{\Sigma}\bar{m}_{\Gamma} - m^{\Sigma}\bar{l}_{\Gamma}). \tag{3.11}$$

There exist four possibilities, depending on particular choices of l^{Σ} and m^{Σ}.

One of the choices is

$$\text{(iii)}\quad (S_T)^{\Sigma}{}_{\Gamma} = \rho_e(h^{\Sigma}\bar{k}_{\Gamma} - k^{\Sigma}\bar{h}_{\Gamma}) \tag{3.12}$$

$$\text{for } h^{\Sigma} = (h^{\sigma},0),\quad k^{\Sigma} = (k^{\sigma},0) \in \mathscr{S}_I.$$

Equation (3.12) satisfies automatically the condition (3.4), since elements in $\mathscr{S}_I$ are orthogonal to each other. The transformation $(S_T)^{\Sigma}{}_{\Gamma}$ maps a complex two-dimensional twistor in $\mathscr{S}_O$ onto $\mathscr{S}_I$, and is the infinitesimal generator of translations.

If we now take as a choice of (3.11) the operator

$$\text{(iv)}\quad (S_K)^{\Sigma}{}_{\Gamma} = \rho_e(p^{\Sigma}\bar{q}_{\Gamma} - q^{\Sigma}\bar{p}_{\Gamma}),$$

$$\text{for } p^{\Sigma} = (p^{\sigma},0),\ q^{\Sigma} = (q^{\sigma},0) \in \mathscr{S}_O, \tag{3.13}$$

then once again (3.4) will be automatically satisfied. However, $(S_K)^{\Sigma}{}_{\Gamma}$ maps a complex two-dimensional twistor in $\mathscr{S}_I$ onto $\mathscr{S}_O$, and is therefore identified with the infinitesimal generator of conformal boosts.

Another choice, obtained by combining transformations of the form (3.11) is

$$\text{(v)}\quad (S_D)^{\Sigma}{}_{\Gamma} = \rho_e[(S_I)^{\Sigma}{}_{\Gamma} - (S_O)^{\Sigma}{}_{\Gamma}], \tag{3.14}$$

where

$$(S_I)^{\Sigma}{}_{\Gamma} = \left[\begin{array}{c|c} (S_I)^{\sigma}{}_{\gamma} & 0 \\ \hline 0 & 0 \end{array}\right], \quad (S_O)^{\Sigma}{}_{\Gamma} = \left[\begin{array}{c|c} (S_O)^{\sigma}{}_{\gamma} & 0 \\ \hline 0 & 0 \end{array}\right], \tag{3.15}$$

and $(S_I)^{\sigma}{}_{\gamma}$ and $(S_O)^{\sigma}{}_{\gamma}$ are given in (2.12) and (2.13). Note that $(S_D)^{\Sigma}{}_{\Gamma}$ has zero trace, as required by (3.4). This infinitesimal transformation is the generator of dilations.

The last possibility for transformations of the form (3.11) corresponds to

$$\text{(vi)}\quad (S_L)^{\Sigma}{}_{\Gamma} = \rho_e(h^{\Sigma}\bar{p}_{\Gamma} - p^{\Sigma}\bar{h}_{\Gamma}), \tag{3.16}$$

for $h^{\Sigma} = (h^{\sigma},0) \in \mathscr{S}_I$, $p^{\Sigma} = (p^{\sigma},0) \in \mathscr{S}_O$ with $h^{\Sigma}\bar{p}_{\Sigma} = 0$. It can be shown that infinitesimal transformations with which (3.16) is associated leave the origin and infinity twistors invariant. Hence the elements $(S_L)^{\Sigma}{}_{\Gamma}$ are the infinitesimal generators of Lorentz transformations. These generators can also be characterized by the conditions

$$\operatorname{gtr}[(S_I)^{\Sigma}{}_{\Gamma}(S_L)^{\Gamma}{}_{\Lambda}] = \operatorname{gtr}[(S_O)^{\Sigma}{}_{\Gamma}(S_L)^{\Gamma}{}_{\Lambda}] = 0. \tag{3.17}$$

We are finally lead to consider anti-Hermitian transformations constructed from (2.42b) and (2.44b). This will be of the form

$$(S_{SS})^{\Sigma}{}_{\Gamma} = \tfrac{1}{2}[B^{\Sigma}{}_{\Gamma} - (B^{\dagger})^{\Sigma}{}_{\Gamma}] = \rho_e\tau^{\Sigma}\bar{\tau}_{\Gamma}. \tag{3.18}$$

Note however that (3.18) is not traceless and, therefore, $(S_{SS})^{\Sigma}{}_{\Gamma}$ will not be an infinitesimal generator of our group. There is, however, one last possibility which consists of combining the nontraceless element $(S_{SS})^{\Sigma}{}_{\Gamma}$ with nontraceless anti-Hermitian elements obtained from (2.42a) and (2.44a) in order to get an infinitesimal generator. Clearly

$$\text{(vii)}\quad (S_{SP})^{\Sigma}{}_{\Gamma} = i\rho_e[(S_I)^{\Sigma}{}_{\Gamma} + (S_O)^{\Sigma}{}_{\Gamma} - 4i\tau^{\Sigma}\bar{\tau}_{\Gamma}] \tag{3.19}$$

is both anti-Hermitian and traceless. It is in fact the infinitesimal generator of phase transformations.

IV. GRADED ALGEBRA

The infinitesimal generators given in the equations marked (i)–(vii) in the previous section can be used to obtain in a straightforward manner the graded algebra su(2,2/1). Thus we can write (with tensor indices deleted) the following Lie brackets[16]:

$$[S_{1e},S_{2e}] = [\xi^i_e E_i, \eta^j_e E_j]$$
$$= \xi^i_e \eta^j_e [E_i,E_j] = \xi^i_e \eta^j_e f_{ij}{}^k E_k, \tag{4.1}$$
$$[S_{1e},S_{2o}] = [\xi^i_e E_i, \rho^s_o Q_s]$$
$$= \xi^i_e \rho^s_o [E_i,Q_s] = \xi^i_e \rho^s_o f_{is}{}^t Q_t, \tag{4.2}$$
$$[S_{1o},S_{2o}] = [\sigma^s_o Q_s, \rho^t_o Q_t]$$
$$= \sigma^s_o \rho^t_o \{Q_s,Q_t\} = \sigma^s_o \rho^t_o f_{st}{}^k E_k, \tag{4.3}$$

where $S_{1e} = \xi^i_e E_i$ and $S_{2e} = \eta^j_e E_j$ are any two different linear combinations of the infinitesimal generators with even Grassmann numbers that we derived above, the E_i represent the even generators of the graded algebra, and $S_{1o} = \sigma^s_o Q_s$, $S_{2o} = \rho^t_o Q_t$ are any two different linear combinations of our infinitesimal generators with odd Grassmann numbers while Q_s denotes the odd generators of the graded algebra. The quantities $f_{ij}{}^k$, $f_{is}{}^t$, and $f_{st}{}^k$ are the structure constants of the graded algebra.

As particular case of (4.3) take two infinitesimal generators of the form given in Eq. (3.9), i.e.,

$$(S_{1o})^\Sigma{}_\Gamma = \sigma_o (Q_1)^\Sigma{}_\Gamma, \quad \text{where } (Q_1)^\Sigma{}_\Gamma = h^\Sigma \bar{\tau}_\Gamma - \tau^\Sigma \bar{h}_\Gamma,$$
$$h^\Sigma = (h^\sigma,0) \in \mathscr{S}_I, \tag{4.4}$$
$$(S_{2o})^\Sigma{}_\Gamma = \rho_o (Q_2)^\Sigma{}_\Gamma, \quad \text{where } (Q_2)^\Sigma{}_\Gamma = k^\Sigma \bar{\tau}_\Gamma - \tau^\Sigma \bar{k}_\Gamma,$$
$$k^\Sigma = (h^\sigma,0) \in \mathscr{S}_I.$$

The Lie bracket is then given by

$$[S_{1o},S_{2o}]^\Sigma{}_\Gamma = [\sigma_o Q_1, \rho_o Q_2]^\Sigma{}_\Gamma$$
$$= (\sigma_o Q_1)^\Sigma{}_\Lambda (\rho_o Q_2)^\Lambda{}_\Gamma - (\rho_o Q_2)^\Sigma{}_\Lambda (\sigma_o Q_1)^\Lambda{}_\Gamma$$
$$= \sigma_o \rho_o [(Q_1)^\Sigma{}_\Lambda (Q_2)^\Lambda{}_\Gamma + (Q_2)^\Sigma{}_\Lambda (Q_1)^\Lambda{}_\Gamma]$$
$$= \sigma_o \rho_o \{Q_1,Q_2\}^\Sigma{}_\Gamma$$
$$= -\sigma_o \rho_o [(ih^\Sigma)\bar{k}_\Gamma - k^\Sigma (\overline{ih})_\Gamma]. \tag{4.5}$$

Note that the last expression in the above equation is of the same form as that given in Eq. (3.12), and thus it represents an infinitesimal generator of a translation. This is one of the well-known results of the graded algebra su(2,2/1), which allows the coupling of fermions and bosons in supersymmetry theories such as the one we present in the accompanying paper.

As a final remark we point out that our formalism can be extended to encompass the group SU(2,2/N) by replacing the one-dimensional space $\mathscr{W}_1$ by an N-dimensional space $\mathscr{W}_N$ with an inner product being i times a positive definite Hermitian form.

ACKNOWLEDGMENTS

This work was supported in part by International Scientific Program Grants, National Science Foundation No. OIP75-09783A01 and Consejo Nacional de Ciencia y Tecnología No. PCAIEUA-790316.

[1] D. V. Volkov and V. P. Akulov, Phys. Lett. B **46**, 109 (1973).
[2] J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974).
[3] See, e.g., F. A. Berezin and G. I. Kac, Math. USSR Sb. **11**, 311 (1970); L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod Phys. **47**, 573 (1975); V. G. Kac, Commun. Math. Phys. **53**, 31 (1977); M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1626, 1640 (1976); F. Gürsey and L. Marchildon, J. Math. Phys. **19**, 942 (1978); and references therein.
[4] A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974).
[5] R. Penrose, in *Quantum Gravity*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Clarendon, Oxford, 1975), p. 293.
[6] M. Daniel, Phys. Lett. B **65**, 246 (1976); cf. also M. Daniel and C. N. Ktorides, Nucl. Phys. B **115**, 313 (1976).
[7] A. Ferber, Nucl. Phys. B **132**, 55 (1978).
[8] R. Penrose, J. Math. Phys. **8**, 345 (1967).
[9] Supertwistors in the form of a five-component column, with four of the complex components being anticommuting numbers and the fifth component a complex even element of a Grassmann algebra, and invariant bilinear forms constructed from them were used by Gürsey and Marchildon (cf. Ref. 3) in their discussion on nonlinear realizations of the superconformal group on coset spaces; see also V. Rittenberg, in *Proceedings of the 1977 Tubingen Conference on Group Theoretic Methods in Physics* (Springer, New York, 1978).
[10] J. Lukierski, J. Math. Phys. **21**, 561 (1980).
[11] C. P. Luehr and M. Rosenbaum, J. Math. Phys. **23**, 1471 (1982).
[12] M. Rosenbaum, M. P. Ryan, and L. Urrutia, Phys. Rev. D **26**, 761 (1982).
[13] C. P. Luehr and M. Rosenbaum, J. Math. Phys. **25**, 380 (1984).
[14] C. P. Luehr and M. Rosenbaum, J. Math Phys. **26**, 1834 (1985).
[15] We have chosen the same statistics assignment as in Ref. 7, first because this choice does not actually violate the spin statistics theorem, and also because it is most convenient for the construction of the gauge theory in the following paper, since the subspace $\mathscr{V}_{2,2}$ of ordinary twistors with even Grassmann coefficients will constitute the building block for the bosonic gravitational part of the supergravity Lagrangian.
[16] See, e.g., M. Roček, in *Proceedings of the Nuffield Workshop* (Cambridge U.P., Cambridge, 1980).

Supertwistor fiber bundles as a formalism for supergravities

C. P. Luehr and M. Rosenbaum
Centro de Estudios Nucleares, Universidad Nacional Autónoma de México, Circuito Exterior, C. U. 04510 México, D. F., Mexico

(Received 13 April 1984; accepted for publication 15 March 1985)

A gauge theory of supergravity which allows one to obtain first-order Lagrangians that are locally gauge invariant by construction is presented. The formalism makes use of supertwistors as a representation space for the construction of a typical fiber of a vector bundle associated with a principal bundle, where the structural group is the super-Poincaré group. The approach proposed provides a means of resolving one of the central problems of gauge field theories of external symmetry groups, that is the satisfactory treatment of translations.

I. INTRODUCTION

In earlier papers[1] we have discussed some of the essential difficulties which result from attempting to apply the Utimaya[2] procedure, appropriate for the gauging of internal symmetry groups, to external groups. In particular we analyzed the work of Kibble,[3] which enlarged Utimaya's formalism to include inhomogeneous Lorentz transformations, and found that some apparently well-established results could not be properly justified by means of such an approach.[4]

The same type of difficulties occur in several of the articles appearing in the current literature which utilize the above-mentioned procedures as a starting point to obtain gauge theories for the supersymmetric groups.[5]

The purpose of the present paper is to show how the twistor bundle formalism, which we developed previously[6] in order to construct gauge theories for noncompact groups correctly, can be extended to supersymmetric groups in a way which leads one to unambiguous supergravity theories.

We believe that through our formalism one of the central problems of gauge field theories that have the Poincaré group as a characteristic subgroup, which is the gauging of the translations in a satisfactory manner, has been adequately resolved.[7]

The essential features of our theory are the use of supertwistors[8] obtained by means of enlarging ordinary twistor space to a graded vector space given as an orthogonal direct sum of elements with even and odd Grassmann coefficients [one adjoins a one-dimensional (N-dimensional) complex vector space with odd Grassmann coefficients for simple (extended) supersymmetries]. Making use of these structures we can generate the required supertwistor bundles by a procedure based on an extension of our results in Ref. 6. Another important feature of our formalism is that the base space of our fiber bundles is a four-dimensional manifold which has no *a priori* additional structure imposed, but which acquires the metric structure of space-time as a consequence of the formalism. Thus, even though supertwistors are closely related to superspace in that the former may be regarded as the basic ingredients from which composite superspace is constructed,[9] the fibers in our theory are not based on superspace and superfields. Rather our approach relies on the formal point of view of considering supertwistor space as the fundamental complex five-dimensional [$(4+N)$-dimensional if the internal symmetry group SU(N) is included] linear space representation of the graded group SU(2,2/1) [or SU(2,2/N) for extended supersymmetries]. This representation space serves to define a typical fiber from which our vector bundles are constructed.

Our procedure then leads to the calculation of a supertwistor curvature, in terms of which all possible Lagrangians permitted by the theory may be expressed as scalar functionals by means of appropriate contractions. These Lagrangians will be gauge invariant by construction; a feature of the theory which is made possible by the specific relations between the gravitino field and the Riemann and torsion tensors, imposed by the form of the supertwistor curvature.

It is interesting to note here that in the case of $N=0$ supertwistors, i.e., of ungraded standard twistor space, our formalism reduces to a gauge theory for ordinary gravitation (arrived at in a somewhat different, and perhaps simpler, manner than the one used in Ref. 6). This is the main reason for the particular choice of assignment of statistics in the construction of our supertwistors.

As a final remark, we point out that in this paper we have introduced the additional requirement that the null cone at infinity be part of the supertwistor structure. Thus we have broken superconformal invariance and retained only the group of super-Poincaré linear transformations which leave the infinity twistor invariant. Also, within this restriction we have only considered in detail the Lagrangian for simple supergravity, although this is only a particular case of the possible invariant scalars which are allowed by the theory.

The plan of the presentation is as follows: In Sec. II we give a very brief summary of the algebra of supertwistors which was developed in the immediately preceding paper (hereafter referred to as I). For details the reader is referred to that work. We also present in this section the essential arguments which lead, by incorporating these structures, to an extension of our formalism for a gauge theory of gravitation based on twistor bundles[6] to the construction of a gauge theory for the super-Poincaré group. The section also contains the basic aspects of supertwistor calculus which enable us to define supertwistor connections and arrive at a supertwistor curvature which is further related to the Riemann curvature, the torsion, and additional terms resulting from the gravitino field.

 0022-2488/85/071834-13$02.50

Section III is dedicated to the construction of the Lagrangian density for simple supergravity, even though, as noted previously, this is only a particular case of the scalars that may be obtained from the more general functional form of the permissible invariant scalar Lagrangians.

We also outline in this section the procedure for obtaining field equations by a variational principle applied to the gauge fields, and show that in our formalism the fundamental quantities to be varied in an action principle are the supersymmetric connections. We further show how the variation of the supersymmetric connection may be split into three independent variations involving the gravitino field, the contorsion tensor, and one more variable, which is associated with the variation of the covariant gradient of the origin twist-tensor field. In Sec. IV we give concluding remarks and suggestions for further work.

At the end of the paper we include two appendices which contain five theorems where we prove several relations that are used throughout the text. In Appendix A, we derive a theorem on the product of skew-symmetric twistors and their duals, which considerably simplifies a number of algebraic manipulations that are performed with these types of twistors. In Appendix B we prove four theorems which serve to establish a natural map between the set of Dirac operators and the set of skew-symmetric twistors $J_i{}^{\alpha\beta} = D_i O^{\alpha\beta}$ (the covariant derivative of the origin twistor) at each point of the manifold. In addition, in the same appendix, we make use of this map to give a representation independent derivation of some properties of the Dirac gamma operators.

Finally some remarks about notation: Throughout the text we use Latin indices to denote space-time variables, while lowercase Greek indices will be reserved for twistors and capital Greek indices for supertwistors in order to follow, as close as possible, the notation conventionally utilized in twistor theory.

II. SUPERTWISTOR SPACE AND BUNDLE STRUCTURES

A. Supertwistors

In I we developed the concept of supertwistors by following a procedure somewhat different to that originally used by Ferber.[9] Here we summarize the essential aspects of the algebra of these structures which will be required throughout the paper.

We define a supertwistor as an element of the graded vector space given by the orthogonal direct sum

$$\mathscr{V}_{3,2} = \mathscr{V}_{2,2} \oplus \mathscr{V}_1 = (\mathscr{G}_e \otimes \mathscr{U}_{2,2}) \oplus (\mathscr{G}_0 \otimes \mathscr{W}_1), \tag{2.1}$$

where $\mathscr{G}_e$ and $\mathscr{G}_0$ are the even and odd subsets, respectively, of a real Grassmann algebra $\mathscr{G}$ of dimension 2^d generated by a d-dimensional real vector space $\mathscr{K}_d$.

The subspace $\mathscr{V}_{2,2}$ is formally the same as ungraded twistor space, however supertwistors in $\mathscr{V}_{2,2}$ will be represented by the triple $(\omega^A, \pi_{A'}, 0)$, where the spinor components $\omega^A, \pi_{A'}$ are now complexified even elements of the Grassmann algebra $\mathscr{G}$. A supertwistor in $\mathscr{V}_{2,2}$ will be denoted by $u^\Sigma = (u^\sigma, 0) \leftrightarrow (\omega^A, \pi_{A'}, 0)$.

Supertwistors in the subspace $\mathscr{V}_1$ are of the form

$$\xi^\Sigma = \xi \tau^\Sigma, \tag{2.2}$$

where ξ is a complexified odd element of the Grassmann algebra $\mathscr{G}$, and τ^Σ is a normalized basis in the one-dimensional complex vector space $\mathscr{W}_1$ such that

$$\bar{\tau}_\Sigma \tau^\Sigma = i, \quad \bar{\tau}_\Sigma u^\Sigma = \bar{u}_\Sigma \tau^\Sigma = 0, \quad \forall\, u^\Sigma \in \mathscr{V}_{2,2}. \tag{2.3}$$

Throughout the discussion, we will leave the dimensionality of the space $\mathscr{K}_d$ undefined in order to allow for all possible supersymmetric theories with increasing degree of complexity as $d \to \infty$.

A supertwistor in $\mathscr{V}_{3,2}$ is written as $Z^\Sigma = u^\Sigma + \xi^\Sigma$, where $u^\Sigma \in \mathscr{V}_{2,2}$ and $\xi^\Sigma \in \mathscr{V}_1$. Furthermore, since the subspaces $\mathscr{V}_{2,2}$ and $\mathscr{V}_1$ are assumed to be orthogonal, a Hermitian inner product in $\mathscr{V}_{3,2}$ may be defined as

$$\langle Z | W \rangle = \bar{Z}_\Sigma W^\Sigma = \bar{u}_\Sigma v^\Sigma + \bar{\xi}_\Sigma \theta^\Sigma, \tag{2.4}$$

where $W^\Sigma = v^\Sigma + \theta^\Sigma$, and $\bar{Z}_\Sigma = \bar{u}_\Sigma + \bar{\xi}_\Sigma$ is the supertwistor conjugate to Z^Σ that has values in the dual space $\mathscr{V}'_{3,2}$.

The anticommutativity of odd Grassmann numbers implies

$$\begin{aligned} \bar{Z}_\Sigma W^\Sigma &= \bar{u}_\Sigma v^\Sigma + \bar{\xi}_\Sigma \theta^\Sigma = v^\Sigma \bar{u}_\Sigma - \theta^\Sigma \bar{\xi}_\Sigma \\ &= \bar{u}_\sigma v^\sigma + i\xi^* \theta = v^\sigma \bar{u}_\sigma - i\theta\xi^*. \end{aligned} \tag{2.5}$$

As pointed out in I, the graded group SU(2,2/1) is the set of linear transformations which leave (2.5) invariant.

We can now construct, in analogy to what is done with ordinary twistors, the tensor space $\mathscr{V}_{2,2}^{\wedge 2}$, $\mathscr{V}_{2,2}^{\wedge 4}$, $\mathscr{V}'^{\wedge 2}_{2,2}$, $\mathscr{V}'^{\wedge 4}_{2,2}$, $\mathscr{V}_{2,2} \otimes \mathscr{V}'_{2,2}$, etc. The only difference is that these spaces are now regarded as even-graded subspaces of the larger supertwistor spaces $\mathscr{V}_{3,2}^{\wedge 2}$, $\mathscr{V}_{3,2}^{\wedge 4}$, $\mathscr{V}'^{\wedge 2}_{3,2}$, $\mathscr{V}'^{\wedge 4}_{3,2}$, $\mathscr{V}_{3,2} \otimes \mathscr{V}'_{3,2}$, etc., respectively.

In particular we will also have here the special elements: (a) the totally antisymmetric supertwistor $\in \mathscr{V}_{2,2}^{\wedge 4}$,

$$\eta^{\Sigma\Gamma\Delta\Lambda} = \begin{cases} \eta^{\sigma\gamma\delta\lambda}, & \text{for } \Sigma,\Gamma,\Delta,\Lambda \text{ all different and} \\ & \text{values ranging from 0 to 3,} \\ 0, & \text{for any of the indices equal to 4;} \end{cases} \tag{2.6}$$

(b) the vertex of the null cone at infinity, infinity supertwistor, or metric supertwistor $\in \mathscr{V}_{2,2}^{\wedge 2}$,

$$I^{\Sigma\Gamma} = \left[\begin{array}{c|c} I^{\sigma\gamma} & 0 \\ \hline 0 & 0 \end{array}\right]; \tag{2.7}$$

and (c) the origin supertwistor $\in \mathscr{V}_{2,2}^{\wedge 2}$,

$$O^{\Sigma\Gamma} = \left[\begin{array}{c|c} O^{\sigma\gamma} & 0 \\ \hline 0 & 0 \end{array}\right]. \tag{2.8}$$

All these special quantities are chosen from the corresponding homogeneous subspaces of degree zero.

The properties of these supertwistors are, up to a Grassmann factor, the same as those for the corresponding ordinary twistors, and so is the property of certain elements of $\mathscr{V}_{2,2}^{\wedge 2}$ being real elements.

B. Gauge theory for the supergroup

The essential aspects of the philosophy that we will adopt for the construction of a gauge theory of the supergroup SU(2,2/1) [or SU(2,2/N) if internal symmetries are included] are fundamentally the same as those described in Ref. 6. For details we refer the reader to that paper. Here we will restrict ourselves to stressing the points of difference between the formalisms.

First, in place of $\mathscr{U}_{2,2}$ we start with $\mathscr{V}_{3,2}$ as the typical fiber of our twistor bundle.

If we also introduce $I^{\Sigma\Gamma}$ as part of the structure of $\mathscr{V}_{3,2}$, superconformal invariance will be broken and we will have a faithful representation of the super-Poincaré group.

Now, beginning with the four-dimensional base manifold $\mathscr{M}$, we can construct the bundles $\mathscr{V}_{3,2}(\mathscr{M})$, $\mathscr{V}_{3,2}^{\wedge 2}(\mathscr{M})$, $\mathscr{V}_{3,2}^{\wedge 4}(\mathscr{M})$ and $\mathscr{V}_{3,2}^{\wedge 5}(\mathscr{M})$. The cross section $I^{\Sigma\Gamma} \in\Gamma(\mathscr{M},\mathscr{V}_{2,2}^{\wedge 2}(\mathscr{M}))\subset\Gamma(\mathscr{M},\mathscr{V}_{3,2}^{\wedge 2}(\mathscr{M}))$ is taken as part of the structure of $\mathscr{V}_{3,2}(\mathscr{M})$ to give the bundle $(\mathscr{V}_{3,2},\ I^{\Sigma\Gamma})$ which from here on will be simply denoted by $\mathscr{V}_{3,2}(\mathscr{M})$.

At each point $q\in\mathscr{M}$, $(\mathscr{V}_{3,2})_q$ is the fiber above q and $(\mathscr{V}_{2,2})_q$ is a subspace of $(\mathscr{V}_{3,2})_q$.

As in the case of the Poincaré group,[1,6] no metric structure is initially assumed for the manifold $\mathscr{M}$. It is by the selection of an origin twistor field $O^{\Sigma\Gamma}$ that it becomes possible to define a map which leads to a unique way of imposing a metric structure and connection on the tangent bundle $\mathscr{T}(\mathscr{M})$.

C. Supertwistor connections

Let D_X^S be a connection on the bundle $\mathscr{V}_{3,2}(\mathscr{M})$, in which $\mathscr{V}_{3,2}$ is a typical fiber. We have that D_X^S satisfies the usual axioms of an arbitrary connection and is compatible with the inner product (2.5), i.e.,

$$X(\overline{Z}_\Sigma W^\Sigma) = (D_X^S\overline{Z}_\Sigma)\, W^\Sigma + \overline{Z}_\Sigma\, D_X^S W^\Sigma . \tag{2.9}$$

Moreover, D_X^S preserves the structure of the typical fiber $(\mathscr{V}_{3,2},I^{\Sigma\Gamma})$ so we also have

$$D_X^S\, I^{\Sigma\Gamma} = 0 . \tag{2.10}$$

If we now recall that two connections can differ only by a transformation linear in the even Grassmann coefficients, we can express the supersymmetric connection as

$$D_X^S Z^\Sigma = D_X Z^\Sigma + B_x{}^\Sigma{}_\Gamma Z^\Gamma , \tag{2.11}$$

where $Z^\Sigma(q)\in(\mathscr{V}_{3,2})_q$, i.e., Z^Σ is a cross section of the $\mathscr{V}_{3,2}(\mathscr{M})$ bundle, D_X is the twistor connection which leaves invariant the subspace $\mathscr{V}_{2,2}$ on which it acts, and $B(q)$ is a tensor field with values in $\mathscr{T}'_q\otimes(\mathscr{V}_{3,2})_q\otimes(\mathscr{V}'_{3,2})_q$ whose components are $B_x{}^\Sigma{}_\Gamma$ and which represents the action of the connection due to the supertranslations.

The connection D_x has the properties of being compatible with the inner product in $\mathscr{V}_{2,2}(\mathscr{M})$ and satisfies the conditions

$$D_X \eta^{\Sigma\Gamma\Delta\Lambda} = 0 , \tag{2.12}$$

$$D_X I^{\Sigma\Gamma} = 0 . \tag{2.13}$$

As an additional property, we shall require that

$$D_X \tau^\Sigma = 0 . \tag{2.14}$$

Note that Eqs. (2.11)–(2.14) serve to determine completely the action of D_X on the bundle $\mathscr{V}_{3,2}(\mathscr{M})$. In particular, it can be readily shown that D_X is also compatible with the inner product in $\mathscr{V}_{3,2}(\mathscr{M})$.

Now let us consider in detail the action of the supertranslations in (2.11). These can be given explicitly as the most general linear combination of the generators of supertranslations as derived in I [see Eq. (3.9)]. Thus we have

$$B_x{}^\Sigma{}_\Gamma = \psi_x{}^\Sigma\bar{\tau}_\Gamma - \tau^\Sigma\bar{\psi}_{x\Gamma} , \tag{2.15}$$

where the supertwistor $\psi_x{}^\Sigma$ is represented by the triple

$$\psi_x{}^\Sigma = (\psi_x{}^\sigma,0)\leftrightarrow(\rho_x{}^A + i\sigma_x{}^A,0,0) , \tag{2.16}$$

and $\rho_x{}^A$, $\sigma_x{}^A$ are two-component real spinors which are odd in the Grassmann algebra $\mathscr{G}$.

D. Lie and exterior derivative connections

In addition to the connections defined so far we will also need the concepts of Lie and exterior derivative connections. This construction will turn out to be most convenient in our later discussion of supertwistor curvatures.

To this end, let us first define a torsionless connection D_X^0 on the bundle $\mathscr{V}_{2,2}(\mathscr{M})$ by

$$D_X^0 u^\Sigma = D_X u^\Sigma + (D_X O^{\Sigma\Gamma})I_{\Gamma\Lambda}u^\Lambda , \quad u^\Lambda(q)\in(\mathscr{V}_{2,2})_q . \tag{2.17}$$

Note that this connection has the following properties [which follow directly from (2.16)]: (a) for $u^\Sigma(q)\in\mathscr{G}_e\otimes(\mathscr{S}_I)_q$,

$$D_X^0 u^\Sigma = D_X u^\Sigma , \tag{2.18}$$

(b) for $u^\Sigma(q)\in\mathscr{G}_e\otimes(\mathscr{S}_O)_q$,

$$\begin{aligned} D_X^0 u^\Sigma &= D_X\,[(S_O)^\Sigma{}_\Gamma u^\Gamma\,] + (D_X O^{\Sigma\Gamma})I_{\Gamma\Lambda}u^\Lambda \\ &= D_X(O^{\Sigma\Gamma}I_{\Lambda\Gamma}u^\Lambda) + (D_X O^{\Sigma\Gamma})I_{\Gamma\Lambda}u^\Lambda \\ &= O^{\Sigma\Gamma}I_{\Lambda\Gamma}D_X u^\Lambda = (S_O)^\Sigma{}_\Lambda D_X u^\Lambda , \end{aligned} \tag{2.19}$$

where

$$(S_O)^\Sigma{}_\Gamma = O^{\Sigma\Lambda}I_{\Gamma\Lambda} \tag{2.20}$$

is the extension of the idempotent defined by (2.13) in I, i.e., the projection of $D_X u^\Sigma$ onto $\mathscr{G}_e\otimes(\mathscr{S}_O)_q$ gives $D_X^0 u^\Sigma$. Also

(c) $$D_X^0\, I^{\Sigma\Gamma} = 0 , \tag{2.21}$$

(d) $$D_X^0\, O^{\Sigma\Gamma} = 0 , \tag{2.22}$$

and

(e) $$[(D_X^0 u)^c]^\Sigma = (D_X^0 u^c)^\Sigma , \tag{2.23}$$

where $(u^c)^\Sigma$ is the charge conjugation of $u^\Sigma(q)$, i.e.,

$$(u^c)^\Sigma = \bar{u}_\Gamma(I^{\Gamma\Sigma} - O^{\Gamma\Sigma}) . \tag{2.24}$$

[Compare to the definition given by Eq. (B56) in Appendix B.]

Recall that in Ref. 6, we defined the torsion tensor on the fiber as the action of the curvature tensor on the origin twistor [see Eq. (3.42) there]. Consequently, if we use the connection D_X^0 in particular, we have, by virtue of (2.22),

$$(\mathbf{T}_{\mathscr{T}}^0)^{\Sigma\Gamma}(\mathbf{x},\mathbf{y}) = (D_X^0\, D_Y^0 - D_Y^0\, D_X^0 - D_{[X,Y\,]}^0)O^{\Sigma\Gamma} = 0 , \tag{2.25}$$

i.e., D_X^0 is indeed torsionless as asserted previously.

We can now introduce an operator $\mathscr{L}_X$, the Lie derivative connection, as a combined Lie derivative and twistor connection acting on fields of differential p forms ($p=0,1,2,\ldots$),

$$u^\Sigma \in (\mathscr{V}_{2,2})_q\,,\quad \psi^\Sigma \equiv \psi_i{}^\Sigma\, dx^i \in \mathscr{T}'_q \otimes (\mathscr{V}_{2,2})_q\,,$$

$$\Phi^\Sigma \equiv \tfrac{1}{2}\, \Phi_{[i,j]}{}^\Sigma\, dx^i \wedge dx^j \in \mathscr{T}'^{\wedge 2}_q \otimes (\mathscr{V}_{2,2})_q\,,$$

etc. The action of $\mathscr{L}_X$ can be defined as

$$\mathscr{L}_X u^\Sigma = D^0_X u^\Sigma\,, \tag{2.26}$$

$$(\mathscr{L}_X \psi^\Sigma)(\mathbf{y}) = \mathscr{L}_X[\psi^\Sigma(\mathbf{y})] - \psi^\Sigma(\mathscr{L}_X\, \mathbf{y}) = D^0_X[\psi^\Sigma(\mathbf{y})] - \psi^\Sigma([\mathbf{x},\mathbf{y}])\,, \tag{2.27}$$

$$(\mathscr{L}_X \Phi^\Sigma)(\mathbf{y},\mathbf{z}) = \mathscr{L}_X[\Phi^\Sigma(\mathbf{y},\mathbf{z})] - \Phi^\Sigma(\mathscr{L}_X\, \mathbf{y},\mathbf{z}) - \Phi^\Sigma(\mathbf{y},\mathscr{L}_X\, \mathbf{z})\,, \text{ etc.,} \tag{2.28}$$

where $\psi^\Sigma(\mathbf{y}) \equiv y^i \psi_i{}^\Sigma$, $\Phi^\Sigma(\mathbf{y},\mathbf{z}) = y^i z^j \Phi_{[ij]}{}^\Sigma$, and $\mathscr{L}_X u^\Sigma$, $\mathscr{L}_X \psi^\Sigma$, $\mathscr{L}_X \Phi^\Sigma$, at q have values in $(\mathscr{V}_{2,2})_q$, $\mathscr{T}'_q \otimes (\mathscr{V}_{2,2})_q$, $\mathscr{T}'^{\wedge 2}_q \otimes (\mathscr{V}_{2,2})_q$, respectively.

An exterior derivative-connection $\mathscr{D}$, is a combined exterior derivative and twistor connection, defined on fields of p forms $u^\Sigma, \psi^\Sigma, \Phi^\Sigma$, etc., where $\mathscr{D} u^\Sigma$, $\mathscr{D} \wedge \psi^\Sigma$, $\mathscr{D} \wedge \Phi^\Sigma$, etc., at q are in $\mathscr{T}'_q \otimes (\mathscr{V}_{2,2})_q$, $\mathscr{T}'^{\wedge 2}_q \otimes (\mathscr{V}_{2,2})_q$, $\mathscr{T}'^{\wedge 3}_q \otimes (\mathscr{V}_{2,2})_q$, etc., respectively.

The action of $\mathscr{D}$ can also be defined by induction as

$$(\mathscr{D} u^\Sigma)(x) = D^0_X u^\Sigma\,, \tag{2.29}$$

$$(\mathscr{D} \wedge \psi^\Sigma)(\mathbf{x},\mathbf{y}) = (\mathscr{L}_X \psi^\Sigma)(\mathbf{y}) - (\mathscr{D}[\psi^\Sigma(\mathbf{x})])(\mathbf{y})\,, \tag{2.30}$$

$$(\mathscr{D} \wedge \Phi^\Sigma)(\mathbf{x},\mathbf{y},\mathbf{z}) = (\mathscr{L}_X \Phi^\Sigma)(\mathbf{y},\mathbf{z}) - (\mathscr{D} \wedge [\Phi^\Sigma(\mathbf{x})])(\mathbf{y},\mathbf{z}), \text{ etc.,} \tag{2.31}$$

where $\Phi^\Sigma(\mathbf{x}) = x^i \Phi_{[ij]}{}^\Sigma\, dx^j$.

From (2.27), (2.29), and (2.30) we get, in addition, the useful result

$$D^0_X[\psi^\Sigma(\mathbf{y})] - D^0_Y[\psi^\Sigma(\mathbf{x})] - \psi^\Sigma([\mathbf{x},\mathbf{y}]) = (\mathscr{D} \wedge \psi^\Sigma)(\mathbf{x},\mathbf{y})\,. \tag{2.32}$$

Note that although the above definitions are given for supertwistors constructed from the subspace $\mathscr{V}_{2,2}$, they can be readily generalized to supertwistors and their tensors derived from $\mathscr{V}_{3,2}$ by making use of (2.14) and (2.17).

E. Supertwistor curvature

We define the supertwistor curvature tensor $\mathfrak{S}^S$, with value at q in $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes (\mathscr{V}_{3,2})_q \otimes (\mathscr{V}'_{3,2})_q$, by means of the expression

$$(\mathfrak{S}^S)^\Sigma{}_\Gamma(\mathbf{x},\mathbf{y}) Z^\Gamma = (D^S_X D^S_Y - D^S_Y D^S_X - D^S_{[X,Y]}) Z^\Sigma\,, \tag{2.33}$$

where Z^Σ is a supertwistor field with $Z^\Sigma(q) \in (\mathscr{V}_{3,2})_q$.

Moreover, making use of (2.1)–(2.3) and of (2.14) and (2.15), one readily obtains

$$\begin{aligned}(D^S_X D^S_Y &- D^S_Y D^S_X - D^S_{[X,Y]}) Z^\Sigma \\ &= (D_X D_Y - D_Y D_X - D_{[X,Y]})\, Z^\Sigma \\ &\quad - \tau^\Sigma \{D_X[\bar\psi_\Gamma(\mathbf{y})] - D_Y[\bar\psi_\Gamma(\mathbf{x})] - \bar\psi_\Gamma([\mathbf{x},\mathbf{y}])\}\, Z^\Gamma \\ &\quad + \{D_X[\psi^\Sigma(\mathbf{y})] - D_Y[\psi^\Sigma(\mathbf{x})] - \psi^\Sigma([\mathbf{x},\mathbf{y}])\}\, \bar\tau_\Gamma\, Z^\Gamma \\ &\quad - i[\psi^\Sigma(\mathbf{x})\bar\psi_\Gamma(\mathbf{y}) - \psi^\Sigma(\mathbf{y})\bar\psi_\Gamma(\mathbf{x})]\, Z^\Gamma\,. \end{aligned} \tag{2.34}$$

Now substituting (2.32) into (2.34), and using the definition in (2.33) (and a similar one for the twistor curvature in $\mathscr{T}' \otimes \mathscr{T}' \otimes \mathscr{V}_{2,2} \otimes \mathscr{V}'_{2,2}$), yields

$$\begin{aligned}(\mathfrak{S}^S)^\Sigma{}_\Gamma(\mathbf{x},\mathbf{y}) &= (\mathfrak{S})^\Sigma{}_\Gamma(\mathbf{x},\mathbf{y}) - \tau^\Sigma(\mathscr{D} \wedge \bar\psi_\Gamma)(\mathbf{x},\mathbf{y}) + (\mathscr{D} \wedge \psi^\Sigma)(\mathbf{x},\mathbf{y})\, \bar\tau_\Gamma \\ &\quad - i\,[\psi^\Sigma(\mathbf{x})\bar\psi_\Gamma(\mathbf{y}) - \psi^\Sigma(\mathbf{y})\bar\psi_\Gamma(\mathbf{x})]\,. \end{aligned} \tag{2.35}$$

In order to relate the twistor curvature tensor $\mathfrak{S}$ appearing on the right-hand side of Eq. (2.35) to the Riemann curvature tensor we recall that [cf. Eq. (3.36) in Ref. 6]

$$\mathbf{R}_{\mathscr{G}}(\mathbf{x},\mathbf{y})(u^\Sigma v^\Gamma - v^\Sigma u^\Gamma) = (D_X D_Y - D_Y D_X - D_{[X,Y]})(u^\Sigma v^\Gamma - v^\Sigma u^\Gamma)\,, \tag{2.36}$$

where u^Σ, v^Γ are taken here to have values in $(\mathscr{V}_{2,2})_q$ in general and $\mathbf{R}_{\mathscr{G}}$ has values in $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes (\mathscr{V}^{\wedge 2}_{2,2})_q \otimes (\mathscr{V}'^{\wedge 2}_{2,2})_q$.

Observe now that

$$\begin{aligned}(D_X D_Y &- D_Y D_X - D_{[X,Y]})(u^\Sigma v^\Gamma - v^\Sigma u^\Gamma) \\ &= [(D_X D_Y - D_Y D_X - D_{[X,Y]}) u^\Sigma]\, v^\Gamma \\ &\quad + u^\Sigma [(D_X D_Y - D_Y D_X - D_{[X,Y]}) v^\Gamma] \\ &\quad - [(D_X D_Y - D_Y D_X - D_{[X,Y]}) v^\Sigma]\, u^\Gamma \\ &\quad - v^\Sigma [(D_X D_Y - D_Y D_X - D_{[X,Y]})\, u^\Gamma]\,, \end{aligned}$$

i.e.,

$$\begin{aligned}\mathbf{R}_{\mathscr{G}}(\mathbf{x},\mathbf{y})(u^\Sigma v^\Gamma - v^\Sigma u^\Gamma) &= [(\mathfrak{S})^\Sigma{}_\Lambda(\mathbf{x},\mathbf{y}) u^\Lambda]\, v^\Gamma + u^\Sigma [(\mathfrak{S})^\Gamma{}_\Lambda(\mathbf{x},\mathbf{y}) v^\Lambda] \\ &\quad - [(\mathfrak{S})^\Sigma{}_\Lambda(\mathbf{x},\mathbf{y}) v^\Lambda]\, u^\Gamma - v^\Sigma [(\mathfrak{S})^\Gamma{}_\Lambda(\mathbf{x},\mathbf{y}) u^\Lambda]\,. \end{aligned} \tag{2.37}$$

Moreover, if we use the symbol

$$E^\Gamma{}_\Pi \equiv (S_I)^\Gamma{}_\Pi + (S_O)^\Gamma{}_\Pi \in \mathscr{V}_{2,2} \otimes \mathscr{V}'_{2,2} \tag{2.38}$$

to represent the identity twistor acting on this subspace, we then have

$$\mathbf{R}_{\mathscr{G}}(\mathbf{x},\mathbf{y})(u^\Sigma v^\Gamma - v^\Sigma u^\Gamma) = [(\mathfrak{S})^\Sigma{}_\Lambda(\mathbf{x},\mathbf{y}) E^\Gamma{}_\Pi + E^\Sigma{}_\Lambda (\mathfrak{S})^\Gamma{}_\Pi(\mathbf{x},\mathbf{y})](u^\Lambda v^\Pi - v^\Lambda u^\Pi) \tag{2.39a}$$

and

$$(\mathbf{R}_{\mathscr{G}})^{\Sigma\Gamma\Omega\Phi}(\mathbf{x},\mathbf{y})\, \eta_{\Omega\Phi\Lambda\Pi} = (\mathfrak{S})^\Sigma{}_{[\Lambda}(\mathbf{x},\mathbf{y}) E^\Gamma{}_{\Pi]} + E^\Sigma{}_{[\Lambda} (\mathfrak{S})^\Gamma{}_{\Pi]}(\mathbf{x},\mathbf{y})\,. \tag{2.39b}$$

Putting the indices $\Lambda = \Gamma$ in (2.39b) leads to

$$(\mathbf{R}_{\mathscr{G}})^{\Sigma\Lambda\Omega\Phi}(\mathbf{x},\mathbf{y})\, \eta_{\Omega\Phi\Lambda\Pi} = -2\, \mathfrak{S}^\Sigma{}_\Pi(\mathbf{x},\mathbf{y}) - E^\Sigma{}_\Pi\, \mathfrak{S}^\Lambda{}_\Lambda(\mathbf{x},\mathbf{y})\,. \tag{2.40}$$

Equation (2.40) is our desired relation between the Riemann curvature and twistor curvature tensors. This expression, however, can be put in more familiar terms by explicitly accounting for torsion. We thus have [cf. Eq. (3.47) in Ref. 6]

$$\begin{aligned}(\mathbf{R}_{\mathscr{G}})^{\Sigma\Lambda\Omega\Phi}(\mathbf{x},\mathbf{y}) &= (\mathbf{R}_{\mathscr{F}})^{\Sigma\Lambda\Omega\Phi}(\mathbf{x},\mathbf{y}) + \tfrac{1}{2}[(\mathbf{T}_{\mathscr{F}})^{\Sigma\Lambda}(\mathbf{x},\mathbf{y}) I^{\Omega\Phi} \\ &\quad - I^{\Sigma\Lambda}(\mathbf{T}_{\mathscr{F}})^{\Omega\Phi}(\mathbf{x},\mathbf{y})]\,, \end{aligned} \tag{2.41}$$

where $\mathbf{R}_{\mathscr{F}}$ is the curvature tensor which has its values in $\mathscr{G}_e \otimes \mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{F}_q \otimes \mathscr{F}_q$; $\mathscr{F}_q$ is the subspace of real twistors orthogonal both to $I^{\Sigma\Gamma}$ and $O^{\Sigma\Gamma}$, and having a Minkowski inner product with signature $(+\,-\,-\,-)$, and $\mathbf{T}_{\mathscr{F}}$ is the torsion tensor with values in $\mathscr{G}_e \otimes \mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{F}_q$.

The curvature $\mathbf{R}_{\mathscr{G}}$ can be projected in turn onto the tangent bundle to yield the tensor $\mathbf{R}_{\mathscr{T}}$, with values in $\mathscr{G}_e \otimes \mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{T}_q \otimes \mathscr{T}_q$, by making use of the field

$$J_i^{\Sigma\Gamma} \equiv D_i O^{\Sigma\Gamma}, \tag{2.42}$$

which acts as a bijective map[6] of $\mathscr{T}_q$ on the subspace $\mathscr{G}_e \otimes \mathscr{F}_q \subset (\mathscr{V}_{2,2}^{\wedge 2})_q$. Thus, introducing an holonomic basis $\{\mathbf{e}_i\}$ in the tangent bundle and the dual basis $\{\mathbf{e}^i\}$ in the cotangent bundle, we get

$$(\mathbf{R}_{\mathscr{T}})^{\Sigma\Lambda\Omega\Phi}(\mathbf{e}_i,\mathbf{e}_j) = R_{ij}{}^{kl} J_k{}^{\Sigma\Lambda} J_l{}^{\Lambda\Phi}. \tag{2.43}$$

Similarly $\mathbf{T}_{\mathscr{G}}$ is related to the torsion tensor $\mathbf{T}_{\mathscr{T}}$ in the tangent bundle, with values in $\mathscr{G}_e \otimes \mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{T}_q$, by

$$(\mathbf{T}_{\mathscr{T}})^{\Sigma\Lambda}(\mathbf{e}_i,\mathbf{e}_j) = T_{ij}{}^k J_k{}^{\Sigma\Lambda}. \tag{2.44}$$

Substituting (2.43) and (2.44) into (2.41), results in

$$(\mathbf{R}_{\mathscr{G}})^{\Sigma\Lambda\Omega\Phi}(\mathbf{e}_i,\mathbf{e}_j) = R_{ij}{}^{kl} J_k{}^{\Sigma\Lambda} J_l{}^{\Omega\Phi} + \tfrac{1}{2} T_{ij}{}^k (J_k{}^{\Sigma\Lambda} I^{\Omega\Phi} - I^{\Sigma\Lambda} J_k{}^{\Omega\Phi}). \tag{2.45}$$

Replacing now the left side of (2.41) by (2.45), yields

$$2R_{ij}{}^{kl} J_k{}^{\Sigma\Lambda} J_{l\Lambda\Pi} + T_{ij}{}^k (J_k{}^{\Sigma\Lambda} I_{\Lambda\Pi} - I^{\Sigma\Lambda} J_{k\Lambda\Pi}) = -2\,\mathfrak{S}^\Sigma{}_\Pi(\mathbf{e}_i,\mathbf{e}_j) - E^\Sigma{}_\Pi\,\mathfrak{S}^\Lambda{}_\Lambda(\mathbf{e}_i,\mathbf{e}_j). \tag{2.46}$$

The value of $\mathfrak{S}^\Lambda{}_\Lambda(\mathbf{e}_i,\mathbf{e}_j)$ in the last term of (2.46) can be obtained by means of one additional contraction. We find

$$2R_{ij}{}^{kl} g_{kl} + T_{ij}{}^k (J_k{}^{\Sigma\Lambda} I_{\Sigma\Lambda} - I^{\Sigma\Lambda} J_{k\Sigma\Lambda}) = 6\,\mathfrak{S}^\Lambda{}_\Lambda(\mathbf{e}_i,\mathbf{e}_j), \tag{2.47}$$

where we have made use of the fact that

$$J_k{}^{\Sigma\Lambda} J_{l\Sigma\Lambda} = g_{kl}$$

[see (3.5a) below].

Moreover, since $J_k{}^{\Sigma\Lambda}$ and $I^{\Sigma\Lambda}$ are real twistors we also have

$$J_k{}^{\Sigma\Lambda} I_{\Sigma\Lambda} = J_{k\Sigma\Lambda} I^{\Sigma\Lambda} = 0,$$

since $J_k{}^{\Sigma\Lambda} \in \mathscr{F}$.

It follows readily that

$$\mathfrak{S}^\Lambda{}_\Lambda(\mathbf{e}_i,\mathbf{e}_j) = 0, \tag{2.48}$$

and (2.46) reduces to

$$\mathfrak{S}^\Sigma{}_\Pi(\mathbf{e}_i,\mathbf{e}_j) = -R_{ij}{}^{kl} J_k^{\Sigma\Lambda} J_{l\Lambda\Pi} + T_{ij}{}^k I^{\Sigma\Lambda} J_{k\Lambda\Pi}. \tag{2.49}$$

In arriving at (2.49) we have made use of

$$I^{\Sigma\Lambda} J_{k\Lambda\Pi} = -J_k{}^{\Sigma\Lambda} I_{\Lambda\Pi}, \tag{2.50}$$

which results directly from the identity (A1), derived in Appendix A.

Finally, as a consequence of (2.49) the expression of the supertwistor curvature tensor given in (2.35) takes the form

$$(\mathfrak{S}^S)_{ij}{}^\Sigma{}_\Gamma \equiv (\mathfrak{S}^S)^\Sigma{}_\Pi(\mathbf{e}_i,\mathbf{e}_j) = -R_{ij}{}^{kl} J_k{}^{\Sigma\Lambda} J_{l\Lambda\Pi} + T_{ij}{}^k I^{\Sigma\Lambda} J_{k\Lambda\Pi} - \tau^\Sigma \mathscr{D}_{[i}\bar{\psi}_{j]\Pi} + \mathscr{D}_{[i}\psi_{j]}^\Sigma \bar{\tau}_\Pi - i\psi_{[i}{}^\Sigma \bar{\psi}_{j]\Pi}, \tag{2.51}$$

where

$$(\mathscr{D} \wedge \boldsymbol{\psi}^\Sigma)_{ij} = (\mathscr{D} \wedge \boldsymbol{\psi}^\Sigma)(\mathbf{e}_i,\mathbf{e}_j) = \mathscr{D}_{[i}\psi_{j]}^\Sigma \tag{2.52}$$

and

$$\psi_i{}^\Sigma = \boldsymbol{\psi}^\Sigma(\mathbf{e}_i), \quad \bar{\psi}_{j\Sigma} = \bar{\psi}_\Sigma(\mathbf{e}_j).$$

III. SUPERGRAVITY LAGRANGIANS AND ACTION PRINCIPLES

Making use of the results in the preceding section we can now set up Lagrangian densities $\mathscr{L}[\mathfrak{S}^S(\mathscr{M})]$ for the gauge fields (i.e., the connection D^S_X) as scalar functionals of $\mathfrak{S}^S(\mathscr{M})$, which by construction will be locally super-Poincaré-invariant. Note that by virtue of Eq. (2.51) these Lagrangian densities will determine the allowed functional form of these Lagrangians in terms of the quantities $R_{ij}{}^{kl}$, $T_{ij}{}^k$, $\mathscr{D}_{[i}\psi_{j]}{}^\Sigma$, $\mathscr{D}_{[i}\bar{\psi}_{j]}^\Sigma$, $\psi_i{}^\Sigma$, and $\bar{\psi}_{i\Sigma}$. Note also that the supertwistor curvature has two files in the contangent bundle and the third and fourth files in $\mathscr{V}_{3,2}(\mathscr{M})$ and $\mathscr{V}'_{3,2}(\mathscr{M})$, respectively, i.e., in the fiber space above the base manifold. Thus, in order to be able to construct scalars[10] from this quantity we must first project (2.51) so that the resulting tensor has all its files in the same space or its dual. We can do this most directly by following a procedure analogous to that designed for such a purpose in Ref. 6 and which involves using the action of the connection on the origin twistor to construct a unique map from the tangent bundle to the supertwistor bundle.

We digress briefly in order to review the essential properties of this J^S map which is explicitly defined by

$$x^i \rightarrow x^i (J^S)_i{}^{\Sigma\Gamma} \equiv D^S_X O^{\Sigma\Gamma}. \tag{3.1}$$

Moreover, making use of (2.14) and (2.15) it can be shown that

$$(J^S)_i{}^{\Sigma\Gamma} = D_i O^{\Sigma\Gamma} + B_i{}^\Sigma{}_\Lambda O^{\Lambda\Gamma} - B_i{}^\Gamma{}_\Lambda O^{\Lambda\Sigma} = J_i{}^{\Sigma\Gamma} - \tau^\Sigma \bar{\psi}_{i\Lambda} O^{\Lambda\Gamma} + \tau^\Gamma \bar{\psi}_{i\Lambda} O^{\Lambda\Sigma}. \tag{3.2}$$

Observe that $(\bar{J}^S)_{i\Sigma\Gamma}$ and $(J^S)_i{}^{\Sigma\Gamma}$ can also be used to map fields $Y^{\Sigma\Gamma}$ and $\bar{Z}_{\Sigma\Gamma}$ with values at q in $(\mathscr{V}_{3,2}^{\wedge 2})_q$ and $(\mathscr{V}'^{\wedge 2}_{3,2})_q$, respectively, into fields y_i and z_i with values at q in $\mathscr{G}_e \otimes \mathscr{T}'_q$ according to

$$y_i = (\bar{J}^S)_{i\Sigma\Gamma} Y^{\Sigma\Gamma}, \tag{3.3a}$$

$$z_i = (J^S)_i{}^{\Sigma\Gamma} \bar{Z}_{\Sigma\Gamma}. \tag{3.3b}$$

A similar procedure, based on induction from Eqs. (3.3), can be used for the mapping of tensors constructed from the spaces $\mathscr{V}_{3,2}^{\wedge 2}$ and $\mathscr{V}'^{\wedge 2}_{3,2}$ into cotangent bundle tensors.

Consider, in particular, the tensor $\frac{1}{2}\bar{\eta}_{\Sigma\Lambda\Phi\Pi}$ which acts as a metric tensor on $(\mathscr{V}_{2,2}^{\wedge 2})_q$ to give the inner product

$$L_{\Sigma\Lambda} M^{\Sigma\Lambda} = L^{\Sigma\Lambda} M_{\Sigma\Lambda}, \quad L^{\Sigma\Lambda}, \quad M^{\Sigma\Lambda} \in \mathscr{V}_{2,2}^{\wedge 2} \tag{3.4}$$

[cf. Eq. (2.6) in I].

Mapping $\frac{1}{2}\bar{\eta}_{\Sigma\Lambda\Phi\Pi}$ with $(J^S)_i{}^{\Sigma\Gamma}$, and making use of (3.2), yields

$$G_{ij} = \tfrac{1}{2}(J^S)_i{}^{\Sigma\Lambda} \bar{\eta}_{\Sigma\Lambda\Phi\Pi} (J^S)_j{}^{\Phi\Pi} = \tfrac{1}{2} J_i{}^{\Sigma\Lambda} \bar{\eta}_{\Sigma\Lambda\Phi\Pi} J_j{}^{\Phi\Pi} = J_i{}^{\Sigma\Lambda} J_{j\Sigma\Lambda} = J_{i\Sigma\Lambda} J_j{}^{\Sigma\Lambda}, \tag{3.5}$$

where $\mathbf{G}$ is valued in $\mathscr{G}_e \otimes \mathscr{T}'_q \otimes \mathscr{T}'_q$.

The tensor $\mathbf{G}$ serves to impose the metric structure given in the fibers on the tangent space $\mathscr{T}_q$, i.e., we can define an inner product in the tangent bundle by

$$\mathbf{x} \cdot \mathbf{y} = x^i G_{ij} y^j. \tag{3.6}$$

In terms of components in an holonomic basis, (3.5) and (3.6) read

$$g_{ij} \equiv \mathbf{e}_i \cdot \mathbf{e}_j = G_{ij} = J_{i\Sigma\Lambda} J_j{}^{\Sigma\Lambda}, \tag{3.5'}$$

$$\mathbf{x}\cdot\mathbf{y}=x^iy^jg_{ij}\,. \tag{3.6'}$$

We return now to the main objective of this section, which is the construction of scalar invariants from the supertwistor curvature tensor. In what follows we shall concentrate on contractions of (2.51) with (3.2) which lead to the Lagrangian for simple supergravity. The extension of the procedure to other permissible Lagrangians, and to Lagrangians for extended supergravities, is suggested by the approach here adopted and is rather straightforward.

First we transvect (2.51) with $(J^S)_m{}^{\Pi\Gamma}$, to get

$$\begin{aligned}
&(\mathfrak{S}^S)_{ij}{}^{\Sigma}{}_{\Pi}(J^S)_m{}^{\Pi\Gamma}\\
&\quad= -R_{ij}{}^{kl}J_k{}^{\Sigma\Lambda}J_{l\Lambda\Pi}\,J_m{}^{\Pi\Gamma}+R_{ij}{}^{kl}J_k{}^{\Sigma\Phi}J_{l\Phi\Pi}O^{\Pi\Lambda}\bar\psi_{m\Lambda}\tau^{\Gamma}\\
&\qquad+T_{ij}{}^{k}I^{\Sigma\Lambda}J_{k\Lambda\Pi}\,J_m{}^{\Pi\Gamma}-T_{ij}{}^{k}I^{\Sigma\Phi}J_{k\Phi\Pi}O^{\Pi\Lambda}\bar\psi_{m\Lambda}\tau^{\Gamma}\\
&\qquad-\tau^{\Sigma}(\mathscr{D}_{[i}\bar\psi_{j]\Pi})\,J_m{}^{\Pi\Gamma}+\tau^{\Sigma}\tau^{\Gamma}(\mathscr{D}_{[i}\bar\psi_{j]\Pi})O^{\Pi\Lambda}\bar\psi_{m\Lambda}\\
&\qquad+i\,(\mathscr{D}_{[i}\psi_{j]}{}^{\Sigma})O^{\Gamma\Pi}\bar\psi_{m\Pi}-i\psi_{[i}{}^{\Sigma}\bar\psi_{j]\Pi}\,J_m{}^{\Pi\Gamma}\,.
\end{aligned}\tag{3.7}$$

The above equation still has two supertwistor indices and three indices in the contangent bundle free. It would seem natural that the next operation should then be a double contraction with $(\bar J^S)_{n\Sigma\Gamma}$. However, if we do this, all the important dynamical information contained in (3.7) relating to the gravitational and the $\psi_i{}^{\Sigma}$ fields will be lost.

To circumvent this problem we transvect (3.7) first on the right with $(\tilde\gamma_5)_{\Gamma}{}^{\Xi}$, where $(\tilde\gamma_5)_{\Gamma}{}^{\Xi}=(\gamma_5)^{\Xi}{}_{\Gamma}$ is the transpose of the Dirac gamma operator defined by [see Eq. (B51a) in Appendix B]

$$(\gamma_5)^{\Xi}{}_{\Gamma}=-i[(S_I)^{\Xi}{}_{\Gamma}-(S_O)^{\Xi}{}_{\Gamma}]\,, \tag{3.8}$$

and

$$(\gamma_5)^{\Sigma}{}_{\Gamma}(\gamma_5)^{\Gamma}{}_{\Lambda}=-E^{\Sigma}{}_{\Lambda}\,, \tag{3.9}$$

and contract the result from the left with $(\bar J^S)_{n\Sigma\Xi}$.

We therefore get

$$\begin{aligned}
&(\bar J^S)_{n\Sigma\Xi}(\mathfrak{S}^S)_{ij}{}^{\Sigma}{}_{\Pi}(J^S)_m{}^{\Pi\Gamma}(\tilde\gamma_5)_{\Gamma}{}^{\Xi}\\
&\quad=-R_{ij}{}^{kl}J_k{}^{\sigma\lambda}J_{l\lambda\alpha}\,J_m{}^{\alpha\beta}(\tilde\gamma_5)_{\beta}{}^{\xi}J_{n\sigma\xi}\\
&\qquad+T_{ij}{}^{k}I^{\sigma\alpha}J_{k\alpha\beta}\,J_m{}^{\beta\lambda}(\tilde\gamma_5)_{\lambda}{}^{\xi}J_{n\sigma\xi}\\
&\qquad+i(\mathscr{D}_{[i}\bar\psi_{j]\sigma})\,J_m{}^{\sigma\alpha}(\tilde\gamma_5)_{\alpha}{}^{\beta}O_{\beta\lambda}\psi_n{}^{\lambda}\\
&\qquad+i(\mathscr{D}_{[i}\psi_{j]}{}^{\sigma})\,J_{n\sigma\alpha}(\gamma_5)^{\alpha}{}_{\beta}O^{\beta\lambda}\bar\psi_{m\lambda}\\
&\qquad+i\psi_{[i}{}^{\sigma}J_{|n\sigma\lambda}(\gamma_5)^{\lambda}{}_{\alpha}\,J_{m|}{}^{\alpha\beta}\bar\psi_{j]\beta}\,.
\end{aligned}\tag{3.10}$$

Note that in writing the right side of (3.10), we have made explicit use of the fact that this expression has no free supertwistor indices left and that the components in $\mathscr{V}_1$ and $\mathscr{V}_1'$ have been contracted out completely. Thus all supertwistors in the right of (3.10) are valued in $\mathscr{V}_{2,2}$, $\mathscr{V}_{2,2}'$ or are tensor constructed from these spaces. Consequently, by virtue of our previous definition (cf. Sec. II), we can use lowercase Greek indices (running from 0 to 3) to denote these supertwistors.

Also note that because of the spaces in which $\psi_i{}^{\sigma}$, $J_j{}^{\sigma\alpha}$, and $(\gamma_5)^{\alpha}{}_{\beta}$ are situated, the last term in (3.10) drops out. It is important to point out here that this cancellation has nothing to do with the fact that the coefficients of the fields $\psi_i{}^{\sigma}$ are Grassmann variables.

Furthermore, making use of the relation between the J's and the Dirac gammas [see Appendix B, Eqs. (B61) and (B62)], we get

$$J_i{}^{\alpha\beta}=\tfrac{1}{2}(\gamma_i)^{\alpha}{}_{\lambda}(\gamma_5)^{\lambda}{}_{\kappa}C^{\kappa\beta}\,, \tag{3.11a}$$

$$J_{i\alpha\beta}=-\tfrac{1}{2}C_{\alpha\lambda}(\gamma_i)^{\lambda}{}_{\kappa}(\gamma_5)^{\kappa}{}_{\beta}\,, \tag{3.11b}$$

where

$$C^{\alpha\beta}=O^{\alpha\beta}-I^{\alpha\beta},\quad C^{\alpha\beta}C_{\beta\kappa}=E^{\beta}{}_{\kappa}\,, \tag{3.12}$$

we have

$$\begin{aligned}
&-J_k{}^{\sigma\lambda}J_{l\lambda\alpha}J_m{}^{\alpha\beta}(\tilde\gamma_5)_{\beta}{}^{\xi}J_{n\sigma\xi}\\
&\quad=-\tfrac{1}{16}(\gamma_k)^{\sigma}{}_{\lambda}(\gamma_l)^{\lambda}{}_{\alpha}(\gamma_m)^{\alpha}{}_{\beta}(\gamma_n)^{\beta}{}_{\xi}(\gamma_5)^{\xi}{}_{\sigma}\\
&\quad=-\tfrac{1}{16}\,\mathrm{tr}(\gamma_k\gamma_l\gamma_m\gamma_n\gamma_5)\\
&\quad=-[1/16(4!)]\,\mathrm{tr}(\gamma_{[k}\gamma_l\gamma_m\gamma_{n]}\gamma_5)\\
&\quad=-[1/16(4!)^2]\epsilon_{klmn}\epsilon^{qrst}\,\mathrm{tr}(\gamma_{[q}\gamma_r\gamma_s\gamma_{t]}\gamma_5)\,.
\end{aligned}\tag{3.13}$$

[In (3.13) we have deleted the tensor indices in the operations with the Dirac gammas, since multiplication of the mixed tensors $(\gamma_i)^{\sigma}{}_{\lambda}$ is just ordinary matrix algebra.]

Now, by virtue of (B43), (B46), and (B47) in Appendix B, we have

$$\begin{aligned}
&\frac{1}{(4!)^2}\epsilon^{qrst}\,\mathrm{tr}(\gamma_{[q}\gamma_r\gamma_s\gamma_{t]}\gamma_5)\\
&\quad=\sqrt{-g}\,[\epsilon^{qrst}/(4!)^2\sqrt{-g}\,]\,\mathrm{tr}(\gamma_{[q}\gamma_r\gamma_s\gamma_{t]}\gamma_5)\\
&\quad=\sqrt{-g}\,\mathrm{tr}(\gamma_5\gamma_5)=-4\sqrt{-g}\,.
\end{aligned}\tag{3.14}$$

Hence,

$$-J_k{}^{\sigma\lambda}\,J_{l\lambda\alpha}\,J_m{}^{\alpha\beta}(\tilde\gamma_5)_{\beta}{}^{\xi}J_{n\sigma\xi}=\tfrac{1}{4}\sqrt{-g}\,\epsilon_{klmn}\,. \tag{3.15}$$

By analogous arguments, we find that

$$\begin{aligned}
&-I^{\sigma\alpha}J_{k\alpha\beta}+J_m{}^{\beta\lambda}(\tilde\gamma_5)_{\lambda}{}^{\xi}J_{n\sigma\xi}\\
&\quad=\mp\tfrac{1}{8}[\,I^{\sigma\alpha}O_{\alpha\beta}(\gamma_k)^{\beta}{}_{\lambda}(\gamma_m)^{\lambda}{}_{\kappa}(\gamma_n)^{\kappa}{}_{\sigma}\,]=0\,.
\end{aligned}\tag{3.16}$$

Consequently, the second term in (3.10), which contains the torsion explicitly, also vanishes.

Taking into account (3.8), (3.9), and (3.11)–(3.16), Eq. (3.10) reduces to

$$\begin{aligned}
&(\bar J^S)_{n\Sigma\Xi}(\mathfrak{S}^S)_{ij}{}^{\Sigma}{}_{\Pi}(J^S)_m{}^{\Pi\Gamma}(\tilde\gamma_5)_{\Gamma}{}^{\Xi}\\
&\quad=+\tfrac{1}{4}\sqrt{-g}\,\epsilon_{klmn}R_{ij}{}^{kl}\\
&\qquad-(i/2)(\mathscr{D}_{[i}\bar\psi_{j]\sigma})(\gamma_m)^{\sigma}{}_{\lambda}\psi_n{}^{\lambda}\\
&\qquad-(i/2)(\mathscr{D}_{[i}\psi_{j]}^{\sigma})(\tilde\gamma_n)_{\sigma}{}^{\lambda}\bar\psi_{m\lambda}\,,
\end{aligned}\tag{3.17}$$

where we have also used the relations [see Eqs. (B28) and (B29) in Appendix B]

$$(\tilde\gamma_i)_{\sigma}{}^{\lambda}=-C_{\sigma\alpha}(\gamma_i)^{\alpha}{}_{\beta}C^{\beta\lambda}\,, \tag{3.18}$$

$$(\tilde\gamma_5)_{\sigma}{}^{\lambda}=C_{\sigma\alpha}(\gamma_5)^{\alpha}{}_{\beta}C^{\beta\lambda}\,. \tag{3.19}$$

Note that in (3.17) all twistor indices are already contracted. Thus, in order to get a scalar density Lagrangian, we have only to contract on the vector indices in the contangent bundle. This can be readily accomplished by multiplication with the tensor element of volume $d\Omega^{ijmn}=d^4x\,\epsilon^{ijmn}$. The result is

$$\begin{aligned}
\mathscr{L}\,d^4x&=d\Omega^{ijmn}(\bar J^S)_{n\Sigma\Xi}(\mathfrak{S}^S)_{ij}{}^{\Sigma}{}_{\Pi}(J^S)_m{}^{\Pi\Gamma}(\tilde\gamma_5)_{\Gamma}{}^{\Xi}\\
&=\{\sqrt{-g}\,R_s+(i/2)\epsilon^{ijmn}[(\mathscr{D}(_{[i}\bar\psi_{j]\sigma})(\gamma_m)^{\sigma}{}_{\lambda}\psi_n{}^{\lambda}\\
&\qquad+(\mathscr{D}_{[i}\psi_{j]}^{\sigma})(\tilde\gamma_n)_{\sigma}{}^{\lambda}\bar\psi_{m\lambda}\,]\}\,d^4x\,,
\end{aligned}\tag{3.20}$$

where R_s is the Ricci scalar $R_s \equiv R_{ij}{}^{ij}$.

Using the fact that the ordinary complex conjugate of a scalar formed from supertwistors is obtained by taking the conjugate of each of the constituent twistors [cf. Eq. (2.39), and (2.40) in I], and also using the properties of the Dirac gammas, it can be easily checked that the Lagrangian density given by (3.20) is real, as required.

We can compare (3.20) with the usual form given in the literature[11] in terms of Majorana spinors, by noting that

$$(\psi^c)_i{}^\sigma = C^{\sigma\alpha}\bar{\psi}_{i\alpha} = (O^{\sigma\alpha} - I^{\sigma\alpha})\bar{\psi}_{i\alpha}\,, \tag{3.21a}$$

$$\psi_i{}^\sigma = (\psi^{cc})_i{}^\sigma\,, \tag{3.21b}$$

and that

$$(\psi^M)_i{}^\sigma \equiv \psi_i{}^\sigma + (\psi^c)_i{}^\sigma \tag{3.22}$$

is a Majorana spinor.

Also note that from (3.21a)

$$\psi_i{}^\sigma \in \mathscr{G}_o \otimes \mathscr{S}_I \rightarrow (\psi^c)_i{}^\sigma \in \mathscr{G}_0 \otimes \mathscr{S}_O\,,$$

and

$$\psi_i{}^\sigma \in \mathscr{G}_o \otimes \mathscr{S}_O \rightarrow (\psi^c)_i{}^\sigma \in \mathscr{G}_o \otimes \mathscr{S}_I\,.$$

Therefore,

$$\begin{aligned}\mathscr{L}_g &\equiv \epsilon^{ijmn}[(\mathscr{D}_{[i}\bar{\psi}_{j]\sigma})(\gamma_m)^\sigma{}_\lambda \psi_n{}^\lambda + (\mathscr{D}_{[i}\psi^\sigma_{j]})(\tilde{\gamma}_n)_\sigma{}^\lambda \bar{\psi}_{m\lambda}]\\ &= \epsilon^{ijmn}\{[(\mathscr{D}_{[i}(\psi^c)^\sigma_{j]}](\tilde{\gamma}_m)_\sigma{}^\beta(\bar{\psi}^c)_{n\beta}\\ &\quad - [\mathscr{D}_{[i}\psi^\sigma_{j]}]\,(\tilde{\gamma}_m)_\sigma{}^\lambda \bar{\psi}_{n\lambda}\}\end{aligned} \tag{3.23}$$

after making use of (3.18) and taking into account that $\mathscr{D}_i C^{\alpha\beta} = 0$, because of Eqs. (2.21), (2.22), and (2.29).

But, since $\mathscr{D}_{[i}(\psi^c)^\sigma_{j]} \in \mathscr{G}_o \otimes \mathscr{S}_O$, and $(S_O)^\alpha{}_\beta = O^{\alpha\gamma}I_{\beta\gamma}$ is the projection operator onto $\mathscr{S}_O$, and $(S_I)^\alpha{}_\beta = I^{\alpha\gamma}O_{\beta\gamma}$ is the projection operator onto $\mathscr{S}_I$, we can write

$$\mathscr{D}_{[i}(\psi^c)^\sigma_{j]} = O^{\sigma\alpha}I_{\beta\alpha}\mathscr{D}_{[i}(\psi^c)^\beta_{j]} = [\mathscr{D}_{[i}(\psi^c)^\beta_{j]}]\,I_{\beta\alpha}\,O^{\sigma\alpha}\,, \tag{3.24}$$

$$\begin{aligned}&(S_I)^\sigma{}_\beta\,[\mathscr{D}_{[i}(\psi^c)^\beta_{j]}]\\ &\quad = 0 = I^{\sigma\alpha}O_{\beta\alpha}[\mathscr{D}_{[i}(\psi^c)^\beta_{j]}] = [\mathscr{D}_{[i}(\psi^c)^\beta_{j]}]O_{\beta\alpha}\,I^{\sigma\alpha}\,.\end{aligned} \tag{3.25}$$

Substracting (3.25) from (3.24) and utilizing (3.8), results in

$$\mathscr{D}_{[i}(\psi^c)^\sigma_{j]} = -i\,[\mathscr{D}_{[i}(\psi^c)^\beta_{j]}](\tilde{\gamma}_5)_\beta{}^\sigma\,. \tag{3.26}$$

By similar arguments we get

$$\mathscr{D}_{[i}\psi^\sigma_{j]} = i\,[\mathscr{D}_{[i}\psi^\beta_{j]}](\tilde{\gamma}_5)_\beta{}^\sigma\,. \tag{3.27}$$

Consequently, substituting (3.26) and (3.27) in (3.23) gives

$$\begin{aligned}\mathscr{L}_g = -i\epsilon^{ijmn}\{&[\mathscr{D}_{[i}(\psi^c)^\sigma_{j]}]\,(\tilde{\gamma}_5)_\sigma{}^\lambda(\tilde{\gamma}_m)_\lambda{}^\beta(\bar{\psi}^c)_{n\beta}\\ &+ [\mathscr{D}_{[i}\psi^\sigma_{j]}]\,(\tilde{\gamma}_5)_\sigma{}^\lambda(\tilde{\gamma}_m)_\lambda{}^\beta\bar{\psi}_{n\beta}\}\,.\end{aligned} \tag{3.28}$$

Furthermore, since

$$[\mathscr{D}_{[i}(\psi^c)^\sigma_{j]}]\,(\tilde{\gamma}_5)_\sigma{}^\lambda(\tilde{\gamma}_m)_\lambda{}^\beta\bar{\psi}_{n\beta} = 0 \tag{3.29}$$

and

$$[\mathscr{D}_{[i}\psi^\sigma_{j]}]\,(\tilde{\gamma}_5)_\sigma{}^\lambda(\tilde{\gamma}_m)_\lambda{}^\beta(\bar{\psi}^c)_{n\beta} = 0\,, \tag{3.30}$$

we can add these two noncontributing terms to (3.28) to get

$$\begin{aligned}\mathscr{L}_g &= -i\epsilon^{ijmn}\{[\mathscr{D}_{[i}(\psi^M)^\sigma_{j]}](\tilde{\gamma}_5)_\sigma{}^\lambda(\tilde{\gamma}_m)_\lambda{}^\beta(\bar{\psi}^M)_{n\beta}\}\\ &= i\epsilon^{ijmn}\{(\bar{\psi}^M)_{n\beta}(\gamma_m)^\beta{}_\lambda(\gamma_5)^\lambda{}_\sigma\mathscr{D}_{[i}(\psi^M)^\sigma_{j]}\}\\ &= -i\epsilon^{ijmn}\{(\bar{\psi}^M)_{n\beta}(\gamma_5)^\beta{}_\lambda(\gamma_m)^\lambda{}_\sigma\mathscr{D}_{[i}(\psi^M)^\sigma_{j]}\}\,.\end{aligned} \tag{3.31}$$

Finally, replacing this expression in (3.20), yields

$$\begin{aligned}\mathscr{L}\,d^4x = \{&\sqrt{-g}\,R_s - \tfrac{1}{2}\epsilon^{ijmn}\\ &\times[(\bar{\psi}^M)_{n\beta}(\gamma_5)^\beta{}_\lambda(\gamma_m)^\lambda{}_\sigma\mathscr{D}_{[i}(\psi^M)^\sigma_{j]}]\}\,d^4x\,.\end{aligned} \tag{3.32}$$

Converted to ordinary Dirac bispinors and matrices, Eq. (3.32) reads

$$\mathscr{L}\,d^4x = \{\sqrt{-g}\,R_s + (i/2)\,\epsilon^{ijmn}\bar{\psi}^M_n\gamma_5\gamma_m\mathscr{D}_{[i}\psi^M_{j]}\}\,d^4x\,, \tag{3.32'}$$

where $\psi_j{}^M$ is a one column Majorana bispinor and $\bar{\psi}_n^M$ is the adjoint Majorana bispinor defined, as usual, by $\bar{\psi}_n^M = (\psi_n^M)^\dagger\gamma^0$.

Aside from dimensionality factors, which can be readily made explicit, Eq. (3.32) is in agreement with the first order formulation of simple supergravity as it is usually presented in the literature.[5,11]

Now we will discuss variational principles. A few remarks concerning the derivation of field equations from our first-order Lagrangians seem to be in order. These equations of motion are obtained in our formalism by noting that the fundamental (gauge) quantities to be varied in an action principle are the connections D^S_X.

On the other hand, by virtue of (2.14) and (2.15), variation of D^S_X is equivalent to independent variations of D_X and the gravitino field $\psi_x{}^\sigma$.

Furthermore, since any two linear connections may differ only by a linear transformation, we have

$$(\delta D_X)z^\sigma = \delta M_x{}^\sigma{}_\gamma z^\gamma\,,\quad \forall z^\sigma(q) \in (\mathscr{V}_{2,2})_q\,, \tag{3.33}$$

where $\delta\mathbf{M}(q) \in \mathscr{T}'_q \otimes (\mathscr{V}_{2,2})_q \otimes (\mathscr{V}'_{2,2})_q$. In order to obtain additional insight into the structure of $\delta\mathbf{M}_x$, we note that

$$\begin{aligned}\delta[X(\bar{z}^\sigma w_\sigma)] = 0 &= [(\delta D_X)\bar{z}_\sigma]w^\sigma + \bar{z}_\sigma[(\delta D_X)w^\sigma]\\ &= \bar{z}_\sigma[\delta M_x{}^\sigma{}_\gamma + (\delta M^\dagger_x)^\sigma{}_\gamma]w^\gamma\,,\end{aligned} \tag{3.34}$$

i.e., $\delta\mathbf{M}_x$ has to fulfill the same requirements as the infinitesimal generators of Poincaré transformations. Hence, the most general form of $\delta\mathbf{M}_x$ must be given as a linear combination of the generators of translations and Lorentz transformations.

Recalling the expressions for the generators of translations and Lorentz transformation given in Eqs. (3.12) and (3.16) of I, we can write in general

$$\delta M_x{}^\sigma{}_\gamma = \delta_1 M_x{}^\sigma{}_\gamma + \delta_2 M_x{}^\sigma{}_\gamma\,, \tag{3.35}$$

where

$$\delta_1\mathbf{M}(q) \in \mathscr{G}_e \otimes \mathscr{T}'_q \otimes [(\mathscr{S}_I)_q \otimes_{\mathrm{aH}} (\mathscr{S}_{\bar{I}})_q]\,, \tag{3.36}$$

$$\delta_2\mathbf{M}(q) \in \mathscr{G}_e \otimes \mathscr{T}'_q \otimes [(\mathscr{S}_I \otimes \mathscr{S}_{\bar{O}}) \otimes_{\mathrm{aH}} (\mathscr{S}_O \otimes \mathscr{S}_{\bar{I}})]_q\,, \tag{3.37}$$

and $\mathscr{S}_{\bar{I}}$ and $\mathscr{S}_{\bar{O}}$ are subspaces of $\mathscr{U}'_{2,2}$, conjugate to the Weyl spinor spaces $\mathscr{S}_I$ and $\mathscr{S}_O$, respectively. That is, the variation of D_X acting on a twistor is made up of the two independent variations $\delta_1 M_x{}^\sigma{}_\gamma$ and $\delta_2 M_x{}^\sigma{}_\gamma$.

In addition, when D_X acts on a real twistor $V^{\alpha\beta} \in \mathscr{G}_e \otimes \mathscr{E}$ we have

$$(\delta D_X)V^{\alpha\beta} = \delta B_x{}^{\alpha\beta}{}_{\gamma\delta} V^{\gamma\delta}, \tag{3.38}$$

where $\delta \mathbf{B}(q) \in \mathscr{G}_e \otimes \mathscr{T}'_q \otimes \mathscr{E}_q \otimes \mathscr{E}_q$.

Also, from the compatibility with the inner product we get

$$\delta[X(V^{\alpha\beta}W_{\alpha\beta})] = 0 = V^{\alpha\beta}[\delta B_{x\alpha\beta\gamma\delta} + \delta B_{x\gamma\delta\alpha\beta}]\, W^{\gamma\delta},$$

i.e,

$$\delta B_{x\alpha\beta\gamma\delta} = -\delta B_{x\gamma\delta\alpha\beta}. \tag{3.39}$$

It follows then from (3.39) that $\delta \mathbf{B}(q) \in \mathscr{G}_e \otimes \mathscr{T}'_q \otimes (\mathscr{E}_q \wedge \mathscr{E}_q)$. Moreover, from

$$(\delta D_X)I^{\alpha\beta} = 0 = \delta B_x{}^{\alpha\beta}{}_{\gamma\delta} I^{\gamma\delta}, \tag{3.40a}$$

$$\delta[X(I^{\alpha\beta}V_{\alpha\beta})] = 0 = I^{\alpha\beta}\delta B_{x\alpha\beta\gamma\delta} V^{\gamma\delta} \Rightarrow I^{\alpha\beta}\delta B_{x\alpha\beta\gamma\delta} = 0, \tag{3.40b}$$

$$2(D_X O_{\alpha\beta})O^{\alpha\beta} = \delta[X(O^{\alpha\beta}O_{\alpha\beta})] = 0$$
$$= 2O^{\alpha\beta}\delta B_{x\alpha\beta\gamma\delta} O^{\gamma\delta} = 2O^{\alpha\beta}\delta J_{x\alpha\beta}, \tag{3.40c}$$

it is easy to conclude that $\delta B_{i\alpha\beta\gamma\delta}$ has to be further restricted to the form

$$\delta B_{i\alpha\beta\gamma\delta} = \tfrac{1}{2}(\delta N)_i{}^j (J_{j\alpha\beta} I_{\gamma\delta} - I_{\alpha\beta} J_{j\gamma\delta}) + (\delta K)_i{}^{kl}(J_{k\alpha\beta} J_{j\gamma\delta}), \tag{3.41}$$

where

$$(\delta K)_i{}^{kl} = -(\delta K)_i{}^{lk}. \tag{3.42}$$

Equation (3.41), which was derived under the assumption that D_X acted on real twistors, applies equally well to the variation of the connection acting on an arbitrary twistors valued in $\mathscr{V}^{\wedge 2}_{2,2}$, since the latter can always be expressed as a sum of elements in $\mathbb{C}\,\mathscr{E}$.

We can now relate the $\delta\mathbf{B}$ and $\delta\mathbf{M}$ variations by noting that

$$(\delta D_X)(u^\alpha v^\beta - v^\alpha u^\beta)$$
$$= \delta B_x{}^{\alpha\beta}{}_{\gamma\delta}(u^\gamma v^\delta - v^\gamma u^\delta)$$
$$= (\delta M_x{}^\alpha{}_\gamma E^\beta{}_\delta + E^\alpha{}_\gamma \delta M_x{}^\beta{}_\delta)(u^\gamma v^\delta - v^\gamma u^\delta). \tag{3.43}$$

Taking into account the antisymmetry of each of the tensor files in $\delta\mathbf{B}_x$, we can then write

$$\delta B_x{}^{\alpha\beta}{}_{\gamma\delta} = \tfrac{1}{4}(\delta M_{x[\gamma}^{\;[\alpha} E^{\beta]}{}_{\delta]} + E^{[\alpha}{}_{[\gamma}\delta M_x{}^{\beta]}{}_{\delta]}). \tag{3.44}$$

Contracting (3.44) on the second and fourth tensor indices yields

$$\delta B_x{}^{\alpha\beta}{}_{\gamma\beta} = \delta M_x{}^\alpha{}_\gamma + \tfrac{1}{2}\delta M_x{}^\beta{}_\beta E^\alpha{}_\gamma, \tag{3.45}$$

where $\delta M_x{}^\beta{}_\beta$ is obtained by one additional contraction in (3.45), and is given by

$$\delta M_x{}^\alpha{}_\alpha = \tfrac{1}{3}\delta B_x{}^{\alpha\beta}{}_{\alpha\beta}. \tag{3.46}$$

Substituting the first term in (3.41) into (3.45) and (3.46) and using a holonomic basis results in

$$\delta_1 M_i{}^\alpha{}_\gamma = -\tfrac{1}{2}(\delta N)_i{}^j (J_j{}^{\alpha\beta} I_{\beta\gamma} - I^{\alpha\beta} J_{j\beta\gamma}). \tag{3.47}$$

Similarly, from the second term in (3.41) we get

$$\delta_2 M_i{}^\alpha{}_\gamma = -(\delta K)_i{}^{kl}(J_k{}^{\alpha\beta} J_{l\beta\gamma}). \tag{3.48}$$

Let us now consider the different quantities in (3.32) which are affected by these two independent variations.

First we have

$$\delta_1 J_i{}^{\alpha\beta} = (\delta_1 D_i)O^{\alpha\beta} = \delta_1 M_i{}^\alpha{}_\gamma O^{\gamma\beta} + \delta_1 M_i{}^\beta{}_\gamma O^{\alpha\gamma}$$
$$= -(\delta N)_i{}^j [O^{\alpha\gamma} I_{\gamma\lambda} J_j{}^{\lambda\beta} + I^{\alpha\gamma} O_{\gamma\lambda} J_j{}^{\lambda\beta}]$$
$$= (\delta N)_i{}^j [(S_O)^\alpha{}_\lambda + (S_I)^\alpha{}_\lambda]\, J_j{}^{\lambda\beta}$$
$$= (\delta N)_i{}^j J_j{}^{\alpha\beta}. \tag{3.49}$$

In analogy,

$$\delta_2 J_i{}^{\alpha\beta} = \delta_2 M_i{}^\alpha{}_\gamma O^{\gamma\beta} + \delta_2 M_i{}^\beta{}_\gamma O^{\alpha\gamma}$$
$$= -(\delta K)_i{}^{kl}(J_k{}^{\alpha\lambda} J_{l\lambda\gamma} O^{\gamma\beta} - J_k{}^{\beta\lambda} J_{l\lambda\gamma} O^{\gamma\alpha})$$
$$= -(\delta K)_i{}^{kl}(O^{\alpha\lambda} J_{k\lambda\gamma} J_l{}^{\gamma\beta} + O^{\alpha\gamma} J_{l\gamma\lambda} J_k{}^{\lambda\beta}) = 0 \tag{3.50}$$

[due to (3.42)].

In arriving at (3.49) and (3.50) we have used repeatedly the identity (A1) given in Appendix A.

Note that because of (3.11), the variation of the Dirac γ_i's is determined by the variation of $J_i{}^{\alpha\beta}$. However, since γ_5 is defined according to (3.8), it will not vary.

Other quantities that need to be varied in (3.32) are the exterior derivative connections $\mathscr{D}$. To obtian an expression for $(\delta\mathscr{D}_{[i})\psi^\sigma_{j]}$, we first need to evaluate δD^0_i. This follows from (2.17)

$$(\delta D^0_X)u^\sigma = (\delta D_X)u^\sigma + \delta J_x{}^{\sigma\gamma} I_{\gamma\lambda} u^\lambda,$$

i.e.,

$$\delta \mathring{M}_x{}^\alpha{}_\gamma = \delta M_x{}^\alpha{}_\gamma + \delta J_x{}^{\alpha\lambda} I_{\gamma\lambda}. \tag{3.51}$$

From (3.47)–(3.50) we have, in particular,

$$\delta_1 \mathring{M}_x{}^\alpha{}_\gamma = 0, \tag{3.52a}$$

and

$$\delta_2 \mathring{M}_x{}^\alpha{}_\gamma = \delta_2 M_x{}^\alpha{}_\gamma. \tag{3.52b}$$

Consequently, making use of (2.32) we get

$$(\delta_1 \mathscr{D}_{[i})\psi^\sigma_{j]} = (\delta_1 \mathring{M}_{[i|}{}^\sigma{}_{\gamma|})\psi^\gamma_{j]} = 0, \tag{3.53}$$

$$(\delta_2 \mathscr{D}_{[i})\psi^\sigma_{j]} = (\delta_2 \mathring{M}_{[i|}{}^\sigma{}_{\gamma|})\psi^\gamma_{j]} = -(\delta K)_{[i}{}^{kl} J_{|k}{}^{\sigma\beta} J_{l|\beta\gamma}\psi^\gamma_{j]}. \tag{3.54}$$

Next consider

$$\delta g = g g^{ij}\delta g_{ij} = g g^{ij}\delta(J_{i\alpha\beta} J_j{}^{\alpha\beta}).$$

From (3.49) and (3.50), it follows immediately that

$$\delta_2 g = 0, \quad \delta_1 g = 2g g^{ij}(\delta N)_{ij}. \tag{3.55}$$

The last quantity that it is necessary to vary in (3.32) is the Ricci scalar R_s. Most of the work needed to obtain δR_s has already been carried out in Ref. 6. Thus, we have only to substitute (3.41), (3.49), and (3.50) into the expression for δR_s given by Eq. (4.30) in that paper. After some fairly straightforward calculations we get

$$\delta_1 R_s = -2(\delta_1 J_l{}^{\alpha\beta}) J_{m\alpha\beta} g^{mj}\delta^i_k R_{ij}{}^{kl}$$
$$= -2(\delta N)_l{}^n g_{nm} g^{mn} R_{ij}{}^{il}$$
$$= -2(\delta N)^{lj} R_{jl}, \tag{3.56}$$

$$\delta_2 R_s = [T_{kl}{}^i + 2T_{lm}{}^m \delta^i_k](\delta K)_i{}^{kl}, \tag{3.57}$$

where R_{jl} is the nonsymmetric Ricci tensor and $T_{kl}{}^i$ is the torsion tensor in the tangent bundle.

Making use of (3.49), (3.50), and (3.53)–(3.57) in an action principle derived from (3.32) leads to the usual expressions for the field equations of simple supergravity.

As a final remark note that from the definition of the torsion

$$T_{ijk} = [(D_i D_j - D_j D_i) O^{\alpha\beta}] J_{k\alpha\beta} = [D_i J_j{}^{\alpha\beta} - D_j J_i{}^{\alpha\beta}] J_{k\alpha\beta}, \tag{3.58}$$

we get

$$\delta_2 T_{ijk} = [\delta_2 B_i{}^{\alpha\beta}{}_{\gamma\delta} J_j{}^{\gamma\delta} - \delta_2 B_j{}^{\alpha\beta}{}_{\gamma\delta} J_i{}^{\gamma\delta}] J_{k\alpha\beta} = [(\delta K)_{ikj} - (\delta K)_{jki}]. \tag{3.59}$$

Recalling (3.42) and the antisymmetry of T_{ijk} in the first two indices, we can invert (3.59) to get

$$(\delta K)_{ijk} = \tfrac{1}{2}[\delta_2 T_{ikj} - \delta_2 T_{kji} - \delta_2 T_{ijk}]. \tag{3.60}$$

The term in the right of (3.60) may be recognized as the variation of the contorsion tensor. Thus, we see that the variation $\delta_2 \mathbf{M}$ is equivalent to the variation of the contorsion in the tangent bundle.

IV. SUMMARY AND CONCLUSIONS

We have proposed a formalism for the gauging of the super-Poincaré algebra which allows one to obtain first-order Lagrangians for supergravity. Our theory, which uses supertwistors as a representation space for the construction of a typical fiber, does not suffer from the conceptual problems which characterize some of the approaches followed in the literature and which are based on a direct attempt to extend to noncompact groups the Utimaya[2] procedure for gauging internal groups.

The above is made possible by means of the basic idea of treating the super-Poincaré group as an internal group. However, the theory differs from internal group theories in some significant aspects.

The first difference is that no metric structure or connection on the tangent bundle is assumed. In typical gauge theories the metric structure of the tangent bundle is given *a priori* together with a connection compatible with the metric. Nevertheless, a natural isomorphism can be achieved by selecting a given origin twistor field and introducing its covariant derivative $(J_i^{S})^{\Sigma\Gamma} = D_i^{S} O^{\Sigma\Gamma}$, as a means to map structures originating in the fibers onto the tangent bundle, inducing in it a metric and connection [cf. Eqs. (3.5) and (3.6) in the text]. Furthermore, it may be shown that the selection of an origin twistor field imposes no special restriction on the theory.

Another way in which our procedure differs from the usual approach to internal gauge theories, is that there one starts with a principal bundle on which connections are defined. Gauge covariant differentiation is then specified on the associated vector bundle of this principal bundle. In our development we start with fiber bundles which may be regarded as vector bundles associated to a principal G bundle, where the structural group G is the super-Poincaré group. Note, however, that we can pass from our vector bundles to their associated frame bundles and vice versa by changing the space on which the transition functions act from the vector space to the group manifold and back.[12]

We contend also that our approach provides a means of resolving one of the central problems of gauge field theories of external symmetry groups, that is the satisfactory treatment of translations.

Another important feature of our theory is that the Lagrangian density is obtained as a functional of the supertwistor curvature. Such a curvature is constructed basically from spinorial entities, as opposed to the Riemann curvature tensor which is built up from four-vectors.

However, the supertwistor curvature may be related to the Riemann tensor, the torsion, and the gravitino field by means of Eq. (2.51) which we derived in Sec. II. The important consequence of this is that (2.51) determines completely the allowed functional form of the Lagrangian in terms of the gravitino field and the Riemann and torsion tensors in the tangent bundle. Thus our theory is gauge invariant by construction and provides a solid structure from which one can investigate all possible specific Lagrangians relating these fields in a gauge invariant fashion.

In addition, note that the covariant derivative $\mathscr{D}_i$, which is required to occur in supergravity defined according to its spin content only, is introduced in our theory in a straightforward manner through the exterior derivative connections defined in Eqs. (2.29)–(2.32) of Sec. II, and appears naturally in our supertwistor curvature tensor.

Equally natural is the incorporation of the concept of Majorana spinors in our formalism. Note, however, that both the Majorana spinors for the gravitino field and the Dirac gammas that we used are completely general, i.e., we do not introduce any specific choice of representation for them.

We remark, finally, that by enlarging the dimensionality of our supertwistor spaces extended supergravity theories can be readily accommodated within our formalism.

ACKNOWLEDGMENTS

This work was supported in part by International Scientific Program Grants, National Science Foundation No. OIP75-09783A01 and Consejo Nacional de Ciencia y Tecnología No. PCAIEUA-790316.

APPENDIX A: A TWISTOR IDENTITY

A useful tensor identity involving contractions of skew symmetric twistors is proved here. This identity is applied repeatedly throughout the text.

Theorem: For $A^{\mu\nu}, B^{\mu\nu} \in \mathscr{U}_{2,2}^{\wedge 2}$,

$$A^{\mu\lambda} B_{\lambda\nu} + B^{\mu\lambda} A_{\lambda\nu} = -\tfrac{1}{2} A^{\alpha\beta} B_{\alpha\beta} \delta^{\mu}_{\nu}. \tag{A1}$$

Proof: Let

$$M^{\mu}{}_{\nu\alpha\beta\gamma\lambda} = \delta^{\mu}_{\nu} \bar{\eta}_{\alpha\beta\gamma\lambda} - \delta^{\mu}_{\alpha} \bar{\eta}_{\nu\beta\gamma\lambda} - \delta^{\mu}_{\beta} \bar{\eta}_{\alpha\nu\gamma\lambda} - \delta^{\mu}_{\gamma} \bar{\eta}_{\alpha\beta\nu\lambda} - \delta^{\mu}_{\lambda} \bar{\eta}_{\alpha\beta\gamma\nu}, \tag{A2}$$

where $\eta^{\alpha\beta\gamma\lambda}$ is the antisymmetric tensor in $\mathscr{U}_{2,2}^{\wedge 4}$ normalized as

$$\eta^{\alpha\beta\gamma\lambda} \bar{\eta}_{\alpha\beta\gamma\lambda} = 4!$$

and is used for constructing duals. Since $\mathscr{U}_{2,2}$ is four dimensional and $M^{\mu}{}_{\nu\alpha\beta\gamma\lambda}$ is antisymmetric in its five subscripts, it follows that

$$M^{\mu}{}_{\nu\alpha\beta\gamma\lambda} = 0 \,. \tag{A3}$$

By (A3) we have

$$0 = A^{\alpha\beta} B^{\gamma\lambda} M^{\mu}{}_{\nu\alpha\beta\gamma\lambda}$$
$$= 2A^{\alpha\beta} B_{\alpha\beta} \delta^{\mu}_{\nu} + 4A^{\mu\lambda} B_{\lambda\nu} + 4B^{\mu\lambda} A_{\lambda\nu} \,,$$

and (A1) follows immediately from this.

APPENDIX B: DIRAC GAMMA OPERATORS VIA TWISTOR THEORY

The space of Dirac γ operators is related by a one-to-one correspondence to Minkowski space. We shall give here an explicit construction of this relationship by means of a map between two twistor spaces. The first space, the domain of the map, is isomorphic to Minkowski space, and it is the twistor space $\mathscr{T} \equiv \mathscr{W}_O$ defined in the preceding paper. The second space, the range of the map, is isomorphic to the space of Dirac γ operators, and is a real subspace of the space of linear transformations $\mathscr{U}_{2,2} \otimes \mathscr{U}'_{2,2}$ on the space $\mathscr{U} \equiv \mathscr{U}_{2,2} \equiv (\mathscr{U}_{2,2}, \eta^{\alpha\beta\gamma\delta}, I^{\alpha\beta}, O^{\alpha\beta})$ defined in Ref. 6. Note that the space $\mathscr{U}_{2,2}$ is the twistor space, but the privileged elements $\eta^{\alpha\beta\gamma\delta}$, $I^{\alpha\beta}$, $O^{\alpha\beta}$ forming part of its structure make it isomorphic to the Dirac bispinor space.

The map which establishes the above-mentioned correspondence is the bijection $T^{\alpha\beta} \in \mathscr{U}^{\otimes 2} \to V^{\alpha}{}_{\beta} \in \mathscr{U} \otimes \mathscr{U}'$ given as

$$V^{\alpha}{}_{\beta} = L(T^{\alpha\beta}) = 2iT^{\alpha\gamma}(I_{\gamma\beta} + O_{\gamma\beta}) \,. \tag{B1}$$

The inverse map $V^{\alpha}{}_{\beta} \in \mathscr{U} \otimes \mathscr{U}' \to T^{\alpha\beta} \in \mathscr{U}^{\otimes 2}$ is

$$T^{\alpha\beta} = L^{-1}(V^{\alpha}{}_{\beta}) = \tfrac{1}{2} i V^{\alpha}{}_{\gamma}(I^{\gamma\beta} + O^{\gamma\beta}) \,. \tag{B2}$$

We now investigate some properties of this map. We shall define subspaces $\mathscr{J}$, $\mathscr{J}_5$, $\mathscr{J}_4$, $\mathscr{E}_{1,4}$, $\mathscr{E}_{1,3}$ of $\mathscr{U}^{\otimes 2}$, and subspaces $\mathscr{H}$, $\mathscr{H}_5$, $\mathscr{H}_4$, $\mathscr{D}_{1,4}$, $\mathscr{D}_{1,3}$ of $\mathscr{U} \otimes \mathscr{U}'$, which are the respective images of the above $\mathscr{U}^{\otimes 2}$ subspaces under the map L.

First define the subspaces $\mathscr{J} \subset \mathscr{U}^{\otimes 2}$ and $\mathscr{H} \subset \mathscr{U} \otimes \mathscr{U}'$ by

$$\mathscr{J} = \mathscr{U}^{\wedge 2} = \{T^{\alpha\beta} \mid T^{\alpha\beta} \in \mathscr{U}^{\otimes 2},\ T^{\beta\alpha} = -T^{\alpha\beta}\} \,, \tag{B3}$$

$$\mathscr{H} = \{V^{\alpha}{}_{\beta} \mid V^{\alpha}{}_{\beta} \in \mathscr{U} \otimes \mathscr{U}',$$
$$(\widetilde{V})_{\beta}{}^{\alpha} = -(I_{\beta\gamma} + O_{\beta\gamma}) V^{\gamma}{}_{\lambda}(I^{\lambda\alpha} + O^{\lambda\alpha})\} \,. \tag{B4}$$

It is easily shown that under the map L

$$\mathscr{J} \to \mathscr{H} = L(\mathscr{J}) \,, \tag{B5}$$

where $L(\mathscr{J})$ denotes the image of $\mathscr{J}$ under the map L, i.e., $L(\mathscr{J}) \equiv \{L(T^{\alpha\beta}) \mid T^{\alpha\beta} \in \mathscr{J}\}$.

To define the other subspaces, we shall make use of the following elements:

$$(E_6)^{\alpha\beta} = \tfrac{1}{2}(I^{\alpha\beta} + O^{\alpha\beta}) \in \mathscr{J} \,, \tag{B6a}$$

$$(E_5)^{\alpha\beta} = \tfrac{1}{2}(I^{\alpha\beta} - O^{\alpha\beta}) \in \mathscr{J} \,, \tag{B6b}$$

$$(U_6)^{\alpha}{}_{\beta} = -iE^{\alpha}{}_{\beta} = -i\,[(S_I)^{\alpha}{}_{\beta} + (S_O)^{\alpha}{}_{\beta}] \in \mathscr{H} \,, \tag{B6c}$$

$$(U_5)^{\alpha}{}_{\beta} = -i\,[(S_I)^{\alpha}{}_{\beta} - (S_O)^{\alpha}{}_{\beta}] \in \mathscr{H} \,, \tag{B6d}$$

where $E^{\alpha}{}_{\beta} = (S_I)^{\alpha}{}_{\beta} + (S_O)^{\alpha}{}_{\beta} \in \mathscr{U} \otimes \mathscr{U}'$ is the identity transformation on $\mathscr{U}$ [defined by (2.12) and (2.13) in I]. Under the map L, it follows that

$$(E_6)^{\alpha\beta} \to (U_6)^{\alpha}{}_{\beta} = L\,[(E_6)^{\alpha\beta}] \,, \tag{B7a}$$

$$(E_5)^{\alpha\beta} \to (U_5)^{\alpha}{}_{\beta} = L\,[(E_5)^{\alpha\beta}] \,. \tag{B7b}$$

Note also that

$$(E_6)^{\alpha\beta}(E_5)_{\alpha\beta} = 0 \,, \tag{B8a}$$

$$(E_6)^{\alpha\beta}(E_6)_{\alpha\beta} = 1 \,, \tag{B8b}$$

$$(E_5)^{\alpha\beta}(E_5)_{\alpha\beta} = -1 \,. \tag{B8c}$$

Now we define the subspace $\mathscr{J}_4 \subset \mathscr{J}_5 \subset \mathscr{J}$ and $\mathscr{H}_4 \subset \mathscr{H}_5 \subset \mathscr{H}$ as

$$\mathscr{J}_5 = \{T^{\alpha\beta} \mid T^{\alpha\beta} \in \mathscr{J},\ T^{\alpha\beta}(E_6)_{\alpha\beta} = 0\} \,, \tag{B9}$$

$$\mathscr{J}_4 = \{T^{\alpha\beta} \mid T^{\alpha\beta} \in \mathscr{J},\ T^{\alpha\beta}(E_6)_{\alpha\beta} = 0,\ T^{\alpha\beta}(E_5)_{\alpha\beta} = 0\} \,, \tag{B10}$$

$$\mathscr{H}_5 = \{V^{\alpha}{}_{\beta} \mid V^{\alpha}{}_{\beta} \in \mathscr{H},\ V^{\alpha}{}_{\beta}(U_6)^{\beta}{}_{\alpha} = 0\}$$
$$= \{V^{\alpha}{}_{\beta} \mid V^{\alpha}{}_{\beta} \in \mathscr{H},\ V^{\alpha}{}_{\alpha} = 0\} \,, \tag{B11}$$

$$\mathscr{H}_4 = \{V^{\alpha}{}_{\beta} \mid V^{\alpha}{}_{\beta} \in \mathscr{H},\ V^{\alpha}{}_{\beta}(U_6)^{\beta}{}_{\alpha} = 0,\ V^{\alpha}{}_{\beta}(U_5)^{\beta}{}_{\alpha} = 0\}$$
$$= \{V^{\alpha}{}_{\beta} \mid V^{\alpha}{}_{\beta} \in \mathscr{H},\ V^{\alpha}{}_{\alpha} = 0,\ V^{\alpha}{}_{\beta}(U_5)^{\beta}{}_{\alpha} = 0\} \,. \tag{B12}$$

Note that for $T^{\alpha\beta} \in \mathscr{J}$ and $V^{\alpha}{}_{\beta} = L(T^{\alpha\beta}) \in \mathscr{H}$,

$$(E_6)_{\alpha\beta} T^{\alpha\beta} = (i/4)(I_{\alpha\beta} + O_{\alpha\beta})[V^{\alpha}{}_{\gamma}(I^{\gamma\beta} + O^{\gamma\beta})]$$
$$= -(i/4)(I_{\beta\alpha} + O_{\beta\alpha}) V^{\alpha}{}_{\gamma}(I^{\gamma\beta} + O^{\gamma\beta})$$
$$= (i/4)(\widetilde{V})_{\beta}{}^{\beta} = (i/4) V^{\beta}{}_{\beta} \,. \tag{B13}$$

Consequently, under the map L we have

$$\mathscr{J}_5 \to \mathscr{H}_5 = L(\mathscr{J}_5) \,. \tag{B14}$$

In addition, we shall need the following theorem [which can be readily proven by writing $A^{\gamma\delta} = -A^{\gamma\sigma}(I_{\sigma\tau} + O_{\sigma\tau})(I^{\tau\delta} + O^{\tau\delta}) = A^{\gamma\sigma}(\widetilde{E})_{\sigma}{}^{\delta}$, and similarly for $B^{\gamma\beta}$ in the left-hand side of the twistor identity (A1) derived in Appendix A, using once more this identity to change the order of the products and finally resorting to (B13)].

Theorem B.1: For $(T_1)^{\alpha\beta}$, $(T_2)^{\alpha\beta} \in \mathscr{J}$, and $(V_1)^{\alpha}{}_{\beta} = L\,[(T_1)^{\alpha\beta}]$, $(V_2)^{\alpha}{}_{\beta} = L\,[(T_2)^{\alpha\beta}] \in \mathscr{H}$,

$$\tfrac{1}{2}[(V_1)^{\alpha}{}_{\beta}(V_2)^{\beta}{}_{\gamma} + (V_2)^{\alpha}{}_{\beta}(V_1)^{\beta}{}_{\gamma}]$$
$$-\tfrac{1}{4}[(V_1)^{\alpha}{}_{\gamma}(V_2)^{\beta}{}_{\beta} + (V_2)^{\alpha}{}_{\gamma}(V_1)^{\beta}{}_{\beta}]$$
$$= (T_1)_{\mu\nu}(T_2)^{\mu\nu} E^{\alpha}{}_{\gamma} \,. \tag{B15}$$

Observe that for $(V_1)^{\alpha}{}_{\beta}$, $(V_2)^{\alpha}{}_{\beta} \in \mathscr{H}_5$, (B15) leads to

$$\tfrac{1}{2}[(V_1)^{\alpha}{}_{\gamma}(V_2)^{\gamma}{}_{\beta} + (V_2)^{\alpha}{}_{\gamma}(V_1)^{\gamma}{}_{\beta}] = (T_1)_{\mu\nu}(T_2)^{\mu\nu} E^{\alpha}{}_{\beta} \,. \tag{B16}$$

Moreover, taking the trace of (B16) results in

$$(T_1)_{\mu\nu}(T_2)^{\mu\nu} = \tfrac{1}{4}(V_1)^{\alpha}{}_{\gamma}(V_2)^{\gamma}{}_{\alpha} \,. \tag{B17}$$

Thus for $(V_1)^{\alpha}{}_{\beta}$, $(V_2)^{\alpha}{}_{\beta} \in \mathscr{H}_5$, (B16) can be restated as

$$(V_1)^{\alpha}{}_{\gamma}(V_2)^{\gamma}{}_{\beta} + (V_2)^{\alpha}{}_{\gamma}(V_1)^{\gamma}{}_{\beta} = \tfrac{1}{2}[(V_1)^{\delta}{}_{\gamma}(V_2)^{\gamma}{}_{\delta}]\, E^{\alpha}{}_{\beta} \,. \tag{B18}$$

Note also that using $(T_2)^{\alpha\beta} = (E_5)^{\alpha\beta}$ and $(V_2)^{\alpha}{}_{\beta} = (U_5)^{\alpha}{}_{\beta}$ in (B17), leads to the result that under the map L we have

$$\mathscr{J}_4 \to \mathscr{H}_4 = L(\mathscr{J}_4) \,. \tag{B19}$$

Next define the real subspaces $\mathscr{E}_{2,4} \subset \mathscr{J}$, $\mathscr{E}_{1,4} \subset \mathscr{J}_5$,

$\mathscr{E}_{1,3} \subset \mathscr{J}_4$, $\mathscr{D}_{1,4} \subset \mathscr{H}_5$, and $\mathscr{D}_{1,3} \subset \mathscr{H}_4$ as

$$\mathscr{E}_{2,4} = \{ T^{\alpha\beta} | T^{\alpha\beta} \in \mathscr{J}, \; \overline{T}^{\alpha\beta} = T^{\alpha\beta} \}, \tag{B20}$$

$$\mathscr{E}_{1,4} = \{ T^{\alpha\beta} | T^{\alpha\beta} \in \mathscr{J}_5, \; \overline{T}^{\alpha\beta} = T^{\alpha\beta} \}, \tag{B21}$$

$$\mathscr{E}_{1,3} \equiv \mathscr{F} = \{ T^{\alpha\beta} | T^{\alpha\beta} \in \mathscr{J}_4, \; \overline{T}^{\alpha\beta} = T^{\alpha\beta} \}, \tag{B22}$$

$$\mathscr{D}_{1,4} = \{ V^\alpha{}_\beta | V^\alpha{}_\beta \in \mathscr{H}_5, \; \overline{\overline{V}}{}^\alpha{}_\beta = V^\alpha{}_\beta \}, \tag{B23}$$

$$\mathscr{D}_{1,3} = \{ V^\alpha{}_\beta | V^\alpha{}_\beta \in \mathscr{H}_4, \; \overline{\overline{V}}{}^\alpha{}_\beta = V^\alpha{}_\beta \}. \tag{B24}$$

The space $\mathscr{E}_{1,3}$ is Minkowski space, and $\mathscr{D}_{1,3}$ is the space of Dirac gamma operators. In order to relate these two spaces under the map L we need the following theorems.

Theorem B.2: For $T^{\alpha\beta} \in \mathscr{J}$,

$$T^{\alpha\beta} \in \mathscr{J}_5 \Leftrightarrow \overline{T}^{\alpha\beta} \in \mathscr{J}_5, \tag{B25}$$

and for $V^\alpha{}_\beta \in \mathscr{H}$,

$$V^\alpha{}_\beta \in \mathscr{H}_5 \Leftrightarrow \overline{\overline{V}}{}^\alpha{}_\beta \in \mathscr{H}_5. \tag{B26}$$

Relation (B25) follows directly from

$$\overline{T}^{\alpha\beta}(E_6)_{\alpha\beta} = \overline{T}_{\alpha\beta}(E_6)^{\alpha\beta} = [T^{\alpha\beta}(E_6)_{\alpha\beta}]^*,$$

where we have taken into account that $(E_6)^{\alpha\beta}$ is a real twistor.

To arrive at (B26) note that

$$\overline{\overline{V}}{}^\alpha{}_\alpha = (V^\alpha{}_\alpha)^*.$$

Theorem B.3: Under the map L restricted to the domain $\mathscr{J}_5$, i.e., $T^{\alpha\beta} \in \mathscr{J}_5 \rightarrow V^\alpha{}_\beta = L(T^{\alpha\beta}) \in \mathscr{H}_5$, we have

$$\overline{T}^{\alpha\beta} \in \mathscr{J}_5 \rightarrow \overline{\overline{V}}{}^\alpha{}_\beta = L(\overline{T}^{\alpha\beta}) \in \mathscr{H}_5, \tag{B27}$$

and, consequently, for $T^{\alpha\beta} \in \mathscr{J}_5$ we have

$$\overline{T}^{\alpha\beta} = T^{\alpha\beta} \Leftrightarrow \overline{\overline{V}}{}^\alpha{}_\beta = V^\alpha{}_\beta. \tag{B28}$$

To prove this theorem we make use of the identity (A1) in order to show that

$$2i\overline{T}^{\alpha\beta}(I_{\beta\gamma} + O_{\beta\gamma}) = 2i(I_{\gamma\beta} + O_{\gamma\beta})\overline{T}^{\beta\alpha} = -2i\overline{T}_{\gamma\beta}(I^{\beta\alpha} + O^{\beta\alpha}) = \overline{\overline{V}}{}^\alpha{}_\gamma. \tag{B29}$$

Using Theorems B.2 and B.3, one can easily show that under the map L we have

$$\mathscr{E}_{1,4} \rightarrow \mathscr{D}_{1,4} = L(\mathscr{E}_{1,4}), \tag{B30}$$

$$\mathscr{E}_{1,3} \rightarrow \mathscr{D}_{1,3} = L(\mathscr{E}_{1,3}). \tag{B31}$$

Equation (B31) expresses the result that the map L establishes a one-to-one correspondence between Minkowski space $\mathscr{E}_{1,3}$ and the space $\mathscr{D}_{1,3}$ of Dirac gamma operators.

Combining the preceding results we can finally arrive at the following theorem.

Theorem B.4: Suppose $(W_5)^\beta{}_\alpha \in \mathscr{U} \otimes \mathscr{U}'$ has the following properties:

$$(\widetilde{W}_5)_\alpha{}^\beta = -(I_{\alpha\gamma} + O_{\alpha\gamma})(W_5)^\gamma{}_\lambda(I^{\lambda\beta} + O^{\lambda\beta}), \tag{B32a}$$

$$(W_5)^\alpha{}_\alpha = 0, \tag{B32b}$$

$$V^\alpha{}_\gamma(W_5)^\gamma{}_\alpha = 0, \quad \text{for all } V^\alpha{}_\gamma \in \mathscr{D}_{1,3}, \tag{B32c}$$

$$(\overline{\overline{W}}_5)^\alpha{}_\beta = (W_5)^\alpha{}_\beta, \tag{B32d}$$

$$(W_5)^\alpha{}_\gamma(W_5)^\gamma{}_\beta = -E^\alpha{}_\beta. \tag{B32e}$$

Then

$$(W_5)^\alpha{}_\gamma = \pm (U_5)^\alpha{}_\gamma. \tag{B33}$$

To prove this theorem, first let $(F_5)^{\alpha\beta} = (i/2)(W_5)^\alpha{}_\gamma(I^{\gamma\beta} + O^{\gamma\beta})$. Then by the map L we have $(W_5)^\alpha{}_\beta = L[(F_5)^{\alpha\beta}] = 2i(F_5)^{\alpha\gamma}(I_{\gamma\beta} + O_{\gamma\beta})$. Furthermore, by (B32a), (B32b), and (B32d) we find that $(W_5)^\alpha{}_\beta \in \mathscr{D}_{1,4} \subset \mathscr{H}_5 \subset \mathscr{H}$, and using (B13) and (B28) leads to $(F_5)^{\alpha\beta} \in \mathscr{E}_{1,4} \subset \mathscr{J}_5$. In addition, (B32c) and (B17) imply that $T_{\alpha\beta}(F_5)^{\alpha\beta} = 0$ for all $T^{\alpha\beta} \in \mathscr{E}_{1,3}$. But, by virtue of (B22) and (B10), $T^{\alpha\beta}(E_5)_{\alpha\beta} = 0$ for all $T^{\alpha\beta} \in \mathscr{E}_{1,3}$. Hence $(F_5)^{\alpha\beta}$ is in the same one-dimensional subspace in which $(E_5)^{\alpha\beta} \in \mathscr{E}_{1,4}$ lies, so $(F_5)^{\mu\nu} = \alpha(E_5)^{\mu\nu}$ for some real number α. If we now make use of (B16), we get $(W_5)^\gamma{}_\delta(W_5)^\delta{}_\lambda = (F_5)_{\mu\nu}(F_5)^{\mu\nu}E^\gamma{}_\lambda$. Substituting (B32e) on the left-hand side of this equation yields $(F_5)_{\mu\nu}(F_5)^{\mu\nu} = -1$. Also since $(E_5)_{\mu\nu}(E_5)^{\mu\nu} = -1$, we arrive at $\alpha = \pm 1$ and $(F_5)^{\alpha\beta} = \pm (E_5)^{\alpha\beta}$. Thus, $(W_5)^\alpha{}_\beta = \pm 2i(E_5)^{\alpha\gamma}(I_{\gamma\beta} + O_{\gamma\beta}) = \pm (U_5)^\alpha{}_\beta$. Q.E.D.

Having established Theorem B.4, we are now in a position to elaborate further on the relation between Minkowski space and the space of Dirac gamma operators. For this purpose let $(E_i)^{\alpha\beta} \in \mathscr{E}_{1,3}$ for $i = 0, 1, 2, 3$ be any basis for $\mathscr{E}_{1,3}$, and $(E^i)^{\alpha\beta} \in \mathscr{E}_{1,3}$ for $i = 0, 1, 2, 3$ be the corresponding reciprocal basis. Thus

$$(E^i)^{\alpha\beta}(E_j)_{\alpha\beta} = \delta^i_j. \tag{B34}$$

The components g_{ij} of the metric tensor for $\mathscr{E}_{1,3}$ are

$$g_{ij} = (E_i)^{\alpha\beta}(E_j)_{\alpha\beta}. \tag{B35}$$

Including $(E_5)^{\alpha\beta}$ with the $(E_i)^{\alpha\beta}$ given above, we have the elements $(E_A)^{\alpha\beta} \in \mathscr{E}_{1,4}$ for $A = 0, 1, 2, 3, 5$, which form a basis for $\mathscr{E}_{1,4}$. The elements $(E^A)^{\alpha\beta} \in \mathscr{E}_{1,4}$ for $A = 0, 1, 2, 3, 5$, where $(E^i)^{\alpha\beta}$ for $i = 0, 1, 2, 3$, is given above and $(E^5)^{\alpha\beta} = -(E_5)^{\alpha\beta}$ form the corresponding reciprocal basis for $\mathscr{E}_{1,4}$ since

$$(E^A)^{\alpha\beta}(E_B)_{\alpha\beta} = \delta^A_B. \tag{B36}$$

With respect to this basis the components of the metric tensor for $\mathscr{E}_{1,4}$ are

$$g_{AB} = (E_A)^{\alpha\beta}(E_B)_{\alpha\beta}, \tag{B37}$$

and we have g_{ij} the same as given above for $i, j = 0, 1, 2, 3$, and $g_{i5} = g_{5i} = 0$, $g_{55} = -1$.

The elements $(U_i)^\alpha{}_\beta = L[(E_i)^{\alpha\beta}] = 2i(E_i)^{\alpha\gamma}(I_{\gamma\beta} + O_{\gamma\beta}) \in \mathscr{D}_{1,3}$, for $i = 0, 1, 2, 3$, form a basis for $\mathscr{D}_{1,3}$. Using $\frac{1}{4}(A^\alpha{}_\gamma B^\gamma{}_\alpha)$ as the inner product for $A^\alpha{}_\beta, B^\alpha{}_\beta \in \mathscr{D}_{1,3}$, the elements $(U^i)^\alpha{}_\beta = L[(E^i)^{\alpha\beta}] = 2i(E^i)^{\alpha\gamma}(I_{\gamma\beta} + O_{\gamma\beta})$ form the corresponding reciprocal basis for $\mathscr{D}_{1,3}$ since

$$\tfrac{1}{4}(U^i)^\alpha{}_\gamma(U_j)^\gamma{}_\alpha = (E^i)^{\alpha\beta}(E_j)_{\alpha\beta} = \delta^i_j. \tag{B38}$$

Also we have

$$\tfrac{1}{4}(U_i)^\alpha{}_\beta(U_j)^\beta{}_\alpha = (E_i)^{\alpha\beta}(E_j)_{\alpha\beta} = g_{ij}. \tag{B39}$$

Similarly, the elements $(U_A)^\alpha{}_\beta = L[(E_A)^{\alpha\beta}] = 2i(E_A)^{\alpha\gamma}(I_{\gamma\beta} + O_{\gamma\beta}) \in \mathscr{D}_{1,4}$ for $A = 0, 1, 2, 3, 5$ form a basis for $\mathscr{D}_{1,4}$. Using $\frac{1}{4}(A^\alpha{}_\beta B^\beta{}_\alpha)$ as the inner product for $A^\alpha{}_\beta$, $B^\alpha{}_\beta \in \mathscr{D}_{1,4}$ the elements $(U^A)^\alpha{}_\beta = L[(E^A)^{\alpha\beta}] = 2i(E^A)^{\alpha\gamma}(I_{\gamma\beta} + O_{\gamma\beta}) \in \mathscr{D}_{1,4}$ form the corresponding reciprocal basis for $\mathscr{D}_{1,4}$, since

$$\tfrac{1}{4}(U^A)^\alpha{}_\beta(U_B)^\beta{}_\alpha = (E^A)^{\alpha\beta}(E_B)_{\alpha\beta} = \delta^A_B. \tag{B40}$$

Also we have

$$\tfrac{1}{4}(U_A)^\alpha{}_\beta(U_B)^\beta{}_\alpha = (E_A)^{\alpha\beta}(E_B)_{\alpha\beta} = g_{AB}. \tag{B41}$$

Note that $(U^5)^\alpha{}_\beta = -(U_5)^\alpha{}_\beta$.

Introducing the customary notation for the Dirac gamma operators, we have

$$(\gamma_i)^\alpha{}_\beta = (U_i)^\alpha{}_\beta, \quad \text{for } i = 0, 1, 2, 3, \tag{B42a}$$

$$(\gamma_5)^\alpha{}_\beta = (U_5)^\alpha{}_\beta, \tag{B42b}$$

$$(\gamma^i)^\alpha{}_\beta = (U^i)^\alpha{}_\beta, \quad \text{for } i = 0, 1, 2, 3, \tag{B42c}$$

$$(\gamma^5)^\alpha{}_\beta = (U^5)^\alpha{}_\beta. \tag{B42d}$$

Note that

$$(\gamma^5)^\alpha{}_\beta = -(\gamma_5)^\alpha{}_\beta. \tag{B42e}$$

Now let

$$\begin{aligned}(W_5)^\alpha{}_\beta &= (4!)^{-1} N^{ijkl} (U_i)^\alpha{}_\kappa (U_j)^\kappa{}_\lambda (U_k)^\lambda{}_\mu (U_l)^\mu{}_\beta \\ &= (4!)^{-1} N^{ijkl} (\gamma_i)^\alpha{}_\kappa (\gamma_j)^\kappa{}_\lambda (\gamma_k)^\lambda{}_\mu (\gamma_l)^\mu{}_\beta,\end{aligned} \tag{B43}$$

where N^{ijkl} are components of an antisymmetric tensor **N** in $\mathscr{E}_{1,3}^{\wedge 4}$ normalized such that

$$N^{ijkl} N_{ijkl} = -4!. \tag{B44}$$

It can be shown that $(W_5)^\alpha{}_\beta$ satisfies the properties given in Theorem B.4, therefore

$$(W_5)^\alpha{}_\beta = \pm (U_5)^\alpha{}_\beta = \pm (\gamma_5)^\alpha{}_\beta. \tag{B45}$$

The choice of **N** is unique up to a sign. Assume a choice of **N** such that Eq. (B45) has a plus sign, i.e.,

$$(W_5)^\alpha{}_\beta = (U_5)^\alpha{}_\beta = (\gamma_5)^\alpha{}_\beta. \tag{B46}$$

For this **N** we say that a basis $(E_0)^{\alpha\beta}$, $(E_1)^{\alpha\beta}$, $(E_2)^{\alpha\beta}$, $(E_3)^{\alpha\beta}$ for $\mathscr{E}_{1,3}$ is right handed if the components N^{ijkl} with respect to this basis satisfy $N^{0123} > 0$.

In this case, we have

$$N^{ijkl} = \epsilon^{ijkl}/\sqrt{-g}, \tag{B47}$$

where ϵ^{ijkl} is the Levi–Civita symbol ($\epsilon^{0123} = \epsilon_{0123} = 1$).

Suppose in addition that we have a basis $(E_i)^{\alpha\beta}$, for $\mathscr{E}_{1,3}$ which is orthonormal, i.e., such that

$$\begin{aligned}&g_{ij} = 0, \quad \text{for } i \neq j, \\ &g_{00} = -g_{11} = -g_{22} = -g_{33} = 1;\end{aligned} \tag{B48}$$

and suppose this basis is also right handed. Then it follows that

$$(\gamma_5)^\alpha{}_\beta = (\gamma_0)^\alpha{}_\kappa (\gamma_1)^\kappa{}_\lambda (\gamma_2)^\lambda{}_\mu (\gamma_3)^\mu{}_\beta. \tag{B49}$$

For convenience we summarize here some of the properties of the Dirac gamma operators which can be readily obtained from the results so far derived:

$$(\gamma_i)^\alpha{}_\beta = L\,[(E_i)^{\alpha\beta}] = 2i(E_i)^{\alpha\gamma}(I_{\gamma\beta} + O_{\gamma\beta}) \in \mathscr{D}_{1,3}, \tag{B50a}$$

$$(\tilde{\gamma}_i)_\beta{}^\alpha = -(I_{\beta\lambda} + O_{\beta\lambda})(\gamma_i)^\lambda{}_\kappa (I^{\kappa\alpha} + O^{\kappa\alpha}), \tag{B50b}$$

$$(\gamma_i)^\alpha{}_\alpha = 0, \tag{B50c}$$

$$(\gamma_i)^\alpha{}_\lambda (\gamma_5)^\lambda{}_\alpha = 0, \tag{B50d}$$

$$(\tilde{\tilde{\gamma}}_i)^\alpha{}_\beta = (\gamma_i)^\alpha{}_\beta, \tag{B50e}$$

$$(\gamma_i)^\alpha{}_\kappa (\gamma_j)^\kappa{}_\beta + (\gamma_j)^\alpha{}_\kappa (\gamma_i)^\kappa{}_\beta = \tfrac{1}{2}(\gamma_i)^\mu{}_\nu (\gamma_j)^\nu{}_\mu E^\alpha{}_\beta = 2 g_{ij} E^\alpha{}_\beta, \tag{B50f}$$

$$(\gamma_i)^\alpha{}_\kappa (\gamma_5)^\kappa{}_\beta + (\gamma_5)^\alpha{}_\kappa (\gamma_i)^\kappa{}_\beta = 0, \tag{B50g}$$

$$\begin{aligned}(\gamma_5)^\alpha{}_\beta &= L\,[(E_5)^{\alpha\beta}] = 2i(E_5)^{\alpha\gamma}(I_{\gamma\beta} + O_{\gamma\beta}) \\ &\quad - i\,[(S_I)^\alpha{}_\beta - (S_O)^\alpha{}_\beta] \in \mathscr{D}_{1,4},\end{aligned} \tag{B51a}$$

$$(\gamma_5)^\alpha{}_\beta = (\gamma_0)^\alpha{}_\kappa (\gamma_1)^\kappa{}_\lambda (\gamma_2)^\lambda{}_\mu (\gamma_3)^\mu{}_\beta \ [\text{if } (E_i)^{\alpha\beta} \text{ is a right-oriented orthonormal basis for } \mathscr{E}_{1,3}] \tag{B51b}$$

$$(\tilde{\gamma}_5)_\alpha{}^\beta = -(I_{\alpha\gamma} + O_{\alpha\gamma})(\gamma_5)^\gamma{}_\lambda (I^{\lambda\beta} + O^{\lambda\beta}), \tag{B51c}$$

$$(\gamma_5)^\alpha{}_\alpha = 0, \tag{B51d}$$

$$(\gamma_i)^\alpha{}_\beta (\gamma_5)^\beta{}_\alpha = 0 \quad (\text{for } i = 0, 1, 2, 3), \tag{B51e}$$

$$(\tilde{\tilde{\gamma}}_5)^\alpha{}_\beta = (\gamma_5)^\alpha{}_\beta, \tag{B51f}$$

$$(\gamma_5)^\alpha{}_\lambda (\gamma_5)^\lambda{}_\beta = -E^\alpha{}_\beta. \tag{B51g}$$

In the main text we have used the space $\mathscr{F} \equiv \mathscr{E}_{1,3}$ as a typical fiber of a bundle $\mathscr{F}(\mathscr{M}) \equiv \mathscr{E}_{1,3}(\mathscr{M})$. The $(E_i)^{\alpha\beta}$ (denoted there by $J_i{}^{\alpha\beta}$), $I^{\alpha\beta} + O^{\alpha\beta}$, and $(\gamma_i)^\alpha{}_\beta$ are cross sections of the bundles $\mathscr{E}_{1,3}(\mathscr{M})$, $\mathscr{E}_{2,4}(\mathscr{M})$, and $\mathscr{D}_{1,3}(\mathscr{M})$, respectively.

Taking note of this notational correspondence, (B50a) can be written as

$$(\gamma_i)^\alpha{}_\beta = 2iJ_i{}^{\alpha\gamma}(I_{\gamma\beta} + O_{\gamma\beta}), \tag{B52}$$

and its inverse, defined by (B2), is

$$J_i{}^{\alpha\beta} = (i/2)(\gamma_i)^\alpha{}_\gamma (I^{\gamma\beta} + O^{\gamma\beta}). \tag{B53}$$

Also, since $J_i{}^{\beta\alpha} = -J_i{}^{\alpha\beta}$, we have

$$J_i{}^{\alpha\beta} = (i/2)(I^{\alpha\gamma} + O^{\alpha\gamma})(\tilde{\gamma}_i)_\gamma{}^\beta. \tag{B54}$$

We now consider another map which we use extensively in the text in conjunction with the γ operators. This is the charge conjugation operator.

We can define explicitly the charge conjugation operator $C^{\alpha\beta} \in \mathscr{E}_{2,4}$ by

$$C^{\alpha\beta} = O^{\alpha\beta} - I^{\alpha\beta}, \tag{B55}$$

and for $\psi^\alpha \in \mathscr{U}$, its charge conjugate $(\psi^c)^\alpha \in \mathscr{U}$ is

$$(\psi^c)^\alpha = (O^{\alpha\beta} - I^{\alpha\beta})\bar{\psi}_\beta. \tag{B56}$$

From Eqs. (2.12) and (2.13) in I, it is trivial to show that

$$C^{\alpha\gamma} C_{\gamma\beta} = E^\alpha{}_\beta, \tag{B57a}$$

$$C_{\alpha\gamma} C^{\gamma\beta} = (\tilde{E})_\alpha{}^\beta. \tag{B57b}$$

We can now express some of the relations derived above in terms of this charge conjugation operator.

First, multiplication of (B50b) by $C^{\lambda\beta}$ on the left and by $C_{\alpha\mu}$ on the right, and the use of (B50g) and (B51a), yields

$$(\gamma_i)^\alpha{}_\beta = -C^{\alpha\lambda}(\tilde{\gamma}_i)_\lambda{}^\kappa C_{\kappa\beta}. \tag{B58a}$$

The inverse of (B58a) is

$$(\tilde{\gamma}_i)_\beta{}^\alpha = -C_{\beta\kappa}(\gamma_i)^\kappa{}_\lambda C^{\lambda\alpha}, \tag{B58b}$$

which is readily obtained from (B58a) by making use of (B57b).

By means of a similar operation on (B51c), we also get

$$(\tilde{\gamma}_5)_\alpha{}^\beta = C_{\alpha\lambda}(\gamma_5)^\lambda{}_\kappa C^{\kappa\beta}. \tag{B59}$$

Finally, note that by virtue of (B51a) and (B51f)

$$I^{\alpha\beta} + O^{\alpha\beta} = -i(\gamma_5)^\alpha{}_\lambda C^{\lambda\beta} = -iC^{\alpha\lambda}(\tilde{\gamma}_5)_\lambda{}^\beta, \tag{B60a}$$

$$I_{\alpha\beta} + O_{\alpha\beta} = i(\tilde{\gamma}_5)_\alpha{}^\lambda C_{\lambda\beta} = iC_{\alpha\lambda}(\gamma_5)^\lambda{}_\beta. \tag{B60b}$$

It follows then, from (B53) and (B60a), that

$$J_i{}^{\alpha\beta} = \tfrac{1}{2}(\gamma_i)^\alpha{}_\lambda (\gamma_5)^\lambda{}_\kappa C^{\kappa\beta} . \tag{B61}$$

Similarly, taking the conjugate of (B54) and making use of (B50e) and (B60b) results in

$$J_{i\alpha\beta} = \tfrac{1}{2}C_{\alpha\lambda} (\gamma_5)^\lambda{}_\kappa (\gamma_i)^\kappa{}_\beta = -\tfrac{1}{2}C_{\alpha\lambda} (\gamma_i)^\lambda{}_\kappa (\gamma_5)^\kappa{}_\beta . \tag{B62}$$

[1]C. P. Luehr and M. Rosenbaum, J. Math. Phys. **21**, 1432 (1980); M. Rosenbaum and C. P. Luehr, Hadronic J. **4**, 153 (1981).
[2]R. Utimaya, Phys. Rev. **101**, 1597 (1956).
[3]T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961).
[4]See also the discussion by M. A. Schweizer ,in *Cosmology and Gravitation: Spin, Torsion, Rotation, and Supergravity,* edited by P. G. Bergmann and V. De Sabbata (Plenum, New York, 1980).
[5]For a review of the literature pertinent to the subject of supergravity see, e.g., S. Deser, in *Relativity and Gravitation, Proceedings of the Third Latin-American Symposium,* edited by S. Hojman, M. Rosenbaum, and M. P. Ryan (UNAM Press, 1982); Y. Ne'eman, in *Cosmology and Gravitation: Spin, Torsion, Rotation and Supergravity,* edited by P. G. Bergmann and V. De Sabbata (Plenum, New York, 1980); P. van Nieuwenhuizen, *ibid.*; P. van Nieuwenhuizen, Phys. Rep. **68**, 189 (1981); *Supergravity,* edited by P. van Nieuwenhuizen and D. Z. Freedman (North-Holland, Amsterdam, 1979); S. Deser, Hadronic J. **4**, 19 (1981); D. Z. Freedman, in *Introduction to Supersymmetry in Particle and Nuclear Physics,* edited by O. Castaños, A. Frank, and L. Urrutia (Plenum, New York, 1984); M. T. Grisaru, *ibid.*; S. Ferrara, *ibid.*; J. Bagger and J. Wess, *Supersymmetry and Supergravity* (Princeton U.P., Princeton, NJ, 1983).
[6]C. P. Luehr and M. Rosenbaum, J. Math. Phys. **23**, 1471 (1982).
[7]See discussion in Ref. 6 and in C. P. Luehr and M. Rosenbaum, J. Math. Phys. **25**, 380 (1984).
[8]See C. P. Luehr and M. Rosenbaum, J. Math. Phys. **26**, 1828 (1985), and references therein.
[9]A. Ferber, Nucl. Phys. B **132**, 55 (1978); J. Lukierski, J. Math. Phys. **21**, 561 (1980).
[10]Although we could hope for tensors or second rank, obtained from (2.51) by contraction of the first two or last two files, these tensors turn out to be zero, as it becomes clear from a brief examination of the terms in the right side of this equation.
[11]S. Deser and B. Zumino, Phys. Lett. B **62**, 335 (1976); D. Z. Freedman and P. van Nieuwenhuizen, Phys. Rev. D **14**, 912 (1976).
[12]T. Eguchi, P. B. Gilkey, and J. Hanson, Phys. Rep. **66**, 213 (1980).

Gauge fixing and Gupta–Bleuler triplets in de Sitter QED

J. P. Gazeau[a)]
Department of Physics, University of California, Los Angeles, California 90024

(Received 26 September 1984; accepted for publication 11 January 1985)

Binegar *et al.* recently developed de Sitter QED within the frame of a particular gauge fixing $c = \frac{2}{3}$. Such a choice leads to the simplest possible structure of the Gupta–Bleuler triplets, with the SO(3,2) representations $D(3,0)$ and $D(1,1)$ appearing to describe special gauge solutions. The present work examines the general case $c \neq \frac{2}{3}$ by considering explicitly the corresponding homogeneous propagators and some fundamental carrier states for the related indecomposable representations of SO(3,2). The latter expressions involve "logarithmic" gauge states which independently carry extensions of the $D(3,0)$ and $D(1,1)$ Lie algebra representations.

I. INTRODUCTION

In a recent paper,[1] Binegar, Fronsdal, and Heidenreich have put in evidence the need for two dynamically independent potentials in de Sitter quantum electrodynamics. The existence of two independent Gupta–Bleuler triplets follows, whereas only one is required in flat space quantum electrodynamics (QED). Each de Sitter Gupta–Bleuler triplet is a certain indecomposable representation of the de Sitter group SO(3,2), where the massless irreducible representation $D(2,1)$ occupies the central part. They are obtained in Ref. 1 by restriction in conformal QED[2] to the de Sitter subgroup.

A first discontinuity with respect to the flat space limit then appears: whereas the notion of helicity is well defined in Minkowski QED, we lose its meaning in de Sitter space. Another unusual singularity occurs in the choice of the gauge, a choice which amounts to giving a precise value to a certain parameter c. In Minkowski QED, the simplest fixing is the Feynman gauge $c = 0$, while the value $c = \frac{2}{3}$ is privileged in de Sitter QED. This discontinuity is an indication of the nontrivial character of the contraction de Sitter →Poincaré with respect to massless field theories.

The aim of this work is to reexamine de Sitter QED with an arbitrary gauge fixing c (however different from 1), and study the differences with the case $c = \frac{2}{3}$; essentially, "logarithmic" states appear in the general case. Our approach, presenting itself as rather pedestrian and pedagogical, insists on the explicit construction of the homogeneous propagators for any c. The resulting expression makes transparent the *raison d'être* of the value $c = \frac{2}{3}$.

It was pointed out in Ref. 1 that there exists another formulation which allows one to overcome the difficulty due to a non-null gauge fixing: the *a priori* removal of the the transversality condition on the de Sitter fields in order to reimpose it after quantization as a Lorentz condition. Nevertheless, it is not without interest to examine these questions of quantization on a curved space by keeping to a minimum the number of field components in order to stay as close as possible to the flat space conditions.

The organization of this paper is as follows. Section II briefly reviews Minkowski QED and related notions of Gupta–Bleuler triplets and gauge fixing through the introduction of the b field. Section III recalls the de Sitter formalism as it was developed by Fronsdal.[3,4] We also write de Sitter QED equations explicitly.

Introducing a b field in both cases can appear as a useless, pedantic complication, but it is done by keeping in mind a future generalization of the present work to higher spins in the spirit of Ref. 5.

Sections IV and V are devoted to the homogeneous propagators and the SO(3,2) content of their expressions. Section VI is a review of some results of Ref. 1 obtained for the case $c = \frac{2}{3}$, and preparation for the last two sections which consider the general case $c \neq \frac{2}{3}$ and display what we call "logarithmic" states. We then examine the indecomposable representations of the de Sitter group associated with different gauge fixing c and the link that they have with the Gupta–Bleuler triplets of Ref. 1.

II. MINKOWSKI QED

A. Gauge fixing

The free field equations of electromagnetism

$$\Box A_\mu - \partial_\mu\, \partial\cdot A = 0 \tag{2.1}$$

are identically satisfied by the gauge fields

$$A_\mu = \partial_\mu A. \tag{2.2}$$

The occurrence of such a large set of solutions brings trouble when one tries to solve for the field A_μ in terms of an external source J_μ

$$\Box A_\mu - \partial_\mu\, \partial\cdot A = J_\mu,$$
$$A_\mu \equiv \Box^{-1} J_\mu \quad \text{(mod. gauge field)}. \tag{2.3}$$

Let us note that the current has to be conserved ($\partial\cdot J = 0$) consistently with (2.3).

Equation (2.1) is derived from a gauge-invariant Lagrangian density

$$\mathscr{L} = -\tfrac{1}{2}\partial^\mu A^\nu \partial_\mu A_\nu + \tfrac{1}{2}(\partial\cdot A)^2 - J\cdot A. \tag{2.4}$$

The gauge fixing can be realized by adding to $\mathscr{L}$ two terms involving the so-called b field[6]:

$$\mathscr{L} = -\tfrac{1}{2}\partial^\mu A^\nu \partial_\mu A_\nu + \tfrac{1}{2}(\partial\cdot A)^2 - b\partial\cdot A - (\alpha/2)b^2 - J\cdot A. \tag{2.5}$$

The field equations now read

[a)] Permanent address: Laboratoire de Physique Théorique et Mathématique, Université Paris 7, 75251 Paris Cedex 05, France.

$$\Box A_\mu - [(\alpha-1)/\alpha]\partial_\mu\, \partial{\cdot}A = J_\mu, \tag{2.6a}$$

$$\partial{\cdot}A = \alpha b. \tag{2.6b}$$

The simplest choice of gauge is taking $\alpha = 1$, the Feynman gauge. The inhomogeneous propagator is simply $\Box^{-1}\delta_{\mu\nu}$ and no higher poles appear.

Examining Eq. (2.6a) more closely,

$$\Box A_\mu - c\,\partial_\mu\, \partial{\cdot}A = J_\mu, \quad c = (\alpha - 1)/\alpha, \quad c \neq 1,$$

we are led to two conclusions.

(1) Since $\partial{\cdot}J = 0$, $(1-c)\Box\,\partial{\cdot}A = 0$ and hence $\partial{\cdot}A$ is a free field. It has been shown[7] that this gauge condition is the simplest one among those which are conformal invariant.

(2) The gauge fields $A_\mu = \partial_\mu \Lambda$ satisfy the free field equations only if $\Box\Lambda = 0$.

B. Gupta–Bleuler triplets

Quantization of free fields is usually carried out after choosing the Feynman gauge $c = 0$ and we are led to consider the space V' of the positive-energy solutions of

$$\Box A_\mu = 0, \tag{2.7a}$$

$$\|A\|^2 \equiv -\int d^3\vec{x}\, A^{\mu *} i \overleftrightarrow{\partial}_\mu A_\mu < \infty. \tag{2.7b}$$

The form (2.7b) is indefinite. It will be positive semidefinite if we restrict ourselves to the subspace $V \subset V'$ of vector fields satisfying the Lorentz condition $\partial{\cdot}A = 0$, and the radical of (2.7b) is precisely the set $V_g \subset V$ of gauge fields $A_\mu = \partial_\mu \Lambda$, $\Box\Lambda = 0$.

We thus see a chain appearing:

$$V_g \subset V \subset V'.$$

The form (2.7b) is definite positive on the quotient space V/V_g: it acquires there the status of a norm and V/V_g carries the physical content of the theory. It is the space of physical states or transverse photons, while V_g is the space of longitudinal photons. Finally the scalar photons form the quotient space V'/V, whose elements are associated with divergences $\partial{\cdot}A$.

A subspace V'^c of the space of solutions of

$$\Box A_\mu - c\,\partial_\mu\, \partial{\cdot}A = 0 \tag{2.8}$$

can be put in correspondence with the space V' when $c \neq 1$. An analogous chain is displayed:

$$V_g{}^c \subset V^c \subset V'^c.$$

To any solution A of (2.7) corresponds at least one element A^c of V'^c:

$$A^c = A + [c/(1-c)]\partial\Box^{-1}\,\partial{\cdot}A. \tag{2.9}$$

Thus A^c is obtained from A through the transformation $A^c = A + \partial\Lambda$ with $\Box^2\Lambda = 0$. The application (2.9) is the identity when restricted to the subspace V determined by the Lorentz condition $\partial{\cdot}A = 0$. Hence, $V_g{}^c = V_g$ and $V^c = V$. On the quotient spaces V'/V and V'^c/V, the corresponding (2.9) is simply described by

$$\partial{\cdot}A^c = [1/(1-c)]\partial{\cdot}A \tag{2.10}$$

and is obviously biunivocal.

Returning to the simplest case $c = 0$, we now focus on the Poincaré invariance of the chain $V_g \subset V \subset V'$. Under the usual Lorentz transformation of the vector field

$$A_\mu(x) \to A'_\mu(x) = \Lambda_\mu{}^\nu A_\nu(\Lambda^{-1}(x-a)), \tag{2.11}$$

the subspace V_g of longitudinal photons is clearly invariant and carries no helicity. It can be turned into a Hilbert space and the representation then involved is equivalent to $\mathscr{P}(0,0)$, $\mathscr{P}(0,\lambda)$ being the Poincaré UIR with zero mass and discrete helicity λ.

The subspace V_g is not invariantly complemented in the invariant subspace V. The completion of the quotient space V/V_g provided with its norm (2.7b) carries the unitary representation $\mathscr{P}(0,1) \oplus \mathscr{P}(0,-1)$. Also, V is not invariantly complemented in V' and, symmetrically to V_g, the space V'/V of scalar photons carries a representation equivalent to $\mathscr{P}(0,0)$. A not completely reducible or even "indecomposable" representation $U_0(a,\Lambda)$ of the Poincaré group has thus been put in evidence:

$$\mathscr{P}(0,0) \to \{\mathscr{P}(0,1) \oplus \mathscr{P}(0,-1)\} \to \mathscr{P}(0,0), \tag{2.12}$$

where the arrow symbolizes a semidirect sum.

The representation (2.12) is the simplest one among a family of nonequivalent indecomposable representations $U_c(a,\Lambda)$ indexed by the gauge fixing $c \neq 1$. Rideau[8] (see also Ref. 9), explicitly wrote them as extensions of one zero mass scalar representation by a zero mass vectorial representation of the Poincaré group, using cohomological methods. He proved they give rise to covariant quantizations of the Maxwell field, defined in the so-called generalized Lorentz gauges, the latter corresponding precisely to different choices of c.

The transformation (2.9) is not an ordinary gauge transformation since Λ, proportional to $\Box^{-1}\,\partial{\cdot}A$, $A \in V'_{c=0}$, belongs to a certain subspace G' of the space of solutions of

$$\Box^2\Lambda = 0. \tag{2.13}$$

The subspace G' contains an invariant subspace G of solutions of $\Box\Lambda = 0$ carrying the representation $\mathscr{P}(0,0)$. But G is not invariantly complemented and the quotient G'/G carries a representation equivalent to $\mathscr{P}(0,0)$. An extension of the representation with helicity zero by itself has thus been put in evidence,

$$\mathscr{P}(0,0) \to \mathscr{P}(0,0). \tag{2.14}$$

and the resulting nondecomposable representation $U_c(a,\Lambda)$, whose fields A^c form a carrier space, can be pictured by the diagram

$$\begin{array}{l} \mathscr{P}(0,0) \to \mathscr{P}(0,1) \oplus \mathscr{P}(0,-1) \to \mathscr{P}(0,0) \\ \qquad\qquad \mathscr{P}(0,0) \xrightarrow[c]{\hphantom{xxxx}} \nearrow \end{array} . \tag{2.15}$$

We should retain from Ref. 8 that a nondegenerate sesquilinear form invariant with respect to $U_c(a,\Lambda)$ exists if, and only if, c is real and $U_c(a,\Lambda)$ is equivalent to $U_{c'}(a,\Lambda)$ if, and only if, $c = c'$.

III. DE SITTER QED EQUATIONS

The de Sitter version of Eqs. (2.1) reads

$$\nabla^\nu\nabla_\nu A_\mu - \nabla_\mu\nabla^\nu A_\nu - 3\rho A_\mu = 0, \tag{3.1}$$

where the metric $g_{\mu\nu}$, the covariant derivatives ∇_μ, and the (positive) curvature ρ are linked together by

$$[\nabla_\mu,\nabla_\nu]A_\lambda = \rho(g_{\mu\lambda}A_\nu - g_{\nu\lambda}A_\mu). \tag{3.2}$$

Equation (3.1) is identically satisfied by the gauge fields

$$A_\mu = \nabla_\mu A. \tag{3.3}$$

A local isometric embedding in a five-dimensional pseudo-Euclidean space PE_5 with signature $(+,-,-,-,+) = \delta_{\alpha\beta}$, $\alpha,\beta = 0,1,2,3,5$ permits one to visualize the de Sitter space as (the covering space of) a hyperboloid:

$$y^2 \equiv \delta_{\alpha\beta}y^\alpha y^\beta = y_0{}^2 - \vec{y}^2 + y_5{}^2 = 1/\rho. \tag{3.4}$$

A five-component vector field $\{k_\alpha(y)\}$ on PE_5 is locally determined by $A_\mu(x)$ through the relations[4]:

$$A_\mu(x) = y_\mu{}^\alpha k_\alpha(y(x)), \tag{3.5a}$$

where $x \to y(x)$ is the embedding map, and $y_\mu{}^\alpha = \partial y^\alpha/\partial x^\mu$;

$$k_\alpha(y) = (\rho y^2)^{N/2} x_\alpha{}^\mu A_\mu(x(y)), \tag{3.5b}$$

where $y \to x(y)$ is the dilatation-invariant projection map, and $x_\alpha{}^\mu = \partial x^\mu/\partial y^\alpha$. Here, k_α is transverse,

$$y^\alpha k_\alpha(y) = y\cdot k(y) = 0, \tag{3.6}$$

and homogeneous with degree N,

$$\hat{N}k = Nk, \quad \hat{N} \equiv y\cdot\partial = y^\alpha \frac{\partial}{\partial y^\alpha}. \tag{3.7}$$

The gauge fields (3.3) now take the form

$$k_\alpha(y) = (\partial_\alpha - \rho y_\alpha \hat{N})A(x(y)). \tag{3.8}$$

Actually, the differential operator $\bar{\partial} \equiv \partial - \rho y \hat{N}$ is the transverse projection of the gradient ∂. Indeed, introducing the transverse projector θ, whose matrix elements are given by

$$\theta_{\alpha\beta} = \delta_{\alpha\beta} - \rho y_\alpha y_\beta, \tag{3.9}$$

we clearly have

$$\bar{\partial}_\alpha = \theta_\alpha{}^\beta \partial_\beta, \tag{3.10}$$

and we will note the properties

$$\bar{\partial}_\alpha y_\beta = \theta_{\alpha\beta}, \quad \bar{\partial}_\alpha y^2 = 0. \tag{3.11}$$

Thus $\bar{\partial}_\alpha$ is intrinsically defined on the hyperboloid $y^2 = 1/\rho$.

Transforming the covariant derivatives ∇ is straightforward. If k_α corresponds to A_μ through (3.5), $\theta_\beta{}^{\beta'}\theta_\alpha{}^{\alpha'}\partial_{\beta'}k_{\alpha'} = \theta_\alpha{}^{\alpha'}\bar{\partial}_\beta k_{\alpha'} = \bar{\partial}_\beta k_\alpha + \rho y_\alpha k_\beta$ will correspond to $\nabla_\nu A_\mu$ and so on:

$$\nabla\nabla\cdots\nabla A \to \mathrm{Trpr}\,\bar{\partial}\,\mathrm{Trpr}\,\bar{\partial}\cdots\mathrm{Trpr}\,\bar{\partial}k. \tag{3.12}$$

Here, Trpr designates the transverse projection of a tensor field with arbitrary rank,

$$(\mathrm{Trpr}\,k)_{\alpha_1\cdots\alpha_2} = \theta_{\alpha_1}^{\alpha_1'}\cdots\theta_{\alpha}^{\alpha_2'}k_{\alpha_1'\cdots\alpha_2'}.$$

One thus obtains the free field equations of de Sitter space:

$$\bar{\partial}^2 k_\alpha + 2\rho y_\alpha\,\partial\cdot k - \bar{\partial}_\alpha\,\partial\cdot k - 2\rho k_\alpha = 0. \tag{3.13}$$

Together with the "orbital" part $M_{\alpha\beta} = i(y_\alpha\,\partial_\beta - y_\beta\,\partial_\alpha)$ of the infinitesimal generators of the de Sitter group SO(3,2), the operators $i\rho^{-1/2}\,\bar{\partial}_\alpha$ form a representation of the conformal algebra so(4,2) (see Ref. 3):

$$[\bar{\partial}_\alpha,\bar{\partial}_\beta] = -i\rho M_{\alpha\beta}, \tag{3.14a}$$

$$[\bar{\partial}_\alpha, M_{\beta\gamma}] = i(\delta_{\alpha\beta}\,\bar{\partial}_\gamma - \delta_{\alpha\gamma}\,\bar{\partial}_\beta). \tag{3.14b}$$

Now, Eqs. (3.13) can be rewritten in terms which show their SO(3,2) content. The irreducible representations of SO(3,2) which are relevant here are denoted $D(E_0,s)$ (see Ref. 4). The infinitesimal generators of the representation are designated by $L_{\alpha\beta} = M_{\alpha\beta} + S_{\alpha\beta}$, where $S_{\alpha\beta}$ is the "spin" part. The "energy" E_0 is the lowest among the eigenvalues E of L_{50} and the "spin" s is the angular momentum of the lowest energy eigenspace. The second-order Casimir operator

$$Q_s = \tfrac{1}{2}L_{\alpha\beta}L^{\alpha\beta} \tag{3.15}$$

is fixed on the carrier space of $D(E_0,s)$,

$$Q_s = \langle Q_s\rangle I = (E_0(E_0-3) + s(s+1))I, \tag{3.16}$$

and in the flat space limit $\rho \to 0$, the limits of $L_{\mu\nu}$ and $P_\mu \equiv \sqrt{\rho}L_{5\mu}$ become the generators of the Poincaré group.

For $s = 1$, the action of the spin generator $S_{\alpha\beta}$ is defined by

$$S_{\alpha\beta}k_\gamma = i(\delta_{\alpha\gamma}k_\beta - \delta_{\beta\gamma}k_\alpha), \tag{3.17}$$

and it can be verified that

$$L_{\alpha\beta}\,\bar{\partial}_\gamma A = \bar{\partial}_\gamma M_{\alpha\beta}A, \tag{3.18}$$

where A is a scalar field on PE_5.

Now, for spins 0 and 1, the Casimir operators Q_0 and Q_1 can be expressed in terms of the transverse derivatives $\bar{\partial}_\alpha$:

$$Q_0 = -\rho^{-1}\bar{\partial}^2, \tag{3.19a}$$

$$Q_1 k_\alpha = Q_0 k_\alpha - 2y_\alpha\,\bar{\partial}\cdot k + 2k_\alpha, \tag{3.19b}$$

and $\bar{\partial}\cdot k = \partial\cdot k$, since k is transverse.

Therefore, Eq. (3.13) reads

$$Q_1 k + \rho^{-1}\,\bar{\partial}\,\partial\cdot k = 0, \tag{3.20}$$

and can be derived from the Lagrangian density:

$$\mathscr{L} = -(1/2y^2)k\cdot Q_1 k + \tfrac{1}{2}(\partial\cdot k)^2. \tag{3.21}$$

Similarly to flat space QED, the gauge fixing is accomplished by adding to (3.21) two terms involving a scalar field b:

$$\mathscr{L} = -(1/2y^2)k\cdot Q_1 k + \tfrac{1}{2}(\partial\cdot k)^2 - b\,\partial\cdot k + (\alpha/2)b^2. \tag{3.22}$$

A variation of $\mathscr{L}$ leads to the equations

$$Q_1 k + c\rho^{-1}\,\bar{\partial}\,\partial\cdot k = 0, \tag{3.23a}$$

$$b = (1/\alpha)\partial\cdot k, \quad c = (\alpha - 1)/\alpha. \tag{3.23b}$$

But, as we now are going to show it, the simplest choice of c is not zero, contrary to the example of flat space QED.

IV. HOMOGENEOUS PROPAGATORS

To display the remarkable feature[1] of the gauge fixing and its discontinuity with respect to the flat space limit, we will examine in this section the homogeneous propagator $K_{\alpha\alpha'}$, the expansion of which will give us complete information about the states solving the equation of motion (3.23a) for an arbitrary $c \neq 1$, and about the representations of the de Sitter group actually involved there.

The propagator matrix element $K_{\alpha\alpha'}(y,y')$ is a de Sitter-invariant function. It must therefore be a function of the invariant $z = \rho y\cdot y'$ only, and has to solve the equation

$$Q_1 K_{\alpha\alpha'}(z) + c\rho^{-1}\,\bar{\partial}_\alpha\,\partial\cdot K_{\alpha'}(z) = 0 \tag{4.1}$$

(for the y variable). Any tensor $K_{\alpha\alpha'}$ transverse with respect to y and y' (in the indices α and α', respectively) is a linear combination of two basic transverse tensors

$$\mathrm{Trpr}(\delta_{\alpha\alpha'}) = \theta_\alpha{}^\beta\theta'{}_{\alpha'}{}^{\beta'}\delta_{\beta\beta'} \equiv \theta_\alpha\cdot\theta'{}_{\alpha'}, \tag{4.2}$$

$$\mathrm{Trpr}(y_{\alpha'}y'_\alpha) = y\cdot\theta'{}_{\alpha'}y'\cdot\theta_\alpha, \tag{4.3}$$

where θ and θ' are defined in terms of y and y', respectively.

Now, for any function $\phi(z)$, it is easy to check that

$$\bar\partial_\alpha\,\bar\partial'{}_{\alpha'}\phi = \rho\theta_\alpha\cdot\theta'{}_{\alpha'}\phi' + \rho^2 y\cdot\theta'{}_{\alpha'}y'\cdot\theta_\alpha\phi''. \tag{4.4}$$

Therefore, it is equivalent to put $K_{\alpha\alpha'}$ under a form which renders apparent its gauge part:

$$K_{\alpha\alpha'}(z) = \theta_\alpha\cdot\theta'{}_{\alpha'}d(z) + \rho^{-1}\bar\partial_\alpha\,\bar\partial'{}_{\alpha'}\phi(z), \tag{4.5}$$

where $d(z)$ and $\phi(z)$ will be determined by Eq. (4.1). To find the differential equations satisfied by d and ϕ, we will make use of the commutation rules below, which statement requires some technical manipulations:

$$Q_1\bar\partial d = \bar\partial Q_0 d, \tag{4.6}$$

$$Q_1\theta_\alpha\cdot\theta'{}_{\alpha'}d = \theta_\alpha\cdot\theta'{}_{\alpha'}(Q_0+2)d + 2\rho^{-1}\bar\partial_\alpha\,\bar\partial'{}_{\alpha'}\int d, \tag{4.7}$$

$$\bar\partial_\alpha\,\partial\cdot\theta\cdot\theta'{}_{\alpha'}d = -\bar\partial_\alpha\,\bar\partial'{}_{\alpha'}\left[zd + 3\int d\right], \tag{4.8}$$

where $\int$ designates the antiderivative operator with respect to z. Equation (4.1) then becomes

$$\theta_\alpha\cdot\theta'{}_{\alpha'}(Q_0+2)d + \rho^{-1}\bar\partial_\alpha\,\bar\partial'{}_{\alpha'}\Big[(1-c)Q_0\phi + (2-3c)\int d - czd\Big] = 0.$$

Therefore,

$$(Q_0+2)d = 0, \tag{4.9}$$

and

$$\phi(z) = \frac{1}{1-c}Q_0{}^{-1}\left[czd(z) + (3c-2)\int^z d(z)\right]. \tag{4.10}$$

In terms of the variable z, the Casimir operator $Q_0 = -\rho^{-1}\bar\partial^2$ takes the simple form

$$Q_0 = -\left[(1-z^2)\frac{d^2}{dz^2} - 4z\frac{d}{dz}\right], \tag{4.11}$$

and, consequently,

$$Q_0{}^{-1}f(z) = -\int^z (1-z^2)^{-2}\int^z (1-z^2)f(z). \tag{4.12}$$

Two independent solutions of (4.9) are simply $(z\pm1)^{-1}\equiv d_\pm(z)$. Now, examining Eq. (4.10) reveals two extreme values of c: $c=0$ and $c=\frac{2}{3}$. At first glance, the case $c=\frac{2}{3}$ is singularized by its simplicity. Indeed, since $zd_\pm(z) = d_\pm(z)+1$ and $Q_0^{-1}d_\pm = -\frac{1}{2}d_\pm$ from (4.9), we obtain for $c=\frac{2}{3}$ the two independent homogeneous propagators

$$K^\pm_{\alpha\alpha'}(z) = \theta_\alpha\cdot\theta'{}_{\alpha'}d_\pm(z) \pm \rho^{-1}\bar\partial_\alpha\,\bar\partial'{}_{\alpha'}d_\pm(z). \tag{4.13}$$

Irrespective of a certain choice of c, terms proportional to

$$Q_0{}^{-1}0 = -\,cste\int^z (1-z^2)^{-2}, \tag{4.14}$$

or to

$$Q_0{}^{-1}cste = -\,cste\int^z (1-z^2)^{-2}\int^z (1-z^2), \tag{4.15}$$

can be arbitrarily added to the expression of ϕ given by (4.10) and will supplement the expression (4.5) to $K^\pm_{\alpha\alpha'}$ with pure gauge field propagators without changing its physical content.

The case $c=0$ introduces logarithms in the expression of K: Equation (4.10) now turns to be a rather intricate expression:

$$\begin{aligned}\phi^\pm(z) &= 2\int^z (1-z^2)^{-2}\int^z (1-z^2)\int^z d_\pm(z)\\ &= \frac{2}{3}\left[\left[\frac{\ln|1\pm z|}{1\mp z} \pm \int^z \frac{\ln|1\pm z|}{1\mp z}\right]\right.\\ &\quad\left. + \frac{5}{6}\left[\ln|1\mp z| - \frac{1}{1\mp z}\right] - \frac{1}{2}\ln|1\pm z|\right].\end{aligned} \tag{4.16}$$

Finally, any other gauge fixing $c\neq1$ will present itself as a mixture of the two previously considered basic ones, $c=0$ and $c=\frac{2}{3}$.

V. GROUP THEORETICAL INTERPRETATION OF THE SOLUTIONS

The first part of the homogeneous propagator $K^\pm_{\alpha\alpha'}(z)$: $\theta_\alpha\cdot\theta'{}_{\alpha'}d_\pm(z)$, has an obvious de Sitter group interpretation, $\theta_\alpha\cdot\theta'{}_{\alpha'} = \mathrm{Trpr}\ \delta_{\alpha\alpha'}$ is the propagator for the five-dimensional representation D_5 of SO(3,2), and $d_\pm(z)$ are solutions of the equation $(Q_0+2)\,d(z)=0$, which is also satisfied by the homogeneous propagators of the scalar representations $D(1,0)$ and $D(2,0)$ (see Ref.3), according to Eq. (3.16). The choice which is left to us concerning the sign $+$ or $-$ reflects the existence in the de Sitter theory of two disconnected parts for the initial data surface.[1] Introducing, besides the space coordinates $\vec y$, the de Sitter time coordinate t,

$$y_5 + iy_0 = Ye^{i\sqrt{\rho}t}, \quad Y = (1/\rho + y^2)^{1/2}, \tag{5.1}$$

they correspond to $t=0$ and $t=\pi/\sqrt{\rho}$, and neither can be ignored. But it has been shown[3] that one cannot accommodate both representations $D(2,0)$ and $D(1,0)$ in the domain of a single self-adjoint Hamiltonian $P_0 = i(\partial/\partial t) = \sqrt{\rho}\,L_{50}$. Taking even and odd combinations of the solutions $d_\pm(z)$ permits us to isolate their respective propagators and to consider only the initial data surface $t=0$. The propagator for $D(2,0)$ is the even combination (even with respect to the transformation $z\to -z$)

$$D_+(z) = (\rho/4\pi^2)(z^2-1)^{-1} = (\rho/8\pi^2)[d_-(z) - d_+(z)]. \tag{5.2}$$

The propagator for $D(1,0)$ is the odd combination

$$D_-(z) = (\rho/4\pi^2)z(z^2-1)^{-1} = (\rho/8\pi^2)[d_-(z) + d_+(z)]. \tag{5.3}$$

Now, $\theta_\alpha\cdot\theta'{}_{\alpha'}D_\pm(z)$ are the propagators of the representations $D_5\otimes D(2,0)$ and $D_5\otimes D(1,0)$, respectively. The Clebsch–Gordan reduction of these tensor products is easily demonstrated to be[1]

$$D_5\otimes D(2,0) = D(1,0)\oplus\{D(3,0)\to D(2,1)\to D(3,0)\}, \tag{5.4}$$

$$D_5 \otimes D(1,0) = D(2,0) \oplus \left\{ \begin{array}{c} D(1,1) \rightarrow D(2,1) \rightarrow D(1,1) \\ \oplus \\ D(0,0) \end{array} \right\}, \tag{5.5}$$

the arrows still denoting semidirect sums.

Even and odd combinations of the propagators $K^{\pm}_{\alpha\alpha'}$ now take the form (where we stress the c dependence)

$$\begin{aligned} P^{c\pm}_{\alpha\alpha'}(z) &= (\rho/8\pi^2)\left[K^{-}_{\alpha\alpha'}(z) \mp K^{+}_{\alpha\alpha'}(z)\right] \\ &= \theta_\alpha \cdot \theta'_{\alpha'} D_\pm(z) + \rho^{-1} \bar\partial_\alpha \bar\partial'_{\alpha'} G^{c}_{\pm}(z), \end{aligned} \tag{5.6}$$

where

$$\begin{aligned} G_\pm{}^c &= \frac{1}{1-c}\left[-\frac{c}{2}D_\mp + (3c-2)Q_0{}^{-1}\int D_\pm\right] \\ &= -D_\mp + \frac{2-3c}{1-c}\left[\frac{1}{2}D_\mp - Q_0{}^{-1}\int D_\pm\right] \\ &= -D_\mp - \frac{2-3c}{1-c}Q_0{}^{-1}\left[D_\mp + \int D_\pm\right]. \end{aligned}$$

VI. THE CASE $c = \frac{2}{3}$: GUPTA–BLEULER TRIPLETS

The case $c = \frac{2}{3}$ is now particularly illuminating. The propagator (5.6) takes the simple form

$$P^{2/3\,\pm}_{\alpha\alpha'}(z) = \theta_\alpha \cdot \theta'_{\alpha'} D_\pm(z) - \rho^{-1}\bar\partial_\alpha \bar\partial'_{\alpha'} D_\mp(z). \tag{6.1}$$

The second terms of the above expression remove the pure gauge fields carrying the representations $D(1,0)$ and $D(2,0)$ appearing in the direct sums (5.4) and (5.5) from the propagators $P^{2/3\,\pm}$. We are thus left with the two distinct Gupta–Bleuler triplets already introduced and discussed in Ref. 1.

$$D(3,0) \rightarrow D(2,1) \rightarrow D(3,0), \tag{6.2}$$

$$\begin{array}{c} D(1,1) \rightarrow D(2,1) \rightarrow D(1,1). \\ \oplus \\ D(0,0) \end{array} \tag{6.3}$$

Because of the behavior of the carrier states under the reflection $y_\alpha \to -y_\alpha$, the inner product associated to each representation is given by integration on $t = 0$ only. Equations (6.2) and (6.3) describe two dynamically independent vector potentials belonging to two different domains of self-adjointness for the Hamiltonian. Understood in terms of conformal QED, they owe their separate existence to a spontaneous breaking of the conformal symmetry. The conformal QED Gupta–Bleuler triplet is[2]

$$\begin{array}{c} D(1,\tfrac{1}{2},\tfrac{1}{2}) \rightarrow D(2,1,0) \oplus D(2,0,1) \rightarrow D(1,\tfrac{1}{2},\tfrac{1}{2}). \\ D(0,0,0,) \end{array}$$

The central representation $D(2,1,0) \oplus D(2,0,1)$ describes physical photons with helicities $\lambda = 1$ and $\lambda = -1$ when interpreted through its restriction to the Poincaré subgroup. Instead, when considered in terms of its reduction to the de Sitter subgroup, the requirement of self-adjointness for the Hamiltonian obliges us to work with even and odd combinations of $\lambda = +1$ and $\lambda = -1$ states, and we are thus led to the independent sets (6.2) and (6.3).

In order to compare them with those introduced in the following section, it is worthy to recall from Ref. 1 cyclic and ground states for the indecomposable representations (6.2) and (6.3). Such states are eigenstates of the energy operator L_{50} and generate the carrier spaces involved when acted on by energy-shifting operators $L_{5i} \mp L_{0i}, i = 1,2,3$.

Tables I and II explicitly show them up to normalization and the addition of gauge states. One introduces a set $\{z_\alpha\}$ of Cartesian coordinates for the tangent space of PE_5 and the complex variables $y_\pm = y_5 \pm iy_0$, $z_\pm = z_5 \pm iz_0$, with energy $E = \mp 1$. The arrows denote the sense of the leaking after application of infinitesimal de Sitter generators. One also specifies the values of the divergence $\partial \cdot k$. Here, $\vec\theta$ designates the tensor with components $\theta_{\alpha i}$, where $i = 1,2,3$ indicates the polarization.

One will note the appearance in the second triplet of the state with energy zero:

$$k_\alpha = \bar\partial_\alpha \ln\sqrt{\rho}\, y_+ = y_+{}^{-1}(\theta_{\alpha 5} + i\theta_{\alpha 0}).$$

This (locally) gauge state is associated with the trival representation $D(0,0)$.

VII. THE CASES $c \neq \frac{2}{3}$: LOGARITHMIC STATES

The terms $Q_0{}^{-1} \int D_\pm$ which appear in the expression (5.6) introduce logarithms through the following combinations:

$$g_\pm(z) \equiv D_\mp(z)\ln|z^2 - 1| + D_\pm(z)\ln|(z+1)/(z-1)|. \tag{7.1}$$

By removing terms like $Q_0{}^{-1}0 \propto [D_- + \int D_+]$ and $Q_0{}^{-1}\,\text{const} \propto [D_+ + \int D_-]$, we obtain for the propagators

$$\begin{aligned} P^{c+}_{\alpha\alpha'}(z) &= P^{2/3\,+}_{\alpha\alpha'}(z) + \rho^{-1}\frac{(\frac{2}{3} - c)}{2(1-c)}\bar\partial_\alpha \bar\partial'_{\alpha'} \\ &\quad \times \left[\frac{\rho}{2\pi^2}\ln\left|\frac{z+1}{z-1}\right| + g_+(z)\int^z g_-(z)\right], \end{aligned} \tag{7.2}$$

$$\begin{aligned} P^{c-}_{\alpha\alpha'}(z) &= P^{2/3\,-}_{\alpha\alpha'}(z) - \rho^{-1}\frac{(\frac{2}{3} - c)}{2(1-c)}\bar\partial_\alpha \bar\partial'_{\alpha'} \\ &\quad \times \left[\frac{\rho}{2\pi^2}\ln|1 - z^2| + g_-(z) + \int^z g_+(z)\right]. \end{aligned} \tag{7.3}$$

Note that $\int^z g_+(z) = (\rho/4\pi^2)\ln|1-z|\ln|1+z|$.

TABLE I. Ground states and cyclic states for the Gupta–Bleuler triplet.

$D(3,0)$		$D(2,1)$		$D(3,0)$
$z \cdot k = y_+{}^{-3}(y_+ z_- - y_- z_+)$	$\rightarrow$	$y_+{}^{-3}(y_+\vec z - z_+\vec y)$	$\rightarrow$	$z \cdot \vec\partial y_+{}^{-3}$
cyclic for the whole Gupta–Bleuler triplet $\partial \cdot k = -6y_+{}^{-3}$		physical state, cyclic for $D(2,1) \rightarrow D(3,0)$ $\partial \cdot k = 0$		absolute ground state for $D(3,0)$ $\partial \cdot k = 0$

TABLE. II. Ground states and cyclic states for the second Gupta–Bleuler triplet.

$D(1,1)$ $z\cdot\vec{k} = y_+^{-1}\vec{\theta}\cdot z$ The arrow on k denotes a polarization $\partial\cdot\vec{k} = -3\rho y_+^{-1}\vec{y}$	$\longrightarrow$	$D(2,1)$ $y_+^{-2}\vec{y}\wedge\vec{z}$ cyclic vector for $D(2,1)\to D(1,1)$ $\partial\cdot k = 0$	$\longrightarrow$	$D(1,1)$ $z\cdot\bar{\partial}(y_+^{-1}\vec{y})$ absolute ground state for $D(1,1)$ $\partial\cdot k = 0$
	$\searrow$	$D(0,0)$ $z\cdot\bar{\partial}\ln\sqrt{\rho}y_+$ $\partial\cdot k = -3\rho$	$\nearrow$	

Note that $\int^z g_+(z) = (\rho/4\pi^2)\ln|1-z|\ln|1+z|$.

One can see from the previous expressions that the carrier states for the Gupta–Bleuler triplets (6.2) and (6.3) are now added to gauge fields coming from the factors of $\frac{2}{3} - c$. A limited expansion of the latter in inverse powers of z permits us to identify these states:

$$\bar{\partial}_\alpha\,\bar{\partial}'_{\alpha'}\left[\frac{\rho}{2\pi^2}\ln\left|\frac{z+1}{z-1}\right| + g_+(z) + \int^z g_-(z)\right]$$

$$\simeq(\rho/6\pi^2)\bar{\partial}_\alpha\,\bar{\partial}'_{\alpha'}\left[2z^{-3}\ln|z| + \tfrac{2}{3}z^{-3}\right], \tag{7.4}$$

$$\bar{\partial}_\alpha\,\bar{\partial}'_{\alpha'}\left[\frac{\rho}{2\pi^2}\ln|z^2-1| + g_-(z) + \int^z g_+(z)\right]$$

$$\simeq(\rho/4\pi^2)\bar{\partial}_\alpha\,\bar{\partial}'_{\alpha'}\left[(\ln|z|)^2 + 4\ln|z|\right]. \tag{7.5}$$

Let us express the variable z in terms of y and y' in order to isolate its lowest energy part:

$$z = \rho y\cdot y' = \frac{\rho}{2}y_+y'_-\left[1 - \frac{2\vec{y}\cdot\vec{y}'}{y_+y'_-} + \frac{y_-y'_+}{y_+y'_-}\right]. \tag{7.6}$$

The factor $y_+ = y_5 + iy_0 = Ye^{i\sqrt{\rho}t}$ is an eigenstate of $M_{50} = (i/\sqrt{\rho})(\partial/\partial t)$ with eigenvalue -1. Because the commutation rule (3.18), examining the energy content of a gauge field $\bar{\partial}_\alpha\Lambda$ comes back to study the group-transformation behavior of the scalar function Λ itself. Thus, the lowest definite-energy state appearing in the expansion (7.4) is $\bar{\partial}_\alpha y_+^{-3}$, with energy 3. Besides, there appears the state $\bar{\partial}_\alpha y_+^{-3}\ln\sqrt{\rho}y_+$ coming from the term

$$z^{-3}\ln|z|\simeq((\rho/2)y_+y'_-)^{-3}(\ln\rho y_+y'_- + \cdots).$$

This "logarithmic" state has no definite energy;

$$L_{50}\,\bar{\partial}_\alpha y_+^{-3}\ln\sqrt{\rho}y_+$$

$$= \bar{\partial}_\alpha\frac{i}{\sqrt{\rho}}\frac{\partial}{\partial t}\,Y^{-3}e^{-i3\sqrt{\rho}t}(\ln\sqrt{\rho}\,Y + i\sqrt{\rho}\,t)$$

$$= 3\bar{\partial}_\alpha y_+^{-3}\ln\sqrt{\rho}y_+ - \bar{\partial}_\alpha y_+^{-3}. \tag{7.7}$$

Actually, considered as a gauge field up to the addition of a $D(3,0)$ gauge field, it is an eigenstate of L_{50} with energy 3. Moreover, it is cancelled by the energy-lowering operator $L_{5i} + iL_{0i}$ and the energy-raising operator $L_{5i} - iL_{0i}$ transforms it into the logarithmic state $\bar{\partial}_\alpha y_+^{-4}y_i\ln\sqrt{\rho}y_+$ up to the addition of a $D(3,0)$ state. An invariant space of gauge fields is thus generated. It carries the indecomposable representation of the Lie algebra so(3,2),

$$D(3,0)\to D(3,0). \tag{7.8}$$

The first term $\bar{\partial}_\alpha\,\bar{\partial}'_{\alpha'}(\ln|z|)^2$ of the expansion (7.5) gives birth to the logarithmic vector gauge field $\bar{\partial}_\alpha y_+^{-1}\vec{y}\ln\sqrt{\rho}y_+$ up to the addition of the $D(0,0)$ state $\bar{\partial}_\alpha\ln\sqrt{\rho}y_+$. To see this, note the following approximations:

$$\bar{\partial}_\alpha\,\bar{\partial}'_{\alpha'}(\ln|z|)^2$$

$$\simeq\bar{\partial}_\alpha\,\bar{\partial}'_{\alpha'}\left[\ln\frac{\rho}{2}y_+y'_- - \frac{2\vec{y}\cdot\vec{y}}{y_+y'_-}\right]^2$$

$$\simeq\bar{\partial}_\alpha\,\bar{\partial}'_{\alpha'}(\ln\sqrt{\rho}y_+)(\ln\sqrt{\rho}y'_-)$$

$$-4\bar{\partial}_\alpha\,\bar{\partial}'_{\alpha'}\frac{\vec{y}\cdot\vec{y}'}{y_+y'_-}\ln\rho y_+y'_-.$$

From the group actions,

$$L_{50}\,\bar{\partial}_\alpha y_+^{-1}\vec{y}\ln\sqrt{\rho}y_+ = \bar{\partial}_\alpha\left[y_+^{-1}\vec{y}\ln\sqrt{\rho}y_+ - y_+^{-1}\vec{y}\right],$$

$$(L_{5i} + iL_{0i})\bar{\partial}_\alpha y_+^{-1}\vec{y}\ln\sqrt{\rho}y_+ = i(\partial_i\vec{y})(\bar{\partial}_\alpha\ln\sqrt{\rho}y_+),$$

$$(L_{5i} - iL_{0i})\bar{\partial}_\alpha y_+^{-1}\vec{y}\ln\sqrt{\rho}y_+$$

$$= 2i\,\bar{\partial}_\alpha y_i\vec{y}\left[y_+^{-2} - y_+^{-2}\ln\sqrt{\rho}y_+\right]$$

$$+ i(\partial_i\vec{y})\bar{\partial}_\alpha y_-y_+^{-1}\ln\sqrt{\rho}y_+, \tag{7.9}$$

we easily conclude that the considered state is eigenstate of L_{50} with energy 1 up to the addition of a $D(1,1)$ gauge state and the action of the energy-lowering operator makes it leak to a $D(0,0)$ state. Another invariant space of gauge fields is thus generated; the carrier space for the indecomposable representation of the Lie algebra so(3,2) is

$$[D(0,0)\to D(1,1)]\to[D(0,0)\to D(1,1)]. \tag{7.10}$$

VIII. DISCUSSION

The adding of extra gauge fields when c differs from $\frac{2}{3}$ can be understood by repeating the argument presented at the end of the second section. The space V'_c of solutions of Eq. (3.23a),

$$Q_1k + \rho^{-1}c\,\bar{\partial}\,\partial\cdot k = 0,$$

when $c\neq\frac{2}{3}$ (and $c\neq 1$), can be put in correspondence with the space $V'_{2/3}$ through the equation

$$k^c = k^{2/3} + \frac{(\frac{2}{3}-c)}{\rho(1-c)}\bar{\partial}Q_0^{-1}\,\partial\cdot k^{2/3}. \tag{8.1}$$

The correspondence is not trivial for those states $k^{2/3}$ such that $\partial\cdot k^{2/3}$ is not zero, and then we have

$$\partial\cdot k^c = [1/3(1-c)]\partial\cdot k^{2/3}.$$

The cyclic state displayed by Table I can be used to build the carrier space of the first Gupta–Bleuler triplet. We have $\partial\cdot k^{2/3}$ proportional to y_+^{-3} and (see the Appendix)

$$\bar{\partial} Q_0{}^{-1}\, \partial \cdot k^{2/3} \propto \bar{\partial} y_+{}^{-3} \ln\sqrt{\rho}\, y_+, \tag{8.2}$$

up to the addition of $D(3,0)$ gauge states. Therefore, a logarithmic gauge field is involved in the expression of k^c as expected in the previous section.

For the cyclic vector state displayed by Table II, $\partial \cdot \vec{k}^{2/3}$ is proportional to $y_+{}^{-1}\vec{y}$ and (see the Appendix)

$$\bar{\partial} Q_0{}^{-1}\, \partial \cdot k^{2/3} \propto \bar{\partial} y_+{}^{-1} \vec{y} \ln\sqrt{\rho}\, y_+, \tag{8.3}$$

up to the addition of $D(1,1)$ gauge states. Here too, we recover an expected result.

Finally, for the $D(0,0)$ gauge state $\bar{\partial} \ln\sqrt{\rho}\, y_+$ of the same table, $\partial \cdot k^{2/3} = -3\rho$,

$$\bar{\partial} Q_0{}^{-1}\, \partial \cdot k^{2/3} = -\rho \bar{\partial} \ln\sqrt{\rho}\, y_+, \tag{8.4}$$

and therefore

$$k^c = [1/3(1-c)]k^{2/3}.$$

Actually, one can build the solution k^c by adding to $k^{2/3}$ a gauge field $\bar{\partial}\Lambda$, where Λ obeys

$$Q_0{}^2 \Lambda = 0, \tag{8.5}$$

if $k^{2/3}$ is a carrier state for the first Gupta–Bleuler triplet, or

$$Q_0{}^2 \Lambda = \text{const} \neq 0, \tag{8.6}$$

if it is a carrier state for the second Gupta–Bleuler triplet. A precise meaning is thus given to the fact that $\bar{\partial}\Lambda$ belongs to the carrier spaces of the semidirect sums

$$D(3,0) \rightarrow D(3,0)$$

or

$$[D(0,0) \rightarrow D(1,1)] \rightarrow [D(0,0) \rightarrow D(1,1)],$$

indecomposable Lie algebra representations under which the spaces of solutions of (8.5) and (8.6), respectively, are invariant.

The appearance of the solutions of (8.5) and (8.6) in the expressions of k^c could already be guessed from the additional terms of the propagators given by Eq. (5.6). Since $Q_0{}^{-1} 0 \propto [D_- + \int D_+]$ and $Q_0{}^{-1}\,\text{const} \propto [D_+ + \int D_-]$, propagators $P^{c\pm}_{\alpha\alpha'}$ take the illuminating form

$$P^{c\pm}_{\alpha\alpha'}(z) = P^{2/3}_{\alpha\alpha'}(z) - \rho^{-1} \frac{2-3c}{1-c} \bar{\partial}_\alpha \bar{\partial}'_{\alpha'} Q_0{}^{-2}\, {}^{0}_{\text{const}}. \tag{8.7}$$

Now, the set of the fields k^c defined by (8.1) forms a carrier space for Lie algebra indecomposable representations pictured by the following diagrams. When there is departure from the case $c = \frac{2}{3}$, the occurrence of the extensions (7.8) and (7.10) brings c-indexed arrows which supplement the Gupta–Bleuler triplets of Sec. VI:

$$\begin{tikzcd} D(3,0) \arrow[r, mapsto] & D(2,1) \arrow[r] & D(3,0), \\ & D(3,0) \arrow[ur, "c"] & \end{tikzcd} \tag{8.8}$$

$$\begin{tikzcd} & D(1,1) \arrow[r, mapsto] & D(2,1) \arrow[r, mapsto] & D(1,1)] \\ & & \oplus & \\ {[D(0,0)} \arrow[r] & D(1,1)] \arrow[r, "c"] & {[D(0,0)} \arrow[uur] & \end{tikzcd} \tag{8.9}$$

A problem immediately arises: Are these c-indexed representations equivalent or not, comparatively, to the flat space situation? A more global, less pedestrian answer of the functional spaces and group actions involved here would actually be necessary to answer that question. Presently, our conclusion is only to assert that the simplest choice of the gauge fixing in de Sitter QED remains $c = \frac{2}{3}$: no logarithmic scalar photon states would appear.

ACKNOWLEDGMENTS

We are indebted to Christian Fronsdal for his decisive part in initiating and developing the content of this paper. We thank Moshe Flato and Guy Rideau for useful discussions. We are also grateful to the UCLA Physics Department for its hospitality.

APPENDIX: A SHORT REVIEW OF THE DE SITTER MACHINERY USED IN THIS PAPER

The (global) de Sitter coordinates t, $\vec{y}$ are defined by $(y_\alpha) = (y_0, \vec{y}, y_5)$,

$$y_\pm = y_5 \pm i y_0 = Y \exp(\pm i\sqrt{\rho}\, t),$$
$$Y = (y_5{}^2 + y_0{}^2)^{1/2} = (1/\rho + y^2)^{1/2}.$$

In these terms, the most commonly used intrinsic operators are as follows (Latin indices $i, j, \ldots$, rank from 1 to 3).

(i) $$\bar{\partial}_\alpha = \frac{\dot{y}_\alpha}{\rho Y^2} \frac{\partial}{\partial t} + \theta_{\alpha j}\, \partial^j,$$

where

$$\dot{y}_\alpha = \sqrt{\rho}(\delta_{0\alpha} y_5 - \delta_{5\alpha} y_0).$$

(ii) $$M_{\alpha\beta} = \frac{i}{\sqrt{\rho}\, Y^2} q_{\alpha\beta} \frac{\partial}{\partial t} + i\delta_{\alpha\beta,\gamma j} y^\gamma \partial^j,$$

where

$$q_{\alpha\beta} = (1/\sqrt{\rho})[y_\alpha \dot{y}_\beta - y_\beta \dot{y}_\alpha], \quad \delta_{\alpha\beta,\gamma\delta} = \delta_{\alpha\gamma}\delta_{\beta\delta} - \delta_{\alpha\delta}\delta_{\beta\gamma}.$$

Explicitly,

$$q_{50} = Y^2, \quad q_{5i} = y_0 y_i, \quad q_{0i} = -y_5 y_i, \quad q_{ij} = 0,$$

$$M_{50} = \frac{i}{\sqrt{\rho}} \frac{\partial}{\partial t}, \quad M_{5i} = i\left[\frac{1}{\sqrt{\rho}\, Y^2} y_0 y_i \frac{\partial}{\partial t} + y_5\, \partial_i\right],$$

$$M_{0i} = i\left[-\frac{1}{\sqrt{\rho}\, Y^2} y_5 y_i \frac{\partial}{\partial t} + y_0\, \partial_i\right],$$

$$M_{ij} = i[y_i \partial_j - y_j\, \partial_i].$$

(iii) Energy-lowering and -raising operators. The energy-raising operator is

$$M_{5i} - iM_{0i} = -y_-\left(\frac{1}{\sqrt{\rho}\, Y^2} y_i \frac{\partial}{\partial t} - i\, \partial_i\right).$$

The energy-lowering operator is

$$M_{5i} + iM_{0i} = y_+\left(\frac{1}{\sqrt{\rho}\, Y^2} y_i \frac{\partial}{\partial t} + i\, \partial_i\right).$$

(iv) The Casimir operator is

$$Q_0 = \frac{1}{2} M_{\alpha\beta} M^{\alpha\beta} = -\frac{1}{\rho^2 Y^2} \frac{\partial^2}{\partial t^2} + \frac{1}{\rho} \Delta_s,$$

where

$$\Delta_s = -\partial_i\, \partial^i + 4\rho y_i\, \partial^i + \rho y_i y_j\, \partial^i \partial^j.$$

(v) The specific actions on scalar functions of the variable y_+ alone are

$$M_{50}f(y_+) = -y_+ f'(y_+),\quad (M_{5i}+iM_{0i})f(y_+) = 0,$$
$$(M_{5i}-iM_{0i})f(y_+) = 2iy_i f'(y_+),$$
$$Q_0 f(y_+) = y_+{}^2 f''(y_+) + 4y_+ f'(y_+),$$
$$Q_0{}^{-1}f(y_+) = \int^{y_+} y_+{}^{-4}\int^{y_+} y_+{}^2 f(y_+),$$

particularly where $Q_0{}^{-1}0 \propto y_+{}^{-3}$ is an absolute ground state for the representation $D(3,0)$, $Q_0{}^{-1}cste \propto \ln\sqrt{\rho}y_+$, and $\bar{\partial}\ln\sqrt{\rho}y_+$ is an absolute ground state for the semidirect sum $D(0,0)\rightarrow D(1,1)$. The specific actions on particular vector functions are

$$Q_0{}^{-1}y_i f(y_+) = y_i y_+{}^{-4}\int^{y_+} y_+{}^2\int^{y_+} f(y_+).$$

For example, $f(y_+) = y_+{}^{-1}$,

$$Q_0{}^{-1}y_i y_+{}^{-1} = y_i\left[\tfrac{1}{3}y_+{}^{-1}\ln\sqrt{\rho}y_+ + cste\, y_+{}^{-1} + cste\, y_+{}^{-4}\right].$$

[1]B. Binegar, C. Fronsdal, and W. Heidenreich, Ann. Phys. **149**, 254 (1983).
[2]B. Binegar, C. Fronsdal, and W. Heidenreich, J. Math. Phys. **24**, 2828 (1983).
[3]C. Fronsdal, Phys. Rev. D **12**, 3819 (1975).
[4]C. Fronsdal, Phys. Rev. D **20**, 848 (1979).
[5]C. Fronsdal and H. Hata, Nucl. Phys. B **162**, 487 (1980).
[6]N. Nakanishi, Progr. Theor. Phys. **35**, 1111 (1966).
[7]F. Bayen and M. Flato, J. Math. Phys. **17**, 1112 (1976).
[8]G. Rideau, Lett. Math. Phys. **2**, 529 (1978); J. Math. Phys. **19**, 1627 (1978); Rep. Math. Phys. **16**, 251 (1979).
[9]F. Strocchi and A. S. Wightman, J. Math. Phys. **15**, 2198 (1974).

A new representation for planar Feynman graphs in terms of strings

R. Hong Tuan
Laboratoire de Physique Théorique et Hautes Energies,[a) Université de Paris-Sud, Bâtiment 211, 91405 Orsay Cedex, France

(Received 13 September 1984; accepted for publication 28 December 1984)

We study the planar graphs of a massless scalar ϕ^3 field theory in d Euclidean dimensions. We show, using a lattice cutoff, that any graph G can be expressed as a function of strings defined on the dual of G (in the sense of graph theory). We furthermore obtain a series expansion of the Feynman graph in powers of $(1/d)$. The powers are proportional to the total length of the strings drawn on d copies of the dual of G for a given subset of these strings which never cross either themselves or each other on the same copy.

I. INTRODUCTION

In the last years a growing interest[1–5] can be noted for the study of planar contributions of field theories and of string models. The main reason for it is that it is generally agreed that the $1/N$ expansion is relevant to the real world. Organizing contributions according to their increasingly complex topology may provide a useful tool in disentangling the enormous difficulties associated with the computations in QCD, for instance. But, even restricting oneself to the leading planar contribution, it is still an important unsolved problem to understand how to evaluate that simple topology. Many people[6] have derived (more or less heuristically) relationships between the planar graphs of field theories and string models. But it is generally impossible to know to what extent these relationships hold. Moreover, we lack an analytical framework where one could hope to compute the planar topology and verify, for example, if string models can reproduce the short distance behavior of field theories.

It is our purpose, in the present paper, to derive a new representation of planar Feynman graphs, using a lattice cutoff. This representation is obtained by using an operatorial formalism where the creation and annihilation operators correspond to degrees of freedom living on the dual (in the sense of graph theory) graph of the original one. In fact, we will see that any planar graph can be expressed as a function of closed and open oriented paths (or strings) defined on the dual of G. It is interesting that, by the same token, we also have a power series expansion of any Feynman graph (with a lattice cutoff) in powers of $1/d$, where d is the dimension of space-time used. The power of $(1/d)$ is then proportional to the total length of the strings with no crossing drawn on d copies of the dual of G.

II. DERIVATION OF THE REPRESENTATION

We now explain in details the derivation of this representation. As explained before, we restrict ourselves to planar graphs, i.e., graphs which can be embedded in a plane without any propagator crossing another. External lines will be attached to the border of the graph. We only consider the case of one continuous external border. That is, no hole can exist to which an internal border line would correspond. We note that, as usual with string models, more complex topologies can be generated by punching holes in a planar graph and sewing together the internal border lines.

[a) Laboratoire associé au Centre National de la Recherche Scientifique.

Let us consider any connected planar Feynman graph for a massless scalar ϕ^3 field theory in an Euclidean d-dimensional space. If it has no loop, it is trivial to evaluate it, because no integration is left over after the conservation of momentum and energy has been taken into account at the vertices. The difficulty in evaluating the Feynman graphs lies in the loops. Now, to each internal loop (or face) of the diagram is associated a d-momentum which has to be integrated over. A propagator inside the graph will be the line of contact between two faces or loops. This propagator will have an argument which is the difference of momenta attached to the loops of which it is the border. One can also say that if two loops are *adjacent*, they will share a propagator. In fact, one can characterize any planar diagram by its incidence matrix. This symmetric matrix will have its rows and columns corresponding to the loops. Whenever a loop j and a loop k will have a propagator in common a nonzero matrix element α_{jk} will exist.

The definition of loops can be extended to "external" ones which are defined in the following way: An external loop is formed by a line of propagators on the border of the graph and the external lines coming at the ends of this line of propagators. Therefore external faces or loops extend to infinity and they correspond to momenta which are *not* integrated over contrary to the internal ones. An internal loop j will correspond to a momentum p_j^μ and an external loop k will correspond to a momentum $(p_k^\mu)_0$. Using the Feynman parametrization, a propagator can be written as

$$(p_k - p_j)^{-2} = \int_0^\infty d\alpha_{jk} \exp[-\alpha_{jk}(p_k - p_j)^2]. \tag{1}$$

(Notice that the Feynman parameters α_{jk} will be the elements of the incidence matrix.)

Now, we make use of a lattice cutoff in order to define the theory in an arbitrary number of dimensions. If the lattice spacing is a, the square of momenta $(p^\mu)^2$ in the propagators will be replaced by $\hat{p}^\mu \hat{p}^{\mu *}$ with

$$\hat{p}^\mu = a^{-1}[\exp(-ip^\mu a) - 1].$$

Moreover, the p_μ's will be integrated over the interval $[-\pi/a, \pi/a]$. Introducing the angle

$$\theta^\mu = p^\mu a,$$

we can write $\Sigma_\mu \hat{p}^\mu \hat{p}_\mu^*$ as

$$\sum_\mu \hat{p}^\mu \hat{p}_\mu^* = a^{-2} \sum_\mu [2 - (e^{-i\theta^\mu} + e^{i\theta^\mu})].$$

Therefore, we write a propagator on the lattice as

$$(\hat{p}_k - \hat{p}_j)^{-2} = a^2 \int_0^\infty d\alpha_{jk} \exp\Big\{ - \alpha_{jk} \sum_\mu [2 - (e^{-i(\theta_k^\mu - \theta_j^\mu)} + e^{i(\theta_k^\mu - \theta_j^\mu)})]\Big\}. \tag{2}$$

Each internal momentum carries an integration factor

$$(2\pi)^{-d} \int_{-\pi/a}^{\pi/a} d^d p = (2\pi a)^{-d} \int_\pi^\pi d^d\theta. \tag{3}$$

The momentum conservation holds on the lattice in the same way as in continuous space because

$$a^d \sum_x \exp\Big\{ i\Big(\sum p\Big) x\Big\} = (2\pi a)^d \sum_{n=-\infty}^{+\infty} \delta^d\Big(\sum\theta - 2\pi n\Big)$$

acts as an ordinary δ function on functions of θ defined on the interval $[-\pi,\pi]$ as the propagators.

Let us now make another change of variable by defining

$$\phi_j^\mu = \rho_j^\mu \exp(i\theta_j^\mu).$$

Of course, the measure will now contain new δ functions in order that $\rho_j^\mu = 1$. Then, the product of all the integrands in (2) of the propagators of a given planar graph G can be written

$$\prod_{\mu=1}^d \exp\{ -\mathbf{1} A \mathbf{1} + \boldsymbol{\phi}^\mu A \boldsymbol{\phi}^{\mu *} + \boldsymbol{\phi}^{\mu *} A_c \boldsymbol{\phi}_0^\mu + \boldsymbol{\phi}_0^{\mu *} A_c^T \boldsymbol{\phi}^\mu - \mathbf{1} A_c \mathbf{1}_0 - \mathbf{1}_0 A_c^T \mathbf{1}\}, \tag{4}$$

where $\boldsymbol{\phi}^\mu$ is a n-component vector, n being the number of internal loops; $\boldsymbol{\phi}_0^\mu$ is the vector corresponding to the external loops. Here, A is the $n\times n$ internal loops incidence matrix and A_c is the $n\times n_0$ incidence matrix connecting internal and external loops; A_c is therefore defined on the border of the graph; and $\mathbf{1}$ and $\mathbf{1}_0$ are vectors with unit elements of dimensions n and n_0, respectively.

The integration factors (3) are replaced by factors

$$\prod_{\mu=1}^d (2\pi a)^{-1} \int_0^\infty 2\rho^\mu \delta(\rho^{\mu^2} - 1) d\rho^\mu \int_\pi^{-\pi} d\theta^\mu = \prod_{\mu=1}^d (2\pi a)^{-1} (\pi)^{-1} \int d^2\phi^\mu \times \int_{-\infty}^{+\infty} \exp\{ -i\beta^\mu(\phi^\mu\phi^{\mu*} - 1)\} d\beta^\mu, \tag{5}$$

for each component of $\boldsymbol{\phi}^\mu$.

Collecting all powers of the coupling constant g and of a^2, we finally get the following expression for a planar one-particle irreducible graph G:

$$I_G = K_G(-ga^2)^{n_0}(ig^2a^6)^{n-1} \int_0^\infty DA \int_0^\infty DA_c \times \prod_{\mu=1}^d \int_{-\infty}^{+\infty} DB^\mu \int D\phi^\mu \times \exp\{ \mathbf{1}(iB^\mu - A)\mathbf{1} - \mathbf{1} A_c \mathbf{1}_0 - \mathbf{1}_0 A_c^T \mathbf{1} - \boldsymbol{\phi}^\mu (iB^\mu - A)\boldsymbol{\phi}^{\mu*} + \boldsymbol{\phi}^{\mu*} A_c \boldsymbol{\phi}_0^\mu + \boldsymbol{\phi}_0^{\mu*} A_c^T \boldsymbol{\phi}^\mu\}, \tag{6a}$$

$$K_G = \prod_{j=1}^{n_0} \Big\{ ia^2 \Big[\sum_\mu 2(1 - \cos(\theta_{0j}^\mu - \theta_{0j+1}^\mu))\Big]^{-1}\Big\} (2\pi a)^d \delta^d(0), \tag{6b}$$

where B^μ is a diagonal n-dimensional matrix with elements β^μ. The definition of the measure is given by

$$D\phi^\mu = \prod_{j=1}^n (\pi)^{-1} d^2\phi_j^\mu,$$

$$DB^\mu = \prod_{j=1}^n (2\pi a)^{-1} d\beta_j^\mu,$$

$$DA = \prod_{i<j} d\alpha_{ij},$$

$$DA_c = \prod_{i<j} d\alpha_{c_{ij}}.$$

Of course, the integration over $D\phi^\mu$ can be performed giving

$$I_G = K_G(-ga^2)^{n_0}(ig^2a^6)^{n-1} \times \int_0^\infty DA \int_0^\infty DA_c \prod_{\mu=1}^d \int_{-\infty}^{+\infty} DB^\mu \times \exp\{\mathbf{1}(iB^\mu - A)\mathbf{1} - \mathbf{1} A_c \mathbf{1}_0 - \mathbf{1}_0 A_c^T \mathbf{1}\} \times \det{}^{-1}(iB^\mu - A) \times \exp\{\boldsymbol{\phi}_0^{\mu*} A_c^T (iB^\mu - A)^{-1} A_c \boldsymbol{\phi}_0^\mu\}. \tag{7}$$

If we define A^μ and A_c^μ by

$$A^\mu = (iB^\mu)^{-1/2} A (iB^\mu)^{-1/2},$$

$$A_c^\mu = (iB^\mu)^{-1/2} A_c,$$

we have

$$\det{}^{-1}(iB^\mu - A)\exp\{\boldsymbol{\phi}_0^{\mu*} A_c^T (iB^\mu - A)^{-1} A_c \boldsymbol{\phi}_0^\mu\} = \det{}^{-1}(iB^\mu)\det{}^{-1}(\mathbb{1} - A^\mu) \times \exp\{\boldsymbol{\phi}_0^{\mu*} A_c^{\mu T} (\mathbb{1} - A^\mu)^{-1} A_c^\mu \boldsymbol{\phi}_0^\mu\}. \tag{8}$$

But we know that[7] (dropping here Lorentz indices for the sake of clarity)

$$\mathrm{Tr}[\exp(\mathbf{a}^+ A_c \boldsymbol{\phi}_0) :\exp\{\mathbf{a}^+(A - \mathbf{1})\mathbf{a}\}: \exp(\boldsymbol{\phi}_0^* A_c^T \mathbf{a})] = \Big[\prod_{j=1}^n (\pi)^{-1} \int d^2 z_j \exp(-|z_j|^2)\Big] \times \langle \mathbf{z}|\exp(\mathbf{a}^+ A_c \boldsymbol{\phi}_0) :\exp\{\mathbf{a}^+(A - \mathbf{1})\mathbf{a}\}: \exp(\boldsymbol{\phi}_0^* A_c^T \mathbf{a})|\mathbf{z}\rangle = \det{}^{-1}(\mathbf{1} - A)\exp\{\boldsymbol{\phi}_0^* A_c^T (\mathbf{1} - A)^{-1} A_c \boldsymbol{\phi}_0\}, \tag{9}$$

where $|\mathbf{z}\rangle = \Pi_{j=1}^n |z_j\rangle$ is a product of coherent states.[7] Creation and annihilation operators a_k^+ and a_j have been introduced which satisfy the usual commutation rules

$$[a_j, a_k^+] = \delta_{jk},$$

all other commutators being equal to zero.

The similarity between the kernel (8) and the right-hand side of (9) allows us to draw a connection between the general expression (7) for planar diagrams and the operator formalism which was used for string models. We therefore write

$$I_G = K_G(-ga^2)^{n_0}(ig^2a^6)^{n-1} \int_0^\infty DA \int_0^\infty DA_c \times \prod_{\mu=1}^d \int_{-\infty}^{+\infty} DB^\mu \Big(\prod_{j=1}^n (i\beta_j^\mu)\Big)^{-1} \times \exp\{\mathbf{1}(i\beta^\mu - A)\mathbf{1} - \mathbf{1}A_c\mathbf{1}_0 - \mathbf{1}_0 A_c^T \mathbf{1}\}\,\mathrm{Tr}\,\hat{O}^\mu, \tag{10a}$$

$$\hat{O}^\mu = \exp(\mathbf{a}^{\mu+} A_c^\mu \boldsymbol{\phi}_0^\mu) :\exp\{\mathbf{a}^{\mu+}(A^\mu - \mathbf{1})\mathbf{a}^\mu\}: \times \exp(\boldsymbol{\phi}_0^{\mu*} A_0^{\mu T} \mathbf{a}^\mu), \tag{10b}$$

$$[a_j^\mu, a_k^{+\nu}] = \delta^{\mu\nu}\delta_{jk}. \quad (10c)$$

(As usual, colons indicate a normal ordered expression.)

We shall now evaluate (10a) by taking into account the fact that (dropping again Lorentz indices)

$$:\exp(-a_j^+ a_j): = |0_j\rangle\langle 0_j|, \quad (11)$$

where $|0_j\rangle$ means the vacuum state for a "particle" of index j.

Expanding the exponentials giving the definition of the operator $\hat{O}$ and gathering for each term in the expansion the powers of a_j^+ and a_j for a definite value of the index j we get factors like (neglecting powers of α_{jk}, the factorials attached to them and Lorentz indices)

$$(i\beta_j)^{-(1/2)(p+q+\cdots)}(a_j^+)^p(a_j^+)^q\cdots|0_j\rangle\langle 0_j|(a_j)^l(a_j)^m \times\cdots(i\beta_j)^{-(1/2)(l+m+\cdots)}$$
$$= (i\beta_j)^{-(1/2)v_j}(a_j^+)^{v_j}|0_j\rangle\langle 0_j|(a_j)^{y_j}(i\beta_j)^{-(1/2)y_j}. \quad (12)$$

Taking the trace through the sandwiching factor,

$$(\pi)^{-1}\int d^2z_j \exp(-|z_j|^2)\langle z_j|\cdots|z_j\rangle,$$

transforms expression (12) into

$$v_j!(i\beta_j)^{-v_j}\delta_{v_j y_j}. \quad (13)$$

Performing the integration over the β_j^μ we get

$$v_j^\mu!(2\pi a)^{-1}\int_{-\infty}^{+\infty} d\beta_j^\mu \exp(i\beta_j^\mu)(i\beta_j^\mu)^{-v_j^\mu-1} = a^{-1}. \quad (14)$$

Remarkably, the dependence over the loop index j is now concentrated in the powers of α_{jk} and the factors $\delta_{v_j y_j}$.

If we associate to each power of α_{jk} an oriented arc **jk**, defined on the graph G' dual to G (see Ref. 8), we see that the factors $\delta_{v_j y_j}$ ensure that to each arc going into j there corresponds an outgoing arc stemming from j. Thus, to α_{jk} and α_{kj} correspond two opposite arcs **jk** and **kj**. This induces a dependence on "elementary" oriented paths. These paths never intersect themselves and each of them uses a given oriented arc only once.[9] There are closed and open paths. Open paths must necessarily end up at "external" indices corresponding to external loops.

Let us define by $\{P_j^\mu\}$ some[10] set of connected elementary oriented paths P_j^μ, defined on G', which do not cross each other, although they may overlap. (A given path can be contained more than once in $\{P_j^\mu\}$.) Here, $n_{\mathbf{kl}}^\mu$ will be the number of times the oriented arc **kl** is used by the paths of $\{P_j^\mu\}$. Using

$$\int_0^\infty d\alpha_{kl}\prod_{\mu=1}^d\left[\exp(-2\alpha_{kl})(\alpha_{kl})^{n_{\mathbf{kl}}^\mu+n_{\mathbf{lk}}^\mu}\right]$$
$$= (2d)^{-n_{kl}-1}n_{kl}!, \quad (15)$$

with $n_{kl} = \Sigma_{\mu=1}^d n_{\mathbf{kl}}^\mu + n_{\mathbf{lk}}^\mu$, one can perform all integrals, and one finds

$$I_G = K_G(-ga^2)^{n_0}(ig^2a^6)^{-1}[ig^2a^{6-d}]^n \times \sum_{\{P_{j_1}^1\},\{P_{j_2}^2\},\ldots,\{P_{j_d}^d\}}\left(\prod_{kl\in G'}(2d)^{-n_{kl}-1}n_{kl}!\right) \times\prod_{\mu=1}^d\left\{\left[\prod_{\mathbf{kl}\in G'}(n_{\mathbf{kl}}^\mu!)^{-1}\right]\left[\prod_{p,q\in\{P_{0j}^\mu\}}(\phi_{0p}^\mu\phi_{0q}^{\mu*})^{n_r^\mu}\right]\right\}, \quad (16)$$

with $\phi_{0p}^\mu = \exp(i\theta_{0p}^\mu), \phi_{0q}^\mu = \exp(i\theta_{0q}^\mu)$. Here, p and q are indices of external loops of G and $\{P_{0j}^\mu\}$ indicates the subset of the open paths of $\{P_{j\mu}^\mu\}$; and n_r^μ counts the number of times a path P_{0r}^μ is taken in $\{P_{0j}^\mu\}$. Let us remark that there are only a finite number of distinct elementary paths on a finite graph. Using this fact one can show the above series is certainly convergent if $d > M$ where M is the number of connected elementary (nonoriented) paths of G'.

One also remarks that the sum of all $n_{\mathbf{kl}}$, $\mathbf{kl}\in G'$, is the total length of all paths of a set $\{P_{j_1}^1\}, \{P_{j_2}^2\},\ldots,\{P_{j_d}^d\}$. Then, any planar Feynman graph G with a lattice cutoff is a function of nonlocal objects which can be interpreted as strings on the dual of G.

[1]G. 't Hooft, Phys. Lett. B **119**, 369 (1982).

[2]A. M. Polyakov, Phys. Lett. B **103**, 207 (1981); Phys. Lett. B **103**, 211 (1981).

[3]A. A. Migdal, Phys. Rep. **102**, 199 (1983), and references therein.

[4]J. L. Gervais and A. Neveu, Preprint LPTENS 83/37.

[5]H. Bohr and H. B. Nielsen, Nucl. Phys. B **227**, 547 (1983).

[6]In our opinion, the cleanest derivations of this kind have been made by G. 't Hooft, Nucl. Phys. B **72**, 461 (1974) and A. A. Migdal, Phys. Lett. B **96**, 333 (1980). See, however, the papers of H. B. Nielsen and P. Olesen, Phys. Lett. B **32**, 203 (1970); B. Sakita and M. A. Virasoro, Phys. Rev. Lett. **24**, 1146 (1970).

[7]V. Alessandrini, D. Amati, M. le Bellac, and D. Olive, Phys. Rep. **1**, 269 (1971). We remind the reader that a coherent state $|z_j\rangle$ is defined as

$$|z_j\rangle = \sum_{n=0}^\infty \frac{z_j^n}{n!}(a_j^+)^n|0\rangle.$$

[8]In graph theory, the graph G' dual to another one G is constructed by putting a vertex inside each loop of G and by drawing new edges across those of G and in one to one correspondence with them. Of course, this definition is only valid for planar graphs.

[9]Note that our definition of an elementary path differs from the usual one where a given elementary path is not allowed to go through a given vertex more than once.

[10]Every term in the expansion of $\hat{O}_\mu$ has to be counted only once, so we put restrictions on the P_j^μ's of a given $\{P_j^\mu\}$. To each area bounded by arcs we associate an integer $m \gtrless 0$ which is the sum of positively oriented elementary circuits minus the sum of negatively oriented elementary circuits surrounding it. If two areas of same index m touch each other at a vertex, no path should be crossed in going from one area to the other one through this vertex. Priority is given to the minimum m in case of conflict. Paths are not allowed to cross either themselves or each other. As a consequence, (i) paths of opposite orientations cannot overlap on any arc, (ii) the only path which contains both orientations of the same arc is the simple circuit formed by these oppositely oriented arcs, and (iii) a negatively oriented circuit cannot touch the inside of a positively oriented circuit or the outside of another negatively oriented circuit.

On the general solution to equations modeling a homogeneously broadened injection laser

I. Lerche
Gulf Research and Development Corporation, P. O. Box 37048, Houston, Texas 77236

(Received 24 May 1984; accepted for publication 1 March 1985)

The strongly coupled, nonlinear, differential equations which describe the amplification of modal intensities for propagation through a homogeneously broadened amplifier are shown to be generally reducible to a linear integral equation which is readily soluble by Laplace transform techniques. The mode intensities are shown to be generally expressible in terms of simple quadratures taken over the solution to the linear integral equation, which we also provide.

I. INTRODUCTION

Cassidy[1] has pointed out that, "To describe fully the spectral properties of an injection laser amplifier[2] or injection laser[3] the nonuniform distribution of the photon and the carrier densities along the length of the device must be taken into account." He goes on to note that "In steady-state, and for a perfect device in which the waveguide scattering/absorption loss is zero, the mth mode intensity $I_m^{\pm}$ in the $(+)$ forward- and $(-)$ backward-traveling directions is found by solving the set of nonlinear coupled differential equations[3,4]

$$\pm\frac{d}{dz}I_m^{\pm} = \Big\{g_m(N-n_1S)I_m^{\pm} + \beta B_m \times\Big[N+n_1\sum_i\sigma_i(I_i^{+}+I_i^{-})\Big]\Big\} \times\Big[S+\sum_i\sigma_i(I_i^{+}+I_i^{-})\Big]^{-1}.\text{"} \tag{1}$$

The notation is that employed previously,[1] but see Ref. 4 for a synopsis.

The set of equations (1) has so far proven very difficult to solve numerically or analytically[2,3] although, quite recently, an analytic solution has been given[1] when the lower level population is zero (i.e., $n_1 = 0$).

The purpose of this paper is to provide the general reduction of the set of equations (1) to a simple, linear integral equation so that these prevailing difficulties are, at the least, diminished and, in particular, removed; and to then provide the most general solution for the mode intensities in terms of simple quadratures over the solution to the integral equation, which we also give.

II. REDUCTION AND SOLUTION

Set

$$U(z)=\sum_i\sigma_i\,[I_i^{+}(z)+I_i^{-}(z)],\quad N'=N-n_1S,$$

and write Eqs. (1) in the form

$$\frac{d}{dz}I_m^{\pm} = \pm N'[g_mI_m^{\pm}(z)+\beta B_m] \times[S+U(z)]^{-1}\pm\beta B_m n_1. \tag{2}$$

Change dependent variables in Eq. (2) by setting

$$I_m^{\pm}(z) = -\beta B_m/g_m + U_m^{\pm}(z), \tag{3}$$

when

$$\frac{dU_m^{\pm}(z)}{dz} = \pm N'g_m U_m^{\pm}(z)[S+U(z)]^{-1}\pm\beta B_m n_1. \tag{4}$$

In place of z as the fundamental independent variable introduce τ with $d\tau = dz\,(S+U(z))^{-1}$ and with $\tau = 0$ on $z=0$. Then from (4) we obtain

$$U_m^{\pm}(\tau) = U_m^{\pm}(0)\exp(\pm N'g_m\tau)\pm\beta B_m n_1 \times\int_0^{\tau}[S+U(\tau')]\exp[\pm N'g_m(\tau-\tau')]d\tau', \tag{5}$$

where $U_m^{\pm}(0) = I_m^{\pm}(0)+\beta B_m/g_m$ and $I_m^{\pm}(0)$ are the initial $(z=0)$ conditions needed to specify the solutions to the first-order equations (1). Note from Eqs. (1) and (5) that $I_m^{-}(-z) = I_m^{+}(z)$ provided the values $I_m^{-}(0)$ and $I_m^{+}(0)$ are interchanged. Hence, without loss of generality, it suffices to consider the solution to Eq. (5) in $\tau\geqslant 0$.

We have

$$U(z)\equiv\sum\sigma_i(I_i^{+}(z)+I_i^{-}(z)) = \rho+\sum\sigma_i(U_i^{+}(z)+U_i^{-}(z)),$$

where $\rho = -2\beta\,\Sigma\sigma_i B_i/g_i$.

Multiply Eq. (5) by σ_m and sum over all modes in both the forward $(+)$ and backward $(-)$ directions to obtain

$$U(\tau) = \rho+F(\tau)+2\beta n_1\int_0^{\tau}U(\tau')H(\tau-\tau')d\tau', \tag{6}$$

with

$$F(\tau) = \sum_m\{\sigma_m\,[I_m^{+}(0)e^{N'g_m\tau}+I_m^{-}(0)e^{-N'g_m\tau}] + (2Sn_1/N')\beta(B_m/g_m)[\cosh(N'g_m\tau)-1]\} \tag{7a}$$

and

$$H(\tau) = \sum_m B_m\sigma_m\sinh[N'g_m\tau]. \tag{7b}$$

We recognize Eq. (6) as being of the convolutional type and subject to simple solution by the Laplace transform method.[5] Let L_σ denote the Laplace transform operation

 0022-2488/85/071858-02$02.50

and L_τ^{-1} its inverse. Then from (6) we obtain[6]

$$U(\tau) = L_\tau^{-1}[L_\sigma(\rho + F(\tau))[1 - 2\beta n_1 L_\sigma H(\tau)]^{-1}]. \quad (8)$$

From (3) and (5), we can express the solutions for the $I_m^\pm$ as functions solely of τ. The final transformation to provide an explicit representation for the $I_m^\pm$ as functions solely of z is provided through the quadrature

$$z = \int_0^\tau (S + U(\tau'))d\tau'. \quad (9)$$

This completes the solution to the problem. The only "messy" part is in computing $U(\tau)$ from Eq. (8) (Ref. 6), and the only numerical arithmetic part in most cases is the construction of the final transformation (9) between τ and z and, perhaps, of the integral in (5).

The analytic solution is valid for all parameter values and all initial conditions on the $I_m^\pm$.

Note that if $n_1 = 0$, $U(\tau)$ is explicitly given by (6) and the integral in (5) can be done simply in closed form, as can the transformation (9) between τ and z leading to the solution in parametric form

$$U(\tau) = \rho + F(\tau), \quad (10)$$

$$U_m^\pm(\tau) = U_m^\pm(0)\exp(\pm N' g_m \tau), \quad (11)$$

$$z = (S + \rho)\tau + \sum_m \frac{\sigma_m}{N' g_m} [I_m^+(0)(\exp(N' g_m \tau) - 1) - I_m^-(0)(\exp(-N' g_m \tau) - 1)], \quad (12)$$

with[1]

$$I_m^\pm = -\beta(B_m/g_m) + U_m^\pm(\tau).$$

ACKNOWLEDGMENTS

We thank Dr. D. Cassidy for catching some errors in a previous version of this paper, and for his critical and insightful comments.

We also thank Gulf Oil Corporation for their interest in this work, and for permission to publish.

[1]D. T. Cassidy, Appl. Phys. Lett. **44**, 489 (1984).
[2]D. Marcuse, J. Quantum Electron. **QE-19**, 63 (1983).
[3]K. Y. Lau and A. Yariv, Appl. Phys. Lett. **40**, 763 (1982).
[4]Cassidy[1] notes that "Equation (1) results from considering the time-dependent multimode rate equations for a two-level system in which the population of the lower level n_1 is taken to be constant and the time dependence of the upper level population n_2 is described by

$$\frac{dn_2}{dt} = N - (n_2 - n_1)\Sigma_i \sigma_i (I_i^+ + I_i^-) - \frac{n_2}{\tau},$$

which may be interpreted as explaining that the time rate of change of the upper level population n_2 equals the pumping rate N minus the loss of population due to stimulated and spontaneous emission events. In the equations the variables have the following definitions: τ is the total lifetime of the upper state to spontaneous transitions, B_m is the spontaneous emission profile and specifies which fraction of the spontaneous events falls within the wavelength interval of the mth mode, and σ_i is a saturation profile which denotes the efficiency of the ith mode intensity in reducing the inversion. Equation (1) results from setting the time derivative dn_2/dt equal to zero, solving for the steady-state population n_2 and letting the gain be proportional to the population inversion $(n_2 - n_1)$ and the spontaneous emission proportional to the upper level population n_2. Equation (1) allows for the possibility that the mth mode gain coefficient g_m may not equal σ_m, sets $\tau^{-1} = S$, and allows for the fact that only a fraction β of the spontaneous emission events couples into the laser mode."
[5]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1959).
[6]With $L_\sigma Q = \int_0^\infty e^{-\sigma\tau} Q(\tau)\, d\tau$, we have

$$L_\sigma H = \tfrac{1}{2}\sum_m \frac{B_m \sigma_m}{N'}\{(\sigma - N' g_m)^{-1} - (\sigma + N' g_m)^{-1}\}$$

and

$$L_\sigma[\rho + F] = \Big[\rho - \beta\sum_m \sigma_m B_m\, 2(Sn_1/N')g_m^{-1}\Big]\sigma^{-1} + \sum_m \sigma_m \times\{[I_m^+(0) + (\beta B_m/g_m)(Sn_1/N')](\sigma - N' g_m)^{-1} + [I_m^-(0) + (\beta B_m/g_m)(Sn_1/N')](\sigma + N' g_m)^{-1}\}$$

in $\mathrm{Re}\,\sigma > N' \max(g_m)$. Thus the inverse transform in Eq. (8), with $L_\tau^{-1} Q \equiv (2\pi i)^{-1}\int_c \exp(\sigma\tau) Q(\sigma)\, d\sigma$, merely amounts to sorting out the pole positions of $L_\sigma(\rho + F)$ and of $(1 - 2\beta n_1 L_\sigma H)^{-1}$, and of then writing down the expression for $U(\tau)$ in terms of the residues at the poles.

Flow reversal symmetry in a convective heat exchange model

Alberto Pignotti
TECHINT S.A., C.M. della Paolera 299, 1001 Buenos Aires, Argentina

(Received 3 January 1985; accepted for publication 22 February 1985)

Under some simplifying assumptions often used in the thermal design of heat exchangers, a system of two coupled linear homogeneous first-order partial differential equations with variable coefficients is used to describe a steady-state convective heat exchange process. Using an integral representation for the outlet temperatures as functions of the inlet ones, a symmetry property is proved for the transformation that inverts the velocity fields of both fluids. In particular, the average thermal effectiveness is shown to coincide for the direct and reversed processes.

I. INTRODUCTION

The thermal design of heat exchangers is often performed on the basis of the following idealized premises.

(a) Two fluids are in thermal contact through a flat two-dimensional surface (exchange surface). If the surface is not flat, a flat model should adequately represent it.

(b) A temperature is assigned to each of the fluids as a function only of the coordinates on the surface. Therefore, this temperature represents a steady-state value averaged over the direction normal to the exchange surface.

(c) Likewise, a two-dimensional velocity field is defined for each fluid on the exchange surface, and represents the corresponding velocity averaged over the direction normal to it. For the purpose of thermal design, these velocities are assumed to be known.

(d) The heat flow per unit area and time across the exchange surface is proportional to the difference between the temperatures of the two fluids at the point considered. The proportionality factor U (heat exchange coefficient) may be a function of the coordinates but not of the temperatures of the fluids.

(e) No change of phase is assumed to occur, and the specific heats of the fluids are temperature independent.

(f) The fluid densities may depend on the coordinates but not on the temperatures.

In this simplified scheme, the solution of the thermal problem is the expression of the outlet temperatures as functions of the inlet ones. Quite often, the inlet and outlet fluids are assumed to be bunched into a finite number of perfectly mixed streams. In such cases, the outlet temperatures can be expressed as linear combinations of the inlet ones. The coefficients of these linear combinations are the elements of a thermal matrix, the knowledge of which amounts to the solution of the thermal problem.[1,2]

It was recently argued[3] that the solution just mentioned is symmetric under flow reversal, i.e., under the inversion of the directions of flow of both fluids, keeping the geometry and the fluid properties unchanged. As a consequence of this symmetry, the matrix which describes the reverse process is, apart from some known factors, the transpose of the matrix which solves the original problem. In Ref. 3, arguments are given to support the validity of this property for all geometries with a finite number of perfectly mixed inlet and outlet streams, and practical consequences of the symmetry are discussed. In this article, we relax the assumption of perfect mixture of inlet and outlet streams, and allow for continuous inlet and outlet stream temperature distributions. The thermal matrix becomes a 2×2 matrix integral kernel, and we formulate and give a rigorous proof of its transformation properties under flow reversal. The equality of the average thermal effectiveness of the direct and reversed processes is obtained as a particular consequence of this more general approach.

II. FORMULATION OF THE PROBLEM

Let Σ be the simply connected two-dimensional exchange surface, σ its boundary, and $\mathbf{n}$ an external unit vector normal to σ on the plane of Σ. The known vector fields $\mathbf{v}_i(x,y)$, $i = 1,2$ are defined on Σ, and physically are equal to the velocities of the corresponding fluid times its density and specific heat. From the assumptions (e) and (f) of the previous section, it follows that

$$\operatorname{div} \mathbf{v}_i(x,y) = 0, \quad i = 1,2. \tag{1}$$

Figure 1 shows the basic features of a configuration obeying the above description.

On the boundary σ we define

$$\rho_i = \mathbf{v}_i \cdot \mathbf{n}, \quad i = 1,2 \tag{2}$$

and we consider two different partitions of σ

$$\sigma \equiv \sigma_1^- \cup \sigma_1^0 \cup \sigma_1^+ \equiv \sigma_2^- \cup \sigma_2^0 \cup \sigma_2^+, \tag{3}$$

such that a point (x,y) on the boundary belongs to either σ_i^-, σ_i^0, or σ_i^+ depending on whether

$$\rho_i(x,y) < 0, \quad \rho_i(x,y) = 0, \quad \text{or } \rho_i(x,y) > 0, \tag{4}$$

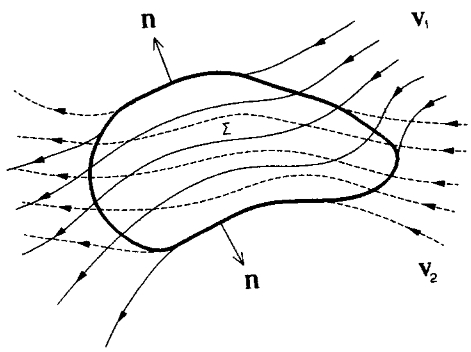


FIG. 1. Basic features of the heat exchange process considered. Field lines corresponding to the vector fields $\mathbf{v}_1$ and $\mathbf{v}_2$ are shown in full and broken lines. The exchange surface Σ and the external unit normal $\mathbf{n}$ are also indicated.

 0022-2488/85/071860-03$02.50

respectively. Therefore, σ_i^- and σ_i^+ denote the parts of the boundary through which the ith fluid enters and leaves the domain Σ.

The equations governing the steady-state solution for the process described in the Introduction are

$$\mathbf{v}_1(x,y)\cdot\nabla T_1(x,y) + U(x,y)[T_1(x,y) - T_2(x,y)] = 0, \tag{5a}$$

$$\mathbf{v}_2(x,y)\cdot\nabla T_2(x,y) + U(x,y)[T_2(x,y) - T_1(x,y)] = 0. \tag{5b}$$

This is a set of coupled, linear, homogeneous, first-order, partial differential equations for the unknown temperatures T_1 and T_2, with variable coefficients $\mathbf{v}_1$, $\mathbf{v}_2$, and U. The boundary conditions are

$$T_i(x,y) = t_i^-(s_i^-) \quad \text{for } (x,y)\in\sigma_i^-, \quad i=1,2, \tag{6}$$

where s_i^- are the arc-length coordinates defined on σ_i^-, and t_i^- are arbitrary but known inlet temperature distributions. In a similar way, the unknown temperatures $T_i(x,y)$ on the outlet boundaries σ_i^+ are written as functions of the variables s_i^+ and are denoted $t_i^+(s_i^+)$.

The solution of the thermal problem can be expressed in the form of the following linear relation between the outlet and inlet temperature distributions:

$$t_i^+(s_i^+) = \sum_{j=1}^{2}\int_{\sigma_j^-} K_{ij}(s_i^+, s_j^-)\, t_j^-(s_j^-)\,\rho_j(s_j^-)\, ds_j^-. \tag{7}$$

This equation is the generalization of the matrix relation between inlet and outlet discrete temperatures, to the case of continuous temperature distributions. The ρ factor, which could be omitted in Eq. (7), thus implying a different definition of the matrix K, has been included explicitly in order to simplify the results to be found later. We do not deal here with the problem of proving Eq. (7) from Eqs. (5a) and (5b), but, rather, assume its validity and proceed to consider a different problem, which may be called a dual problem.

The variables of the dual problem are labeled as the original variables with a ^ symbol on top. The dual problem, depicted in Fig. 2, is defined as having the same geometry as the original one ($\hat\Sigma \equiv \Sigma$, $\hat{\mathbf{n}} = \mathbf{n}$) and the same properties of the fluids involved, but opposite velocity fields. How this inversion of the velocity fields can be achieved is not discussed here. With these assumptions

$$\hat{\mathbf{v}}_i = -\mathbf{v}_i, \quad \hat\rho_i = -\rho_i, \quad \hat\sigma_i^+ \equiv \sigma_i^-, \quad \hat\sigma_i^0 \equiv \sigma_i^0,$$
$$\hat\sigma_i^- \equiv \sigma_i^+, \quad \hat s_i^+ = s_i^-, \quad \hat s_i^- = s_i^+. \tag{8}$$

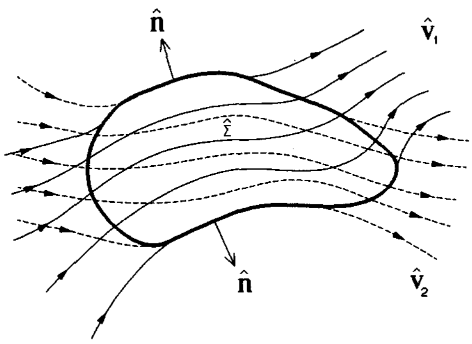


FIG. 2. Basic features of the dual problem corresponding to the exchange process of Fig. 1.

The boundary conditions are now the functions $\hat t_i^-(\hat s_i^-)$, defined on $\hat\sigma_i^-$, and the solution can be expressed, in complete analogy to Eq. (7), as

$$\hat t_i^+(\hat s_i^+) = \sum_{j=1}^{2}\int_{\hat\sigma_j^-} \hat K_{ij}(\hat s_i^+, \hat s_j^-)\, \hat t_j^-(\hat s_j^-)\,\hat\rho_j(\hat s_j^-)\, d\hat s_j^-. \tag{9}$$

It is clear that, because of the generality of the geometry and the vector fields $\mathbf{v}_i$ involved, the original and the dual problems are quite different, except in particular instances in which the original configuration possesses a geometrical symmetry. Indeed, examples can be found in the literature in which considerable efforts were devoted to independently solving a heat exchange problem and its dual problem.[4,5] The aim of this work is to prove that, once the direct problem has been solved, and, therefore, $K_{ij}(s_i^+, s_j^-)$ is known, the solution of the dual problem is just given by

$$\hat K_{ij}(s_i^-, s_j^+) = K_{ji}(s_j^+, s_i^-). \tag{10}$$

This equation is the expression of flow-reversal symmetry, i.e., it gives the transformation property of the matrix K under the inversion of the velocity fields.

III. PROOF OF THE SYMMETRY

In the dual problem the unknown temperatures $\hat T_i$ satisfy equations analogous to (5a) and (5b). Remembering $\hat{\mathbf{v}}_i = -\mathbf{v}_i$ and omitting the arguments we can write

$$\mathbf{v}_1\cdot\nabla\hat T_1 - U(\hat T_1 - \hat T_2) = 0, \tag{11a}$$

$$\mathbf{v}_2\cdot\nabla\hat T_2 - U(\hat T_2 - \hat T_1) = 0. \tag{11b}$$

Multiplying Eqs. (5a) by $\hat T_1$, (5b) by $\hat T_2$, (11a) by T_1, and (11b) by T_2 and adding the resulting equations, we find

$$\mathbf{v}_1\cdot\nabla(T_1\hat T_1) + \mathbf{v}_2\cdot\nabla(T_2\hat T_2) = 0, \tag{12}$$

whence, using Eq. (1),

$$\mathrm{div}(T_1\hat T_1\mathbf{v}_1 + T_2\hat T_2\mathbf{v}_2) = 0. \tag{13}$$

Integrating over Σ and using Gauss' theorem we obtain

$$\sum_{i=1}^{2}\left[\int_{\sigma_i^-} t_i^-\hat t_i^+\rho_i\, ds_i^- + \int_{\sigma_i^+} t_i^+\hat t_i^-\rho_i\, ds_i^+\right] = 0. \tag{14}$$

Replacing t_i^+ *and* $\hat t_i^+$ from Eqs. (7) and (9) into (14) and using Eq. (8) yields

$$\sum_{i,j=1}^{2}\int_{\sigma_i^+} \hat t_i^-(s_i^+)\rho_i(s_i^+)\, ds_i^+ \int_{\sigma_j^-} K_{ij}(s_i^+, s_j^-)\, t_j^-(s_j^-)$$
$$\times\rho_j(s_j^-)\, ds_j^- - \sum_{i,j=1}^{2}\int_{\sigma_i^-} t_i^-(s_i^-)\rho_i(s_i^-)\, ds_i^-$$
$$\times\int_{\sigma_j^+}\hat K_{ij}(s_i^-, s_j^+)\,\hat t_j^-(s_j^+)\rho_j(s_j^+)\, ds_j^+ = 0. \tag{15}$$

Interchanging the dummy indices i and j in the first summation, we finally obtain

$$\sum_{i,j=1}^{2}\int_{\sigma_i^-} t_i^-(s_i^-)\rho_i(s_i^-)\, ds_i^- \int_{\sigma_j^+}\hat t_j^-(s_j^+)\rho_j(s_j^+)$$
$$\times[K_{ji}(s_j^+, s_i^-) - \hat K_{ij}(s_i^-, s_j^+)]\, ds_j^+ = 0. \tag{16}$$

From this equation, because the inlet temperatures $t_i^-(s_i^-)$ and $\hat t_j^-(s_j^+)$ are arbitrary, Eq. (10) follows.

IV. CONSEQUENCES OF THE SYMMETRY

Equation (10) has the immediate consequence of providing the solution of the dual problem once the solution of the original problem is known. In particular, this relation between the two problems implies that the average effectiveness of both processes is the same. Before reaching this conclusion, however, two integral properties of the representation (7) should be discussed.

It is apparent that Eqs. (5) are invariant under the transformation $T_i(x,y) \to T_i(x,y) + \delta$, which, when applied to the solution (7), implies the first integral relation

$$\sum_{j=1}^{2} \int_{\sigma_j^-} K_{ij}(s_i^+, s_j^-) \rho_j(s_j^-)\, ds_j^- = 1, \quad i = 1,2, \tag{17}$$

valid for all values of s_i^+.

The same argument applied to the solution of Eqs. (11a) and (11b) using Eqs. (8) and (10), gives

$$\sum_{i=1}^{2} \int_{\sigma_i^+} K_{ij}(s_i^+, s_j^-) \rho_i(s_i^+)\, ds_i^+ = -1, \quad j = 1,2. \tag{18}$$

Multiplying this equation by $\rho_j(s_j^-)$, integrating over s_j^-, using Eq. (17) and the relation

$$\int_{\sigma_j^+} \rho_j(s_j^+) ds_j^+ + \int_{\sigma_j^-} \rho_j(s_j^-) ds_j^- = 0, \tag{19}$$

which follows from Eqs. (1) and (2), we find

$$\int_{\sigma_1^+} \rho_1(s_1^+) ds_1^+ \int_{\sigma_2^-} K_{12}(s_1^+, s_2^-) \rho_2(s_2^-)\, ds_2^- = \int_{\sigma_2^+} \rho_2(s_2^+) ds_2^+ \int_{\sigma_1^-} K_{21}(s_2^+, s_1^-) \rho_1(s_1^-)\, ds_1^- . \tag{20}$$

This is the second integral relation sought. It establishes that K, when averaged over its variables with weights given by ρ_i, is a symmetric matrix.

We now define the average effectiveness of the heat exchange process as

$$P = (\bar{t}\,_1^+ - t\,_{1,c}^-)/(t\,_{2,c}^- - t\,_{1,c}^-), \tag{21}$$

where

$$\bar{t}\,_1^+ = \int_{\sigma_1^+} t\,_1^+(s_1^+) \rho_1(s_1^+)\, ds_1^+ \left(\int_{\sigma_1^+} \rho_1(s_1^+) ds_1^+ \right)^{-1} \tag{22}$$

is the average outlet temperature of fluid 1, and $t\,_{1,c}^-$ and $t\,_{2,c}^-$ are any constant inlet temperatures. Therefore, for the purpose of defining P, the inlet fluids are assumed to be separately perfectly mixed. The average effectiveness so defined turns out to be independent of the inlet temperatures. Indeed, from Eq. (21), (22), and (17) we obtain

$$P = \int_{\sigma_1^+} \rho_1(s_1^+) ds_1^+ \int_{\sigma_2^-} K_{12}(s_1^+, s_2^-) \rho_2(s_2^-)\, ds_2^- \times \left(\int_{\sigma_1^+} \rho_1(s_1^+) ds_1^+ \right)^{-1}. \tag{23}$$

In the same fashion, for the dual process we have

$$\begin{aligned} \hat{P} &= \int_{\hat{\sigma}_1^+} \hat{\rho}_1(\hat{s}_1^+) d\hat{s}_1^+ \int_{\hat{\sigma}_2^-} \hat{K}_{12}(\hat{s}_1^+, \hat{s}_2^-) \hat{\rho}_2(\hat{s}_2^-)\, d\hat{s}_2^- \times \left(\int_{\hat{\sigma}_1^+} \hat{\rho}_1(\hat{s}_1^+) d\hat{s}_1^+ \right)^{-1} \\ &= -\int_{\sigma_1^-} \rho_1(s_1^-) ds_1^- \int_{\sigma_2^+} K_{21}(s_2^+, s_1^-) \rho_2(s_2^+)\, ds_2^+ \times \left(\int_{\sigma_1^-} \rho_1(s_1^-) ds_1^- \right)^{-1}. \end{aligned} \tag{24}$$

From Eqs. (19), (20), (23), and (24), the equality $P = \hat{P}$ follows.

V. CONCLUSION

This work provides the first mathematically rigorous proof of the validity of the flow-reversal symmetry originally proposed in Ref. 3. In addition, the treatment is generalized to continuous temperature distributions and proved to hold for arbitrary geometries under the assumptions spelled out in the Introduction.

Though simple to state, the symmetry property discussed here is by no means intuitive, because, except for trivial cases, it relates configurations that are apparently quite different, such as those illustrated in Figs. 1 and 2. A particular consequence of this analysis is the invariance of the average thermal effectiveness under the inversion of the velocities of both fluids.

ACKNOWLEDGMENTS

The author wants to acknowledge very valuable suggestions from Professor Alberto Calderón of the University of Buenos Aires. He is also indebted to Professor Sergio Pissanetzky of the Instituto de Física Balseiro and Professor Richard Macamy of Carnegie Mellon University for stimulating discussions.

[1]J. D. Domingos, Int. J. Heat Mass Transfer **12**, 537 (1969).
[2]A. Pignotti, ASME J. Heat Transfer **106**, 352 (1984).
[3]A. Pignotti, ASME J. Heat Transfer **106**, 361 (1984).
[4]A. J. V. Underwood, J. Inst. Pet. Technol. **20**, 145 (1934).
[5]K. Gardner and J. Taborek, AIChE J. **23**, 777 (1977).