Gaussian states on the algebra of the infinite classical system

J. Naudts
Universiteit Antwerpen, B-2610 Wilrijk, Antwerp, Belgium

(Received 28 December 1979; accepted for publication 28 March 1980)

A class of states on the algebra of the infinite classical system is characterized by the vanishing of the higher order ($n > 2$) truncated correlation functions. The states are called Gaussian. It is shown how liquids and crystals can be described by Gaussian states.

PACS numbers: 02.10. – v, 61.20.Gy, 61.50.Em

I. INTRODUCTION

An infinite system of classical particles can be described in terms of correlation functions, or in terms of states on a C^*-algebra.[1] Especially interesting is the C^*-algebra of Ref. 2. It allows to construct states in an elegant and concise way. Quasifree states on this algebra are determined by a positive linear functional on a testspace $\tilde{D}$. It is shown in the present paper that Gaussian states are determined by a linear functional on $\tilde{D}$ describing the density of particles and a sesquilinear form on $\tilde{D}$ describing the correlations. A form of Wick's theorem is proved.

In Sec. 4 liquids are considered. For any density ρ_0 and radial distribution function $g(r)$ a Gaussian state exists provided that the static structure factor is strictly positive everywhere.

In Sec. 5 a description of crystals is given in terms of a periodic density function $\rho^{(1)}(q)$ and of the correlation function $\rho^{(2)}(q_1,q_2)$. The static structure factor alone contains not enough information to calculate this pair correlation and one is forced to introduce a family of "structure factors" $K_{bb'}(k)$ which are functions over the Brillouin zone and are labelled by vectors b,b' of the reciprocal lattice. The existence of a Gaussian state then is assured under conditions similar to the ones in the case of liquids.

The harmonic crystal model is treated as example. A simple description of this model in terms of densities and density correlations is only possible in the one-phonon approximation.

Finally let us indicate the similarities of the present work with that of van Hemmen[3] who describes the harmonic crystal in terms of Gaussian processes. Essential in that approach is a generalized version of the finite dimensional Bochner theorem. It allows to introduce a Gaussian measure on a phase space of the system with infinitely many degrees of freedom. In the present approach such a measure on a phase space is absent. Its counterpart is obtained by the integral decomposition of the Gaussian state into pure states.

A direct link exists between the theory of stochastic processes[4] and the formalism of Ref. 2. It is shortly discussed at the end of Sec. 3.

2. ALGEBRA OF OBSERVABLES

An infinitely extended system with nonvanishing density of particles can be described in the algebraic approach. The C^*-algebra of observables which we will use is the classical analog of the Weyl algebra in quantum mechanics. See Ref. 2.

Let K be the set of subsets X of $\mathbb{R}^3$ which satisfy the condition that for each bounded part V of $\mathbb{R}^3$ the intersection $V \cap X$ contains an at most finite number of points. Then K is the *configuration space* of the infinite system. Remark that unlike in Ref. 2 the particles are described by their coordinates only.

The *testspace* D is the set of real C^∞-functions with bounded support in $\mathbb{R}^3$. For each f in D a function Sf on K is defined by:

$$(Sf)(X) = \sum_{x \in X} f(x).$$

Because the support of f is bounded at most a finite number of terms in the summation do not vanish.

For each f in D a *Weyl function* $W(f)$ is defined by:

$$W(f) = \exp iSf.$$

In fact W is a map of D into the von Neumann algebra $B(K,\mathbb{C})$ of bounded functions on configuration space. One has:

$$W(f)^* = W(-f), \quad W(0) = 1,$$

and

$$W(f)W(g) = W(f+g).$$

The C^*-algebra $\mathcal{A}$ of the infinite classical system now is generated by the function $W(f), f \in D$. For more details see Ref. 2.

3. GAUSSIAN STATES

A class of states on $\mathcal{A}$ is introduced and a characterization is given in terms of truncated correlation functions. Denote $\tilde{D}$ the complex algebra generated by D:

$$\tilde{D} = \{ f + ig \mid f, g \in D \}.$$

Following Ref. 2 we say that a state ω on $\mathcal{A}$ is C^∞ if for all f and g in D the map $\lambda \in \mathbb{R} \to \omega(W(f+\lambda g))$ is infinitely differentiable.

Theorem 1: Let $\omega_0 : D \to \mathbb{R}$ be a linear map. Let s be a nonnegative definite sesquilinear form on $\tilde{D}$. Then a state ω on $\mathcal{A}$ is uniquely defined by the relation

$$\omega(W(f)) = \exp\{ i\omega_0(f) - \tfrac{1}{2}s(f,f) \} \tag{1}$$

for all $f \in D$. Moreover the state ω is C^∞.

Proof: By linearity and continuity ω extends to a linear map on $\mathcal{A}$. Normalization follows from $\omega(1) = \omega(W(0)) = 1$. It rests to show that ω is positive. Let $\xi_1, \cdots, \xi_n$ in $\mathbb{C}$ and $f_1, \cdots, f_n$ in D. One has:

 0022-2488/81/020221-05$1.00

$$\sum_{ij} \xi_i \bar{\xi}_j \ \omega(W(f_i)^* \ W(f_j))$$

$$= \sum_{ij} \xi_i \bar{\xi}_j \ \omega(W(f_j - f_i))$$

$$= \sum_{ij} \eta_i \eta_j \exp s(f_i, f_j), \tag{2}$$

with

$$\eta_i = \xi_i \exp(-i\omega_0(f_i) - \tfrac{1}{2}s(f_i, f_j)).$$

From Schur's theorem follows that (2) is positive. Hence ω is a state on $\mathcal{A}$. One immediately verifies that it has the C^∞-property.

Definition 2: A state ω on $\mathcal{A}$ is called *Gaussian* if there exists a linear map $\omega_0: D \to \mathbb{R}$ and a nonnegative definite sesquilinear form s on D such that relation (1) is satisfied for all f in D.

Let $(\Pi, \mathcal{H}, \Omega)$ be the G.N.S.-representation of a Gaussian state ω. Because ω is C^∞ all *field operators* [2] $B(f), f \in D$ exist and satisfy: $\Pi(W(f)) = \exp iB(f)$ and $B(f_1) \cdots B(f_n)\Omega$ belongs to the domain of $B(f)$.

Truncated correlation functions $\omega_T(f_1, \cdots, f_n)$ are defined by the recurrence relation:

$$(\Omega, B(f_1) \cdots B(f_n)\Omega) = \Sigma \omega_T(f_{i_1}, \cdots) \cdots \omega_T(\cdots, f_{i_n}),$$

where the summation is over all possible partitions $(i_1, \cdots) \cdots (\cdots, i_n)$ of $\{1, \cdots, n\}$ with the original order preserved within each cluster. In particular

$$(\Omega, B(f)\Omega) = \omega_T(f)$$

and

$$(\Omega, B(f_1)B(f_2)\Omega) = \omega_T(f_1)\omega_T(f_2) + \omega_T(f_1, f_2).$$

Lemma 3: For $n \geqslant 3$ one has

$$(\Omega, \prod_{j=1}^{n} B(f_j)\Omega) = \omega_0(f_n) \ (\Omega, \prod_{j=1}^{n-1} B(f_j)\Omega) + \sum_{j=1}^{n-1} s(f_j, f_n) \left(\Omega, \prod_{k=1, k \neq j}^{n-1} B(f_k)\Omega\right).$$

The proof of the lemma is straightforward. The following characterization of Gaussian states is obtained. It is a version of Wick's theorem.

Theorem 4: Let ω be a C^∞-state on $\mathcal{A}$. The following two statements are equivalent: (a) ω is Gaussian; (b) $\omega_T(f_1, \cdots, f_n) = 0$ for $n \geqslant 3$. Let ω_0 and s be the linear map respectively the sesquilinear form which appears in the defining relation (1). Then one has for all $f_1, f_2 \in D$: (c) $\omega_1(f_1) = \omega_0(f_1)$ and $\omega_1(f_1 f_2) = s(f_1, f_2)$.

Proof: (a)$\Rightarrow$(b) and (a)$\Rightarrow$(c) follows from the lemma. (b)$\Rightarrow$(a).

Let ω_0 be given by

$$\omega_0(f) = \omega_T(f) = (\Omega, B(f))\Omega).$$

Clearly ω_0 is a linear function on $\tilde{D}$. This follows from the linearity of the field operator.

Let s be given by

$$s(f_1 + if_2, g_1 + ig_2) = \omega_T(f_1, g_1) + \omega_T(f_2, g_2) + i\{\omega_T(f_2, g_1) - \omega_T(f_1, g_2)\}.$$

Then s is a sesquilinear form on $\tilde{D}$.

Now one has $\omega_T(f_1, f_2) = \omega_T(f_2, f_1)$ because the field operators commute. Hence:

$$s(f_1 + if_2, f_1 + if_2) = \omega_T(f_1, f_1) + \omega_T(f_2, f_2).$$

But from Schwarz inequality one has:

$$\omega_T(f_i, f_i) = (\Omega, B(f_i)^2\Omega) - (\Omega, B(f_i)\Omega)^2 \geqslant 0.$$

Therefore s is nonnegative definite.
Because ω is C^∞ one has

$$\omega(W(f)) = \sum_{n=0}^{\infty} (i^n/n!)(\Omega, B(f)^n\Omega).$$

From condition (b) and with the former definition of ω_0 and s one obtains

$$\omega(W(f)) = \exp\{i\omega_0(f) - \tfrac{1}{2}s(f, f)\}.$$

Therefore ω is Gaussian.
Finally we indicate a relation with the the theory of stochastic processes. Let $f_1, \cdots, f_n$ be test functions in D.
The function

$$x = (x_1, \cdots, x_n) \to \omega(W(\sum_i x_i f_i))$$

is the characteristic function of a Gaussian process with random variables $f_1, \cdots, f_n$. It can be written as

$$\exp(imx - \tfrac{1}{2}xLx),$$

with the mean m given by $m_i = \omega_0(f_i)$ and the covariance matrix L given by $L_{ij} = s(f_i, f_j)$.

The underlying probability space is the one-particle configuration space $\mathbb{R}^3$ equipped with the measure induced by the linear functional ω_0.

4. LIQUIDS

The physical way of describing a liquid or a crystal will be by use of a density functional $\rho^{(1)}(q)$ and a density correlation function $\rho^{(2)}(q_1, q_2)$. The following relations give the formal connection with the linear functional ω_0 and the sesquilinear form s which by definition determine a Gaussian state of the system:

$$\omega_0(f) = \int dq f(q)\rho^{(1)}(q)$$

and

$$s(f_1, f_2) = \omega_0(f_1 f_2) - \int dq_1 \int dq_2 \ f_1(q_1) f_2(q_2) \times [\rho^{(1)}(q_1)\rho^{(1)}(q_2) - \rho^{(2)}(q_1, q_2)].$$

An apparent advantage of working with test functions f_i in D is that one can allow the functions $\rho^{(1)}$ and $\rho^{(2)}$ to be generalized functions. Hence the definition of ω_0 and s by foregoing relations in most applications will not be too difficult to prove. The main requirement is the positivity of the sesquilinear form s.

In the case of a dense gas or of a liquid one has translational invariance of the system. The density correlation functions $\rho^{(1)}$ and $\rho^{(2)}$ are of the form

$$\rho^{(1)}(q) = \rho_0$$

and

$$\rho^{(2)}(q_1, q_2) = \rho_0^2 g(q_1 - q_2).$$

ρ_0 is the density of the system. $g(q)$ is called the radial distribution function, and usually it is assumed to be positive.

It is conventional to denote $h(q) = g(q) - 1$. For convenience we will work with Fourier transformed quantities:

$$\tilde{h}(Q) = \int dq\, e^{iqQ} h(q).$$

The static structure factor[5] is given by:

$$S(Q) = 1 + \rho_0 \tilde{h}(Q).$$

Remark that if f is a test function in D then its Fourier transform $\tilde{f}$ is a bounded function which falls off at infinity faster than any polynomial.

Proposition 5: Let ρ_0 be a constant. Let $\tilde{h}(Q)$ be a real function on $\mathbb{R}^3$ which is locally integrable. Assume there exists a positive integer n such that $|Q|^{-n}\tilde{h}(Q)$ is integrable outside the unit ball of $\mathbb{R}^3$. Assume also that $S(Q) > 0$ almost everywhere. Then a linear functional ω_0 on D and a positive definite sesquilinear form s on $\tilde{D}$ are given by the relations

$$\omega_0(f) = \rho_0 \tilde{f}(0)$$

and

$$s(f_1, f_2) = \omega_0(f_1 f_2) + (2\pi)^{-3}\rho_0^2 \int dQ\, \tilde{f}_1(Q)\, \tilde{f}_2(-Q)\tilde{h}(Q).$$

In particular it corresponds a Gaussian state on $\mathcal{A}$.

Proof: Let us show that the integration in the defining relation s is meaningful. Denote Λ the unit ball of $\mathbb{R}^3$. Let M be a bounded region in $\mathbb{R}^3$. Then $\int_M dQ\, \tilde{f}_1(Q)\tilde{f}_2(-Q)\,\tilde{h}(Q)$ is well defined because $\tilde{h}$ is locally integrable and $\tilde{f}_1$ and $\tilde{f}_2$ are Fourier transforms of infinitely differentiable functions with compact support and hence are continuous and bounded. Suppose now $M \cap \Lambda = \phi$. Then one has:

$$\left|\int_M dQ\, \tilde{f}_1(Q)\tilde{f}_2(-Q)\tilde{h}(Q)\right| \leqslant \chi(M)\|\tilde{f}_2\|_\infty \int_{C\Lambda} dQ\, |Q|^{-n}|\tilde{h}(Q)|,$$

with $\chi(M) = \sup\{|Q|^n |\tilde{f}_1(Q)| : Q \in M\}$. The convergence of the integral in the definition of s follows because the function $\tilde{f}_1(Q)$ falls off at infinity faster than any polynomial. The proof of the proposition now is straightforward. In particular the positivity of s follows from the relation

$$s(f_1, f_2) = (2\pi)^{-3}\rho_0 \int dQ\, \tilde{f}_1(Q)\tilde{f}_2(-Q)\, S(Q).$$

Example: A gas is described by the two-body interaction potential $V(r)$. An obvious choice of the radial distribution function g is

$$g(q) = \exp(-V(|q|)/kT)$$

(k is the Boltzmann factor, and T is the temperature). It is the low density limit of the solution of the Kirkwood equation.

Under reasonable assumptions on the potential V the function $h' = g - 1$ is integrable. At high temperature and low density one has

$$\rho_0 \int dq\, |h(q)| = \rho_0 \int dq\, |\exp(-V(|q|)/kT) - 1| < 1.$$

The Fourier transformed function $\tilde{h}(Q)$ is bounded. It satisfies the conditions of the foregoing proposition. Hence a Gaussian state on $\mathcal{A}$ describes the gas.

The energy density U is given by

$$U(q_1) = \tfrac{1}{2}\int dq_2\, \rho^{(2)}(q_1, q_2) V(|q_1 - q_2|).$$

One finds

$$U = \tfrac{1}{2}\rho_0^2 \int dq \exp(-V(|q|)/kT)\, V(|q|).$$

5. CRYSTALS

We consider Gaussian states on $\mathcal{A}$ which are periodic under translations of $\mathbb{R}^3$. For example they can describe a crystal within the harmonic approximation.

The description of a crystal in terms of a density function $\rho^{(1)}(q)$, which is a superposition of Gaussians centered on lattice sites, and a density correlation function $\rho^{(2)}(q_1, q_2)$ is unusual. The quantities which are accessible by experiment (x ray and neutron scattering) are $\rho^{(1)}(q)$ and the average of $\rho^{(2)}(q_1, q_2)$ over one unit cell Λ_0:

$$\frac{1}{V(\Lambda_0)} \int_{\Lambda_0} ds\, \rho^{(2)}(q_1 + s, q_2 + s).$$

Expressions for the latter quantity can be found in the literature.[5]

We consider a Bravais lattice generated by the vectors a_1, a_2, and a_3. They form a non-degenerate matrix A. The lattice sites are given by: An, $n \in \mathbb{Z}^3$. Denote B the (first) *Brillouin zone*. It is (isomorphic to) the dual of the lattice group. A pair of Fourier transforms is defined by

$$\tilde{c}(k) = \sum_n e^{ikAn} c_n,$$

$$c_n = \frac{1}{V(B)} \int_B dk\, e^{-ikAn} \tilde{c}(k).$$

Denote Γ the reciprocal lattice. Denote Λ_0 the unit cell of the lattice. The density $\rho^{(1)}(q)$ is a continuous periodic function and hence it can be written as

$$\rho^{(1)}(q) = \sum_{b \in \Gamma} e^{ibq} \tilde{\rho}_b,$$

with

$$\tilde{\rho}_b = \frac{1}{V(\Lambda_0)} \int_{\Lambda_0} dq\, e^{-ibq} \rho^{(1)}(q).$$

The density correlation $\rho^{(2)}(q_1, q_2)$ has the periodicity property

$$\rho^{(2)}(q_1 + An, q_2 + An) = \rho^{(2)}(q_1, q_2).$$

One therefore can write:

$$\rho^{(2)}(q_1, q_2) - \rho^{(1)}(q_1)\rho^{(1)}(q_2) = (2\pi)^{-3} \sum_{b, b' \in \Gamma} e^{-ibq_1} e^{+ib'q_2} \times \int_B dk\, e^{ik(q_1 - q_2)} K_{bb'}(k),$$

with

$$K_{bb'}(k) = [1/V(\Lambda_0)] \int_{\mathbf{R}^3} dq'\, e^{iq'(k - b')} \int_{\Lambda_0} dq\, e^{iq(b - b')} \times [\rho^{(2)}(q, q + q') - \rho^{(1)}(q)\rho^{(1)}(q + q')].$$

In particular one has

$$[1/V(\Lambda_0)] \int_{\Lambda_0} ds\, [\rho^{(2)}(q_1 + s, q_2 + s) - \rho^{(1)}(q_1 + s)\rho^{(1)}(q_2 + s)]$$

$$= (2\pi)^{-3} \sum_b \int_B dk\, e^{i(k-b)(q_1-q_2)} K_{bb}(k).$$

Proposition 6: Let $(\tilde{\rho}_b)_{b\in\Gamma}$ be a bounded sequence of reals. Let $\{K_{bb'}(k): b,b' \in \Gamma\}$ be a matrix of integrable functions on B. Assume a uniform bound M exists:

$$\int_B |K_{bb'}(k)|\, dk \leqslant M$$

independent of $b,b' \in \Gamma$. Assume for almost all $k \in B$ that the matrix $\tilde{\rho}_{b'-b} + K_{bb'}(k)$ is positive definite. Then a linear functional ω_0 on D and a positive definite sesquilinear form s on $\tilde{D}$ are given by the relations:

$$\omega_0(f) = \sum_b \tilde{\rho}_b \tilde{f}(b) \tag{3}$$

and

$$s(f_1,f_2) = \omega_0(f_1 f_2) + (2\pi)^{-3} \sum_{b,b'\in\Gamma} \int_B dk\, K_{bb'}(k) \times \tilde{f}_1(k-b)\tilde{f}_2(-k+b'). \tag{4}$$

In particular it corresponds a Gaussian state on $\mathcal{A}$ follows.

Proof: For a test function f the Fourier coefficients $\tilde{f}(b)$ are summable. By assumption the $\tilde{\rho}_b$ are bonded. Hence a linear functional ω_0 is defined by relation (3).

In order to prove that expression (4) is well defined we need to show that for a test function of $\tilde{D}$ the quantity $\|f\|_* = \Sigma_{b\in\Gamma} \sup_{k\in B} |\tilde{f}(k-b)|$ is bounded. It follows because the Fourier transform of a function which is infinitely differentiable with compact support tends to zero at infinity faster than any polynomial. One then has:

$$\sum_{b,b'\in\Gamma} \left| \int_B dk\, K_{bb'}(k) \tilde{f}_1(k-b)\tilde{f}_2(-k+b') \right| \leqslant M \|f_1\|_* \|f_2\|_*.$$

So one has absolute convergence of the summations in (2). From the relation

$$s(f_1,f_2) = (2\pi)^{-3} \sum_{bb'\in\Gamma} \int_B dk\, [\tilde{\rho}_{b'-b} + K_{bb'}(k)] \times \tilde{f}_1(k-b)\tilde{f}_2(-k+b')$$

the positivity of s follows.

Remark: Denote $\rho_0 = V(\Lambda_0)^{-1}$. Let $\tilde{h}(Q)$ be a locally integrable function. Assume that a bound M exists such that

$$\int_B |\tilde{h}(k+b)|\, dk \leqslant M,$$

independently of $b \in \Gamma$. Assume that $\rho_0 \tilde{h}(Q) > -1$ almost everywhere. Let $\tilde{\rho}_b = \delta_{b0}\rho_0$ and $K_{bb'}(k) = \delta_{bb'}\rho_0^2 \tilde{h}(k-b)$. Under these assumptions Prop. 6 becomes a particular case of Prop. 5.

Example: A harmonic crystal is described by the dynamical matrix $N(k)$. The wavevector k belongs to the Brillouin zone B. The inverse matrix is denoted $X(k)$. It is positive definite almost everywhere. The matrix elements $X_{\alpha\alpha'}(k)$ are integrable.

Denote

$$W_{\alpha\alpha'}(p) = kT\, V(B)^{-1} \int_B dk\, e^{ikAp} X_{\alpha\alpha'}(k).$$

Let

$$\tilde{\rho}_b = V(\Lambda_0)^{-1} \exp\left(-\tfrac{1}{2} \sum_{\alpha\alpha'} W_{\alpha\alpha'}(0) b_\alpha b_{\alpha'} \right)$$

and

$$K_{bb'}(k) = V(\Lambda_0)^{-1} \times \exp\left(-\tfrac{1}{2} \sum_{\alpha\alpha'} W_{\alpha\alpha'}(0)[(b+k)_\alpha \times (b+k)_{\alpha'}(b'+k)_\alpha(b'+k)_{\alpha'}] \right) \times \sum_{p\neq 0} e^{-ikAp} \left[\exp \sum_{\alpha\alpha'} W_{\alpha\alpha'}(p)(k+b)_\alpha (k+b')_{\alpha'} - \right.$$

It is shown in the Appendix that $\tilde{\rho}_{b'-b} + K_{bb'}(k)$ is positive-definite. In the approximation that multi-phonon processes of order n larger than a given number n_0 are neglected one can show that the conditions of Prop. 6 are satisfied.

ACKNOWLEDGMENTS

I am indebted to Professor A. Verbeure for helpful discussions on the subject. Financial support of the neutron scattering project of the Belgian National Science Foundation is acknowledged.

APPENDIX

Let $W_{\alpha\alpha'}(p)$, $\tilde{\rho}_b$, and $K_{bb'}(k)$ be as in the example of Sec. 5. We show that the matrix $K_{bb'}(k) + \tilde{\rho}_{b'-b}$ is positive definite and that a uniform bound M exists such that $\int_B dk\ |K_{bb'}(k)| \leqslant M$.

The expression

$$H_{bb'}(k) = V(B)^{-1} \sum_p e^{-ikAp} \times \left\{ \exp \sum_{\alpha\alpha'} W_{\alpha\alpha'}(p)(k+b)_\alpha (k+b')_{\alpha'} - 1 \right\}$$

is expanded in a Taylor series. One obtains

$$H_{bb'}(k) = V(B)^{-1} \sum_p e^{-ikAp} \sum_{n=1}^{\infty} (1/n!) \times \left(\sum_{\alpha\alpha'} W_{\alpha\alpha'}(p)(k+b)_\alpha (k+b')_{\alpha'} \right)^n.$$

Using the defining relation

$$W_{\alpha\alpha'}(p) = kT\ V(B)^{-1} \int_B dk\ e^{ikAp} X_{\alpha\alpha'}(k)$$

one obtains

$$H_{bb'}(k) = \sum_{n=1}^{\infty} (1/n!)(kT)^n V(B)^{-1-n} \times \int_B dk_1 \cdots \int_B dk_n \times \left(\sum_p e^{-ikA} e^{ik_1 Ap} \cdots e^{ik_n Ap} \right) \times Y_{bb'}(k_1) \cdots Y_{bb'}(k_n),$$

with

$$Y_{bb'}(k_i) = \sum_{\alpha\alpha'} X_{\alpha\alpha'}(k_i)(k+b)_\alpha (k+b')_{\alpha'}.$$

Hence

$$H_{bb'}(k) = \sum_{n=1}^{\infty} (1/n!)(kT)^n V(B)^{-n} \times \int_B dk_2 \cdots \int_B dk_n$$

$$\times Y_{bb'}(k_2)\cdots Y_{bb'}(k_n)Y_{bb'}\left(k-\sum_{i=2}^{n}k_i\right).$$

The terms in the foregoing expression can be interpreted as coming from n-phonon processes. It follows by use of Schur's theorem that each of the contributions is a positive definite matrix. From the relation

$$\begin{aligned}K_{bb'}(k)+\tilde{\rho}_{b'-b}=(2\pi)^3V(\Lambda_0)^{-2}&\\ \times\exp\Big(-\tfrac{1}{2}\sum_{\alpha\alpha'}W_{\alpha\alpha'}(0)(b+k)_\alpha(b+k)_{\alpha'}\Big)&\\ \times\exp\Big(-\tfrac{1}{2}\sum_{\alpha\alpha'}W_{\alpha\alpha'}(0)(b'+k)_{\alpha'}(b'+k)_{\alpha'}\Big)&\\ \times(H_{bb'}(k)+V(B)^{-1})&\end{aligned}$$

now follows the required positivity condition.

Let us now make the approximation that the series in the expression for $H_{bb'}(k)$ is terminated after a finite number of terms. From Fubini's theorem then follows that $H_{bb'}(k)$ is an integrable function on B. It is then straightforward to show that the conditions of Prop. 6 are fulfilled.

[1]D. Ruelle, *Statistical Mechanics* (Benjamin, New York/Amsterdam, 1969).
[2]J. V. Pulè and A. Verbeure, J. Math. Phys. **18**, 862 (1977).
[3]J. L. van Hemmen, "*Dynamics and ergodicity of the infinite harmonic crystal*," thesis (Groningen, 1976).
[4]I. I. Gihman and A. V. Skorohod, *The Theory of Stochastic Processes, I* (Springer, Berlin/Heidelberg/New York, 1974).
[5]L. Van Hove, Phys. Rev. **95**, 249 (1954).

Realization, extension, and classification of certain physically important groups and algebras

Nikos Salingaros
Physics Department, University of Massachusetts, Boston, Massachusetts 02125

(Received 10 October 1979; accepted for publication 7 December 1979)

An associative algebra of differential forms with division has been constructed. The algebra of forms in each different space provides a practical realization of the universal Clifford algebra of that space. A classification of all such algebras is given in terms of two distinct types of algebras N_k and S_k. The former include the dihedral, quaternion, and Majorana algebras; the latter include the complex, spinor, and Dirac algebras. The associative product expresses Hodge duality as multiplication by a basis element. This makes possible the realization of higher order algebras in a calculationally useful algebraic setting. The fact that the associative algebras, as well as the enveloped Lie algebras, are precisely those arising in physics suggests that this formalism may be a convenient setting for the formulation of basic physical laws.

PACS numbers: 02.10.Tq

1. INTRODUCTION

The search for an appropriate algebraic setting for the description of physical laws has been a central problem of mathematical physics. Clifford, following Hamilton's success with the quaternions $\mathbb{H}$, generalized them to any dimension in the universal Clifford algebras.[1] As an example of their applicability, he gave a specific realization of one of these algebras, the biquaternions $\mathbb{H} \oplus \mathbb{H}$, and proceeded to apply them to specific problems.[2] His untimely death prevented him from fully realizing this end, and the biquaternions were not used again to any extent.

Physicists developed the Spinor,[3] Majorana,[4] and Dirac[5] algebras as defined by specific matrix representations but only later Hestenes[6] gave a realization of the Spinor algebra that was consistent with Clifford's original conception. More recently, there have been some attempts to apply larger Clifford algebras to physics.[7-10]

A classification of universal Clifford algebras was given by Atiyah, Bott, and Shapiro,[11] without however providing a specific representation in each case which could be used for algebraic manipulations.

This paper gives precisely such a practical algebraic setting by using as a basis the differential forms of each space. In doing this, the larger algebras in the classification[11] are realized exactly, and turn out to include those associative algebras that traditionally appear in physics. In addition, the unfamiliar algebras may be useful in constructing physical models that exhibit symmetry in an intrinsic manner.[12]

The universal Clifford algebra is defined by the multiplication of vectors a and b whose components are expanded on a basis σ^i as $a = \Sigma\, a^i\sigma^i$ and $b = \Sigma\, b^i\sigma^i$. The product is defined with the quadratic scalar form $(a,b) = \Sigma\, g^{ij}a^i b^j$, as $ab + ba = 2(a,b)$, and $a^2 = (a,a)$. Here, g is the metric $g^{ij} = (\sigma^i,\sigma^j)$.

In contrast to this, our approach is to provide an explicit realization of the associative product (directly), which is here defined as the "vee" product avb.[13,14] This is given in terms of the exterior product $a \wedge b$ and the scalar product (a,b), and satisfies $avb + bva = 2(a,b)$ as a condition (and not as a definition). Furthermore, we extend the vee product to all higher rank forms such as $A = \frac{1}{2}\Sigma A^{ij}\sigma^i \wedge \sigma^j$, etc.

We next show that the Hodge duality[15] is expressible as the vee product with the volume element. This result considerably simplifies the manipulation of higher rank forms, and allows a classification of finite groups generated by the vee product among the differential forms in each space (Tables I–IV). These groups are in turn used to provide a representation and classifications of universal Clifford algebras (Table VI).

The Lie and Jordan algebras contained in each universal Clifford algebra (and therefore representable by this construction) are given in Sec. 6 (Ref. 16 and Table V). An explicit realization is described for the Lorentz–Minkowski–Clifford algebra $\mathbb{N}_4$[12,14] in spacetime $M^{1,3}$, by giving the commutation and anticommutation relations (16), (18), (20), and (21).

Of the several unfamiliar (to physicists) universal Clifford algebras, the smallest of these, the "dihedral Clifford algebra" in two dimensional spacetime $\mathbb{N}_1$, is examined. It is shown that $\mathbb{N}_1$ provides an example of a distinct associative algebra of the same dimension as the quarternions, which is not a division algebra. The formalism can be used to demonstrate that Clifford algebras are not in general division algebras.

The utility in actual manipulations follows from a sectional divisibility, namely the existence in each algebra of a unique inverse to each homogeneous form. One can indeed divide in the case of each antisymmetric tensor field.

In the Conclusion, the results obtained from an associative derivative are mentioned, as well as further extensions of the material presented in this paper.

2. ASSOCIATIVE MULTIPLICATION OF DIFFERENTIAL FORMS

Consider the n differential one forms $\sigma^\alpha = dx^\alpha$, $\alpha = 1,\ldots,n$ of an n-dimensional carrier space M^n. We can construct a set of 2^n basis p forms via the Cartan exterior product $\wedge$.[15] These are

 0022-2488/81/020226-07$1.00

$$\{1,\sigma^\alpha,\sigma^\alpha\wedge\sigma^\beta,\ldots\omega^n\},\quad \alpha,\beta=1,\ldots,n\,. \tag{1}$$

Here the volume element in n dimensions is labeled

$$\omega^n=\sigma^1\wedge\ldots\wedge\sigma^n\,. \tag{2}$$

The number of basis forms of each rank is given by a binomial coefficient.[15] In the carrier space M^n, define a metric as the scalar form $g^{\alpha\beta}$[14]:

Definition:

$$g^{\alpha\beta}=(\sigma^\alpha,\sigma^\beta)\,. \tag{3}$$

The carrier space can be characterized by this metric (which we assume for the purposes of this paper to be diagonal and nondegenerate) as $M^{p,q}$ when

$$g^{\alpha\alpha}=(+1,+1,\ldots,-1,-1,\cdots)=g(p,q)\,, \tag{4}$$

p plus signs, q minus signs, and $p+q=n$.

We define a multiplication v between a basis r form and a basis $(s-r)$ form as the sum of permutations of basis forms in Eq. (1):

Definition:

$$\begin{aligned}(\sigma^{\lambda_1}\wedge\ldots\wedge\sigma^{\lambda_r})\mathrm{v}(\sigma^{\lambda_{r+1}}\wedge\ldots\wedge\sigma^{\lambda_s})&\\ =\frac{1}{r!(s-r)!}\sum_{\pi_r}\sum_{\pi_{s-r}}\sum_{k=0}^{r}(-1)^{\pi_r}(-1)^{\pi_{s-r}}&\\ \times g^{\lambda_1\lambda_2}\ldots g^{\lambda_{2k-1}\lambda_{2k}}\sigma^{\lambda_{2k+1}}\wedge\ldots\wedge\sigma^{\lambda_s}.&\end{aligned} \tag{5}$$

Here $(-1)^{\pi_r}$ is the sign of the permutation

$$\begin{pmatrix}1 & 2\ldots r\\ \lambda_1 & \lambda_2\ldots\lambda_r\end{pmatrix}. \tag{6}$$

It is easy to show that Eq. (5) defines an *associative* product (this can be shown in general, and verified directly for simple cases; see Ref. 13).

The vee product in Eq. (5) can be used to define an associative product between antisymmetric tensor fields. These are real-component fields expanded on the differential form basis (1). For example, the vector fields a and b are described as

$$a=\sum_{\lambda_1=1}^{n}a^{\lambda_1}\sigma^{\lambda_1}\ \text{and}\ b=\sum_{\lambda_2=1}^{n}b^{\lambda_2}\sigma^{\lambda_2}\,.$$

The vee product between them is

$$\begin{aligned}a\mathrm{v}b&=\sum_{\lambda_1,\lambda_2=1}^{n}a^{\lambda_1}b^{\lambda_2}\sigma^{\lambda_1}\mathrm{v}\sigma^{\lambda_2}\\ &=\sum_{\lambda_1,\lambda_2=1}^{n}a^{\lambda_1}b^{\lambda_2}(g^{\lambda_1\lambda_2}+\sigma^{\lambda_1}\wedge\sigma^{\lambda_2})\\ &=(a,b)+a\wedge b\,.\end{aligned} \tag{7}$$

This simple case of vee multiplication of vectors serves to illustrate the fact that we can multiply *any* rank antisymmetric tensor fields in the associative vee multiplication. The actual manipulations of vectors in three dimensions are presented in Ref. 13, and are a generalization and extension of the usual vector algebra. The product of higher rank tensor fields contains as special cases the tensor, matrix, and exterior products. Manipulations of fields in Minkowski spacetime are described in Refs. 12 and 14.

In this paper, we omit the machinery for the algebraic manipulations of tensor fields, and concentrate instead on the intrinsic algebraic structure. (See however examples of inverses of tensor fields in Sec. 7.) It will be shown that this "associative algebra of forms" provides a realization of the universal Clifford algebra in each carrier space $M^{p,q}$. For discussions on Clifford algebras, see Refs. 6 to 19, inclusive.

3. DUALITY AND THE GROUP STRUCTURE

One of the most useful properties of the vee product is that the duality of Hodge[20] can be expressed algebraically. This leads to an ease of manipulations which makes this formalism useful for calculations.

Theorem: The dual of any form is the vee product of that form with the volume elements

$$\underset{r}{*}f^{(p)}=\pm\omega^r\mathrm{v}f^{(p)}\,. \tag{8}$$

Here, $f^{(p)}$ is a p form and $\overset{*}{r}$ the duality in r dimensions ($r\leqslant n$). The sign depends on p,r, the dimension n, and the signature t of the metric. The expression (8) can be used to express the embedding of duality in different dimensions as follows:

$$\underset{r}{*}f=\left(\underset{r-1}{*}f\right)\mathrm{v}\sigma^r. \tag{9}$$

In practice, we can label the higher rank forms by their duals. Then, the identities (8) and (9) allow their manipulation in a simple algebraic manner. For example, in three dimensions we can label the twoforms as the duals of one forms:

$$\underset{3}{*}\sigma^1=\sigma^2\wedge\sigma^3,\ \underset{3}{*}\sigma^2=\sigma^3\wedge\sigma^1,\ \underset{3}{*}\sigma^3=\sigma^1\wedge\sigma^2\,. \tag{10}$$

The vee product is defined in Eqs. (4) and (6) with arbitrary indices, because the product between two elements of each algebra (which are antisymmetric tensor fields) involves the summation over these indices [see Eq. (7)]. For specific indices, however, the product of two basis forms of any rank will be a single basis form of some other rank. Hence, the vee multiplication associates a unique form in Eq. (2) to each pair of forms in Eq. (1). This property defines a finite group under the vee multiplication. In the following section, these groups are identified, and make possible a classification of each universal Clifford algebra in terms of its underlying group structure.

Using the identification (10), we can give the multiplication (Table I) of all basis forms in three dimensions. These define a group of forms $G^{0,3}$ under the vee multiplication.

We display only the positive forms in the table. The finite group includes the negatives of these forms as well,

TABLE I. The vee product between all basis forms in three dimensions.

v	σ^1	σ^2	σ^3	$*\sigma^1$	$*\sigma^2$	$*\sigma^3$	ω^3
σ^1	-1	$*\sigma^3$	$-*\sigma^2$	ω^3	σ^3	$-\sigma^2$	$-*\sigma^1$
σ^2	$-*\sigma^3$	-1	$*\sigma^1$	$-\sigma^3$	ω^3	σ^1	$-*\sigma^2$
σ^3	$*\sigma^2$	$-*\sigma^1$	-1	σ^2	$-\sigma^1$	ω^3	$-*\sigma^3$
$*\sigma^1$	ω^3	σ^3	$-\sigma^2$	-1	$*\sigma^3$	$-*\sigma^2$	$-\sigma^1$
$*\sigma^2$	$-\sigma^3$	ω^3	σ^1	$-*\sigma^3$	-1	$*\sigma^1$	$-\sigma^2$
$*\sigma^3$	σ^2	$-\sigma^1$	ω^3	$*\sigma^2$	$-*\sigma^1$	-1	$-\sigma^3$
ω^3	$-*\sigma^1$	$-*\sigma^2$	$-*\sigma^3$	$-\sigma^1$	$-\sigma^2$	$-\sigma^3$	1

TABLE II. Classification of all spaces $M^{p,q}$ by their dimension $n = p + q$, and their metric signature (p,q).

$n = 0$					(0,0)					
$n = 1$				(1,0)		(0,1)				
$n = 2$			(2,0)		(1,1)		(0,2)			
$n = 3$		(3,0)		(2,1)		(1,2)		(0,3)		
$n = 4$	(4,0)		(3,1)		(2,2)		(1,3)		(0,4)	
⋮					⋮					

which is but a trivial extension of Table I. This specific example has the Euclidean metric $g^{ii} = (-1, -1, -1)$. Note that even though the overall minus sign is usually a trivial convention, the related Clifford algebra is distinct.[11] Specifically, the Clifford algebra for $g^{ii} = (-1, -1, -1)$ is $\mathbb{H} \oplus \mathbb{H}$, Clifford's biquaternions, while the corresponding Clifford algebra for $g^{ii} = (+1, +1, +1)$ is $\mathbb{S}$, the Pauli spinor algebra (see Table VI).

To analyze the group structure, we evaluate the order of each element in the group in the context of the vee product.

An element σ of the group is of order two, if $\sigma \vee \sigma = (\sigma)^2 = 1$, and an element is of order four, if $\sigma \vee \sigma = (\sigma)^2 = -1$, i.e., $(\sigma)^4 = 1$. The unit 1 is of order one. The order of the group is given by the order of the elements; when there are k elements of order 2 and m elements of order 4, this is denoted as $(1, k, m)$. The order of the group is just $1 + k + m$. From Table I, it is easy to see that the group of forms $G^{0,3}$ is of order (1, 3, 12), and is therefore isomorphic to the group of Biquaternions of Clifford $B = Q_4 \otimes Z_2$. We have used the fact that small finite nonabelian groups are isomorphic if the group order is the same. (This is however not the case for arbitrarily large groups.[21] In general, one has to identify the subgroup structure in detail.)

Following this example, it is easy to see that every set of basis forms (1) defines a finite group under the vee multiplication. These groups are labeled by the signature of the metric of the carrier space; in general, a set of one forms satisfying the metric (4) will in this manner give rise to a group of forms $G^{p,q}$.

The collection of all such groups is now examined, and in each case the isomorphism to a known finite group is given whenever possible.

4. CLASSIFICATION OF A CLASS OF FINITE GROUPS

To each space $M^{p,q}$ of dimension $n = p + q$ and metric signature (p,q) [see Eq. (4)] there corresponds a finite group of dimension 2^{n+1} labeled $G^{p,q}$. We can classify all such groups by displaying them in a triangular array, where the position (p,q) corresponds to the space $M^{p,q}$ with metric (4) (Table II).

Our procedure following this is to explicitly calculate the order $(1, k, m)$ of each of these groups, and use this as the basis for studying their structure.

It is straightforward to evaluate the order of each element of a given group, using the vee multiplication. One uses the fact that the total number of basis forms of a particular rank is given by a binomial coefficient.[15] (Details of these calculations are omitted from this paper.)

Since the square of any basis form in the vee multiplication is either $+1$ or -1, then no element of the group has order higher than 4. The order of the group can therefore always be denoted by $(1, k, m)$.

This procedure reveals isomorphisms of these groups among themselves, which include the known identities $G^{p,q} = G^{q+1,p-1}$, $G^{p,q+4} \otimes G^0 = G^{p,q} \otimes G^{0,4}$, as well as the periodicity of Bott $G^{p,q+8} \otimes G^0 = G^{p,q} \otimes G^{0,8}$.[11] Other relations between the groups appear new.

All these identities can be succinctly and usefully displayed by identifying the nonisomorphic finite groups as the members of two series of groups N_k and S_k, $k = 0,1,2,...$ The groups $G^{p,q}$ can then be listed in terms of the N_k and S_k notation in a manner that explicitly shows the above isomorphisms. In Table III, we have listed the finite groups up to order 512. The correspondence of the two notations employed for the groups is given by comparing an entry of Table II to the corresponding entry of Table III. For example, the group $G^{3,0}$ from the entry (3,0) of Table II is the group lebeled S_1 in Table III. The usefulness of this labeling will be apparent when we specify the isomorphisms of these groups with the commonly known finite groups.

The matter of identifying the groups N_k and S_k with known finite groups is a straightforward one, and employs simple group-theoretical techniques.[21] An example was given previously in the case of $G^{0,3} = N_2 \otimes N_0 = Q_4 \otimes Z_2$. Here we merely list the results in Table IV, along with some useful identities. The notation employed is the following:

Z_n are the cyclic groups of order n. D_n are the dihedral groups. Z_4 is also the complex group. $Z_2 \otimes Z_2$ is the Gauss–Klein veergruppe, which is isomorphic to D_2. Q_4 is the quaternion group of Hamilton. $S = Z_4 \otimes Z_2 \otimes Z_2$ is the spinor group, which was first discovered by Hamilton as the "complex quaternions." The group M is the group of all possible

TABLE III. Classification of finite groups corresponding to the spaces $M^{p,q}$.

$n = 0$									N_0											
1								$N_0 \otimes N_0$		S_0										
2							N_1		N_1		N_2									
3						S_1		$N_1 \otimes N_0$		S_1		$N_2 \otimes N_0$								
4					N_4		N_3		N_3		N_4		N_4							
5				$N_4 \otimes N_0$		S_2		$N_3 \otimes N_0$		S_2		$N_4 \otimes N_0$		S_2						
6			N_5		N_5		N_6		N_6		N_5		N_5		N_6					
7		S_3		$N_5 \otimes N_0$		S_3		$N_6 \otimes N_0$		S_3		$N_5 \otimes N_0$		S_3		$N_6 \otimes N_0$				
8	N_7		N_8		N_8		N_7		N_7		N_8		N_8		N_7		N_7			
⋮									⋮											

TABLE IV. Isomorphisms with known finite groups.

n		Order	Elements
0	$N_0 = Z_2$	2	(1,1,0)
1	$N_0 \otimes N_0 = Z_2 \otimes Z_2 = D_2$	4	(1,3,0)
	$S_0 = Z_4$	4	(1,1,2)
2	$N_1 = D_4$	8	(1,5,2)
	$N_2 = Q_4$	8	(1,1,6)
3	$S_1 = Z_4 \otimes (Z_2)^2$	16	(1,7,8)
	$N_1 \otimes N_0 = D_4 \otimes Z_2$	16	(1,11,4)
	$N_2 \otimes N_0 = Q_4 \otimes Z_2 = B$	16	(1,3,12)
4	$N_3 = M$	32	(1,19,12)
	N_4	32	(1,11,20)
5	$S_2 = Z_4 \otimes (Z_2)^4 = \mathscr{D}$	64	(1,31,32)
	$N_4 \otimes N_0 = N_1 \otimes N_2$	64	(1,23,40)
	$N_3 \otimes N_0 = N_1 \otimes N_1 = N_2 \otimes N_2$	64	(1,39,24)
6	N_5	128	(1,55,72)
	N_6	128	(1,71,56)
7	$S_3 = Z_4 \otimes (Z_2)^6 = \mathscr{D} \otimes Z_4$	256	(1,127,128)
	$N_5 \otimes N_0 = N_3 \otimes N_2 = N_4 \otimes N_1$	256	(1,111,144)
	$N_6 \otimes N_0 = N_3 \otimes N_1 = N_4 \otimes N_2$	256	(1,143,112)
8	N_7	512	(1,271,240)
	N_8	512	(1,239,272)

combinations of the 4×4 real Majorana matrices,[4] which define a group of order 32. The 16 familiar combinations of the real gamma matrices along with their negatives are a representation of M.[22] In distinction, the Dirac group $\mathscr{D}$ defined by the 4×4 complex Dirac matrices is of order 64, since one has to include the complex elements separately. We note that Dirac's original matrices do not define $\mathscr{D}$, since he added the metric of the quadratic form (1,3) by hand.[5]

The relationship between the Majorana and Dirac groups can be understood as a "complexification" in the context of the vee multiplication. i.e., $\mathscr{D} = M \oplus iM$. This is the reason why M is one half as large as $\mathscr{D}$. The complex unit i corresponds in this case to the five-dimensional volume element ω^5, which commutes with all elements of $\mathscr{D}$ (and M) and has square $\omega^5 \vee \omega^5 = -1$. This follows as a special case of some general theorems, which are given in the following section (see also Ref. 14).

The groups N_k are seen to be the building blocks of the classification. They are the groups arising from defining the vee structure on the differential forms of an *even* dimensional space. The first four are well known as follows: $N_0 = Z_2$; $N_1 = D_4$; $N_2 = Q_4$; and $N_3 = M$.

The groups S_k appear only when the dimension of the underlying space is odd, and may be called "generalized spinor groups." They have the intrinsically complex structure $S_k = Z_4 \otimes (Z_2)^{2k}$. The first three are well known as $S_0 = Z_4$; $S_1 = S$; and $S_2 = \mathscr{D}$.

The series N_k and S_k contain many of the finite groups that have traditionally appeared in physics. This is an indication of the intrinsic physical interest of this classification. The group structure in turn determines the algebraic structure, which will be in Sec. 7. In the following section we give a few useful theorems on the group structure.

Remark: The labeling of the groups S_k as generalized spinor groups reflects the historical development of generalizing the complex numbers to the spinor algebra,[3] then subsequently generalizing the spinor algebra to the Dirac algebra.[5]

5. THE GROUPS AS CENTRAL EXTENSIONS OF CYCLIC GROUPS

It is instructive to give two theorems first derived by Clifford[1] on properties of the volume element ω^n in each universal Clifford algebra. These can then be applied to evaluate the center of each group of forms.

Theorem: The volume element ω^n commutes with all elements σ of G^n for n = odd, and anticommutes for n = even:

$$\omega^n \vee \sigma = (-1)^{n+1} \sigma \vee \omega^n . \tag{11}$$

(This theorem is particularly useful in practical manipulations.[12,13])

Theorem: The square of the volume element in the vee multiplication is

$$(\omega^n)^2 = \omega^n \vee \omega^n = (-1)^{n(n-1)/2} \det g = \pm 1 . \tag{12}$$

The proofs are direct;[1,16,19] g is the metric tensor (4).

These two theorems can now be used to give the center of G^n.

Theorem: The center of the group of forms G^n is isomorphic to the finite group Z_2, when n = even; Z_4, when n = odd and $(\omega^n)^2 = -1$; or $Z_2 \otimes Z_2 = D_2$, when n = odd and $(\omega^n)^2 = +1$. (13)

The proof is as follows: We see that the center of each group of forms is generated by the elements $\{1, -1\}$ for n = even and $\{1, -1, \omega^n, -\omega^n\}$ for n = odd (see also Ref. 16). These sets define finite groups which are isomorphic to Z_2, Z_4, or $Z_2 \otimes Z_2$, by using Eq. (12). This theorem in turn leads to a key result in the group structure.

Theorem: The factor group G modulo, the center of G, is the Abelian group $(Z_2)^n = Z_2 \otimes \ldots \otimes Z_2$ (n times). The different cases are

$$n = \text{even}: \quad G^n/Z_2 = (Z_2)^n , \tag{14a}$$

$$n = \text{odd}: \quad (\omega^n)^2 = -1 , \quad G^n/Z_4 = (Z_2)^{n-1} , \tag{14b}$$

$$n = \text{odd}: \quad (\omega^n)^2 = 1 , \quad G^n/Z_2 \otimes Z_2 = (Z_2)^{n-1} . \tag{14c}$$

This theorem, presented here without proof (although it is very easy to check for the first few cases), can be rewritten as a corollary.[23]

Corollary: The groups of forms G are the central extensions of Z_2, Z_4, and $Z_2 \otimes Z_2$ by the Abelian group $(Z_2)^n$. (15)

It seems, therefore, that those groups appearing in this classification scheme, which includes many of the groups used in physics, fall into a rather special class of finite groups having the above property.

6. REPRESENTATION OF A CLASS OF LIE AND JORDAN ALGEBRAS

The groups of forms G^n can be used to represent certain Lie and Jordan algebras, by using a bracket operation defined with the vee product. We define commutators and anticommutators as follows:

Definitions:

$$[\sigma^\alpha,\sigma^\beta] = \sigma^\alpha \mathrm{v} \sigma^\beta - \sigma^\beta \mathrm{v} \sigma^\alpha , \tag{16a}$$

$$\{\sigma^\alpha,\sigma^\beta\} = \sigma^\alpha \mathrm{v} \sigma^\beta + \sigma^\beta \mathrm{v} \sigma^\alpha . \tag{16b}$$

The set of forms (1) with either of these bracket products is closed, and is isomorphic to some Lie or Jordan algebra, respectively. It is a straightforward matter to evaluate the specific Lie and Jordan algebras, so the details are omitted. The first few cases are known, and the larger examples can at least be verified by evaluating the dimension.

The Lie algebras obtained in this manner are given in the compact case of each manifold M^n by the following theorem:

Theorem: The Lie algebra corresponding to the commutator vee structure on the space M^n is

$$\mathrm{SL}(2^{n/2};\mathbb{R}), \quad \text{for } n = \text{even},$$
$$\mathrm{SL}(2^{(n-1)/2};\mathbb{C}), \quad \text{for } n = \text{odd}. \tag{17}$$

What is of interest is the ability to use the explicit representation provided by the vee product between all the differential forms in order to evaluate the Killing–Cartan metric form contained in each Lie algebra. This is obtained by identifying the orthogonal group SO(*r*,*s*) which is covered by each Lie group in Eq. (17). This is also rather straightforward once the representation is known. For example, the Lorentz–Minkowski–Clifford group N_4 in spacetime $M^{1,3}$ contains the orthogonal conformal group SO(1,5).

In Table V, we classify the orthogonal Lie groups SO(*r*,*s*) contained in each group of forms by giving the Killing–Cartan form (*r*,*s*). We note that the table contains the rotation, Lorentz, and conformal Lie groups.

It is to be emphasized that the commutator of the group of forms gives a *specific* representation of the particular Lie group. This representation is precisely the fundamental representation, which can be verified by computing the Cartan–Casimir invariant.

For example, the realization of the SO(3) subgroup by the three generators J^{12}, J^{23}, J^{13} given below satisfies $j(j+1) = -(J^{12})^2 - (J^{23})^2 - (J^{13})^2 = 3/4$, and is therefore a spin one half representation. Similarly, the Lorentz group SO(1,3) is realized by six generators in a $(j,j) = (\frac{1}{2},\frac{1}{2})$ representation. This feature is indicative of the very specific nature of this algebraic scheme. The consequences of this particular characteristic to physics will be examined separately.

As a specific example, we describe the Lie algebra generated from the Lorentz–Minkowski spacetime $M^{1,3}$ (see Ref. 12). The relevant group of forms is $G^{1,3} = N_4$ of order 32. We can label the 15 nonscalar basis forms in a 6×6 antisymmetric array, utilizing the duality in both three and four dimensions in the labeling [see Eq. (10)], as well as the volume elements (2):

$$J^{\alpha\beta} = \tfrac{1}{2}\begin{pmatrix} 0 & \overset{*}{3}\sigma^3 & -\overset{*}{3}\sigma^2 & \sigma^1\wedge\sigma^4 & \overset{*}{4}\sigma^1 & \sigma^1 \\ & 0 & \overset{*}{3}\sigma^1 & \sigma^2\wedge\sigma^4 & \overset{*}{4}\sigma^2 & \sigma^2 \\ & & 0 & \sigma^3\wedge\sigma^4 & \overset{*}{4}\sigma^3 & \sigma^3 \\ & & & 0 & \omega^3 & \sigma^4 \\ & & & & 0 & -\omega^4 \\ & & & & & 0 \end{pmatrix} . \tag{18}$$

The indices α,β run from 1,...,6. The Killing–Cartan form in this case is easily verified from the techniques of Sec. 2 to be

$$\bar{g}_{\alpha\alpha} = (-1,-1,-1,+1,-1,-1). \tag{19}$$

The commutation relations are given in canonical form as

$$[J^{\alpha\beta},J^{\gamma\delta}] = \bar{g}^{\beta\gamma}J^{\alpha\delta} - \bar{g}^{\beta\delta}J^{\alpha\gamma} - \bar{g}^{\alpha\gamma}J^{\beta\delta} + \bar{g}^{\alpha\delta}J^{\beta\gamma} . \tag{20}$$

Even though it is known that the Clifford bivectors of the space $M^{1,3}$ (which are here represented by the basis forms labeled $J^{\alpha\beta}$ with $\alpha,\beta = 1,2,3,4$) provide a representation of the Lorentz group SO(1,3);[24] the construction here presented gives a representation of the much larger conformal group SO(1,5).

Furthermore, the identification of each orthogonal Lie algebra of Table V as the one maximally contained in the universal Clifford algebra of the corresponding space M provides a set of examples to the Poincare–Birkhoff–Witt theorem.[25] This important theorem endows each Lie algebra with a unique universal associative enveloping algebra. Representations of these algebras are not known in general, but we can provide the universal associative enveloping algebra for each of the entries of Table V. The fact that these orthogonal groups include those that are of physical importance makes this restricted result of some interest to physical applications.

In a similar vein to the preceeding analysis of the Lie algebraic structure induced by defining the commutator product (16a), we can analyze the Jordan structure[26] induced by the anticommutator product (16b). As in the preceeding case, we give as an example the anticommunication relations of the 15 nonscalar bases (18) corresponding to the space $M^{1,3}$. The Killing–Cartan form $\bar{g}^{\alpha\beta}$ [Eq. (19)] is the same as in the commutation relations (20):

$$\{J^{\alpha\beta},J^{\gamma\delta}\} = \tfrac{1}{2}\bar{g}^{\alpha\delta}\bar{g}^{\beta\gamma} - \tfrac{1}{2}\bar{g}^{\alpha\gamma}\bar{g}^{\beta\delta} - \epsilon^{\alpha\beta\gamma\delta\epsilon\xi}J^{\epsilon\xi} . \tag{21}$$

Here $\epsilon^{\alpha\beta\gamma\delta\epsilon\xi}$ is the entirely antisymmetric symbol of Levi-Civita in six indices, and $\epsilon^{123456} = 1$.

The anticommutation relations for the basis forms labeled $J^{\alpha 6}$, $\alpha = 1,2,3,4$ are trivially known, since they define the Jordan algebra as

$$\{\sigma^\mu,\sigma^\nu\} = 2g^{\mu\nu} . \tag{22}$$

TABLE V. Killing–Cartan form of Lie groups SO(*r*,*s*) realized by each group of forms $G^{p,q}$.

$n = 0$ (Abelian)					SO(1)					
1 (Abelian)				(2,0)		(0,2)				
2			(2,1)		(1,2)		(0,3)			
3		(3,1)		(2,2)		(1,3)		(0,4)		
4	(5,1)		(3,3)		(3,3)		(1,5)		(1,5)	
⋮					⋮					

TABLE VI. Classification of universal Clifford algebras.

n																	
$n=0$									$\mathbb{R}$								
1								$\mathbb{R}\oplus\mathbb{R}$		$\mathbb{C}$							
2							$\mathbb{N}_1$		$\mathbb{N}_1$		$\mathbb{H}$						
3						$\mathbb{S}$		$\mathbb{N}_1\oplus\mathbb{N}_1$		$\mathbb{S}$		$\mathbb{H}\oplus\mathbb{H}$					
4					$\mathbb{N}_4$		$\mathbb{M}$		$\mathbb{M}$		$\mathbb{N}_4$		$\mathbb{N}_4$				
5				$\mathbb{N}_4\oplus\mathbb{N}_4$		$\mathbb{D}$		$\mathbb{M}\oplus\mathbb{M}$		$\mathbb{D}$		$\mathbb{N}_4\oplus\mathbb{N}_4$		$\mathbb{D}$			
6			$\mathbb{N}_5$		$\mathbb{N}_5$		$\mathbb{N}_6$		$\mathbb{N}_6$		$\mathbb{N}_5$		$\mathbb{N}_5$		$\mathbb{N}_6$		
7		$\mathbb{S}_3$		$\mathbb{N}_5\oplus\mathbb{N}_5$		$\mathbb{S}_3$		$\mathbb{N}_6\oplus\mathbb{N}_6$		$\mathbb{S}_3$		$\mathbb{N}_5\oplus\mathbb{N}_5$		$\mathbb{S}_3$		$\mathbb{N}_6\oplus\mathbb{N}_6$	
8	$\mathbb{N}_7$		$\mathbb{N}_8$		$\mathbb{N}_8$		$\mathbb{N}_7$		$\mathbb{N}_7$		$\mathbb{N}_8$		$\mathbb{N}_8$		$\mathbb{N}_7$		$\mathbb{N}_7$
⋮									⋮								

[Here $g^{\mu\nu}$ is the Lorentz–Poincare metric $(-1,-1,-1,+1)$ and not Eq. (19).] The inclusion of the other 11 basis forms in the expression (21), however, enlarges the anticommutation structure, and displays an interesting "skew" character.

Since Eq. (22) is also the defining relation of the Clifford algebra by the extension theorem,[11,18,19] the Atiyah–Bott–Shapiro construction[11] gives the universal associative enveloping algebra of each Jordan algebra. This is precisely the universal Clifford algebra, and provides the analog to the Poincare–Birkhoff–Witt theorem for the case of Jordan algebras.

7. CLASSIFICATION OF ASSOCIATIVE ALGEBRAS

The present formalism allows an identification and classification of universal Clifford algebras in terms of their underlying group structure. This analysis places the associative algebras used in physics in this scheme, as well as providing a practical realization for use in actual applications.[13,14] The first three algebras are the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$, and the quaternions $\mathbb{H}$. In addition, it is possible to include the Spinor algebra $\mathbb{S}$, the Majorana algebra $\mathbb{M}$, and the Dirac algebra $\mathbb{D}$.

Each group of forms shown in Table III defines a realization of a universal Clifford algebra which is displayed in Table VI. We have used the group terminology of Sec. 4 to label the algebras $\mathbb{N}_k$ and $\mathbb{S}_k$, $k = 0,1,2,\ldots$, thereby singling out the nonisomorphic algebras by the notation. The isomorphisms to familiar associative algebras are the following:

$$\mathbb{N}_0 = \mathbb{R},\quad \mathbb{N}_2 = \mathbb{H},\quad \mathbb{N}_3 = \mathbb{M};$$

$$\mathbb{S}_0 = \mathbb{C},\quad \mathbb{S}_1 = \mathbb{S},\quad \mathbb{S}_2 = \mathbb{D}.$$

The remaining algebras are not well known in physics. (References 8–10 and 27 are among the few to examine these larger Clifford algebras with regard to physical applications.)

A comparison of Table VI with the classification of universal Clifford algebras given in Atiyah, Bott, and Shapiro[11] will illustrate the differences resulting from this approach. The major advantage of this formalism, however, is in the relative ease with which practical manipulations can be realized. This is succinctly expressed by the following theorem, which is a key point of the construction.

Theorem: Homogeneous forms are invertible in each algebra. (23)

Thus, any antisymmetric tensor field (which is a linear combination of basis forms of equal rank) will have a unique inverse in any given algebra of Table VI. The inverse is of course defined in terms of the vee multiplication. Again, examples from the Lorentz–Minkowski–Clifford algebra $\mathbb{N}_4$ [12,14] in spacetime $M^{1,3}$ are presented. For a vector field

$$a = \sum a^\mu \sigma^\mu, \quad \mu = 1,2,3,4\,. \tag{24a}$$

Then

$$a^{-1} = a / \} a_\mu a^\mu\,. \tag{24b}$$

For a rank two tensor field, the duality in three dimensions can be used for elegance (see Ref. 13)

$$F = \tfrac{1}{2} \sum F^{\mu\nu} \sigma^\mu \wedge \sigma^\nu = \mathbf{E} \wedge \sigma^4 - \overset{*}{3}\mathbf{B}, \quad \mu,\nu = 1,2,3,4\,, \tag{25a}$$

with

$$E^i = F^{i4}, \text{ and } B^i = \tfrac{1}{2} \sum \epsilon^{ijk} F^{jk}, \quad i,j = 1,2,3\,.$$

Then

$$F^{-1} = \frac{(k_1\mathbf{E} + k_2\mathbf{B}) \wedge \sigma^4 - \overset{*}{3}(k_1\mathbf{B} - k_2\mathbf{E})}{k_1^2 + k_2^2}, \tag{25b}$$

where $k_1 = (E,E) - (B,B)$, and $k_2 = 2(E,B)$, with $(E,E) = \Sigma E^i E^i$.

(The constants k_1, k_2, and $k_1^2 + k_2^2$ are recognized to be the invariants of the Lorentz and duality transformations of the electromagnetic field. This is analyzed in detail elsewhere,[12] and serves to make the connection with physics.) The identity $F \vee F = k_1 - k_2\omega$ can be used to verify the inverse of the field F [Eq. (25b)].

Clearly, one must have the full framework of the construction in each particular case in order to use the algebra for practical manipulations. Details for the algebras $\mathbb{N}_4$ [12,14] and $\mathbb{S}$ [13] have been given elsewhere. Note that a related but distinct description for $\mathbb{S}$ was given by Hestenes.[6]

The other algebras of Table VI may be of value in certain physical problems, and will be presented in detail in the appropriate physical context in forthcoming publications. We comment here on the appearance of the associative algebra $\mathbb{N}_1$, which we call the "dihedral Clifford algebra in a two-

dimensional space–time" (from the corresponding dihedral group in Table IV).

The algebra, although associative, and of the same size as the quaternions $\mathbb{H}$, does not appear to have been used in the past. A reason for this may be that $\mathbb{N}_1$ is neither normed, nor a division algebra. (This is easily verified directly.) We note that, in general, the associative algebras of Table VI are not division algebras, as this would violate the celebrated Hurwitz theorem.[28] Thus, the inverse of a linear combination of *all* the basis forms (i.e., different rank tensor fields) may not exist. The only division algebras in the classification VI are the usual ones: $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$.

The algebra $\mathbb{N}_1$ is therefore an example of a relatively small associative algebra which is not a division algebra, yet has possible physical significance because it arises from two-dimensional spacetime. Physical applications of this interesting and curious algebra will be presented separately.

8. CONCLUSIONS AND EXTENSIONS

This paper has illustrated how universal Clifford algebras can be generated by an associative multiplication of differential forms in a computationally useful setting. The introduction of Hodge duality, in particular, is a key tool in the algebraic manipulations. The simplicity of the reduction from the universal Clifford algebra to the Jordan and Lie algebras in each case illustrated how the topological enveloping is related to the algebraic structure via the commutator and anticommutator vee products.

It is hoped that this formalism may provide a novel and useful framework to describe both internal and external symmetries in physics. The appearance of many physically significant groups and algebras in this classification is indicative of the inherent applicability of the scheme.

For most of the results on group structure, the algebraic basis can be considered in an abstract manner. However, the identification of the differential one forms as the basis is indeed crucial, as will be demonstrated in a subsequent communication. There, an associative derivative in each space will give an expression of generalized holomorphy as well as a set of field equations. The equations in the case of Lorentz–Minkowski spacetime are precisely the Maxwell equations (with electric sources).

ACKNOWLEDGMENTS

I am indebted to Dr. K.C. Hannabuss of the Mathematical Institute, Oxford, for his clarifying remarks, and in particular for those on the Abelian structure. I would like to thank Professor R. Penrose and Professor E.A. Paschos for their hospitality at Oxford and Dortmund, respectively, as well as for their encouraging comments during the summer of 1979.

[1] W.K. Clifford, "On the Classification of Geometric Algebras," paper XL III in *Mathematical papers of W.K. Clifford*, edited by R. Tucker (MacMillan, London, 1882).
[2] W.K. Clifford, Am. J. Math. **1**, 350 (1878).
[3] W. Pauli, Z. Phys. **43**, 601 (1927).
[4] E. Majorana, Nuovo Cimento **5**, 171 (1937).
[5] P.A.M. Dirac, Proc. R. Soc. (London) Ser. A **117**, 610 (1928).
[6] D. Hestenes, *Spacetime Algebra* (Gordon and Breach, New York, 1966).
[7] E. Caianiello, Nuovo Cimento Supp. **14**, 177 (1959); see also Ref. 17.
[8] S.A. Basri and L.P. Horwitz, Phys. Rev. D **11**, 572 (1975); see also Refs. 10 and 27.
[9] C.N. Ktorides, J. Math. Phys. **16**, 2123 (1975).
[10] L.P. Horwitz and L.C. Biedenharn, J. Math. Phys. **20**,269 (1979).
[11] M. Atiyah, R. Bott, and A. Shapiro, Topology **3**, 3 (1964); see also Refs. 18 and 19.
[12] N. Salingaros and M. Dresden, "An associative algebra of Tensor Fields" to appear in Proc. R. Soc. (London).
[13] N. Salingaros, Hadronic J. **3**, 339 (1979).
[14] N. Salingaros and M. Dresden, Phys. Rev. Lett. **43**, 1 (1979).
[15] H. Flanders, *Differential Forms* (Academic, New York, 1963).
[16] C. Chevalley *Theory of Lie Groups* (Princeton U.P., Princeton, 1946).
[17] E. Caianiello, *Combinatorics and Renormalizations in Quantum Field Theory* (Benjamin, Reading, Mass., 1973).
[18] I. Porteous, *Topological Geometry* (Van Nostrand Reinhold, London, 1969).
[19] M. Karoubi, *K-Theory* (Springer, Berlin, 1979).
[20] W.V.D. Hodge, *The Theory and Application of Harmonic Integrals* (Cambridge University, Cambridge, England, 1952).
[21] B. Higman, *Applied Group-Theoretic and Matrix Methods* (Oxford University, London, 1955).
[22] W. Pauli, Ann. Inst. Henri Poincaré **6**, 137 (1936).
[23] K.C. Hannabuss (private communication).
[24] H. Boerner, *Representations of Groups* (North-Holland, Amsterdam, 1967).
[25] N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).
[26] P. Jordan, Z. Phys. **80**, 285 (1933).
[27] L.P. Horwitz, D. Sepunaru, and L.C. Biedenharn, Tel-Aviv preprint TAUP 752–79.
[28] C.W. Curtis, in *Studies in Modern Algebra*, edited by A.A. Albert (Prentice-Hall, Englewood Cliffs, New Jersey 1963).

The Racah algebra for groups with time reversal symmetry

J. D. Newmarch[a] and R. M. Golding
Department of Physical Chemistry, University of New South Wales, P.O. Box 1, Kensington, New South Wales 2033, Australia

(Received 6 February 1979; accepted for publication 11 September 1979)

The elements of the Racah algebra for a general compact group with time reversal symmetry are developed. As time reversal is antilinear it is not possible to treat these groups by representation theory but we may instead use Wigner's theory of corepresentations. The results we obtain often parallel those for linear groups but there are some important divergences. We illustrate these with the grey double point groups.

PACS numbers: 02.20. + b, 03.65.Fd

1. INTRODUCTION

Over the years an extensive range of mathematical tools generally subsumed under the heading of "Racah algebra" has been developed to handle the problem of many electrons moving in a spherically symmetric potential.[1-4] These methods may be interpreted group theoretically as follows: (a) The set of vectors $\{|jm\rangle : -j \leqslant m \leqslant j\}$ form a basis for an irreducible representation (IR) of SU(2). (b) If we couple two electrons, each with a sharp angular momentum, we are group theoretically forming the direct product of two IR's. This may be transformed by a unitary matrix into a direct sum of IR's. The elements $\langle j_1 j_2 jm | j_1 j_2 m_1 m_2\rangle$ are termed coupling coefficients and give

$$|j_1 m_1\rangle |j_2 m_2\rangle = \sum_{jm} \langle j_1 j_2 jm | j_1 j_2 m_1 m_2\rangle |jm\rangle.$$

This may be inverted as

$$|(j_1 j_2) jm\rangle = \sum_{m_1 m_2} \langle j_1 j_2 m_1 m_2 | j_1 j_2 jm\rangle |j_1 m_1\rangle |j_2 m_2\rangle,$$

where

$$\langle j_1 j_2 m_1 m_2 | j_1 j_2 jm\rangle = \langle j_1 j_2 jm | j_1 j_2 m_1 m_2\rangle^*.$$

(c) A high symmetry coefficient—the $\overline{V}$ coefficient of Fano and Racah[5]—may be constructed from this

$$\overline{V}\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = (-1)^{2j_3 + j_3 + m_3} [j_3]^{-1/2} \times \langle j_1 j_2 m_1 m_2 | j_1 j_2 j_3 - m_3\rangle,$$

where $[j_3] = 2j_3 + 1$, which at most suffers from a sign change under permutations and time reversal. It is related to the equally symmetric $3j$ symbol of Wigner[6] by

$$\overline{V}\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = (-1)^{j_1 + j_2 + j_3}\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}.$$

An equivalent approach in representation theory relates this to the reduction of the triple direct product to the identity representation:

$$\overline{V}\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = (-1)^{2j_3} \langle j_1 j_2 j_3 m_1 m_2 m_3 | 00\rangle.$$

(d) If we couple three angular momenta together, the coupling may be carried out in different ways. The recoupling coefficient gives the transformation between these, and again there is a high symmetry coefficient related to this—which is independent of the azimuthal quantum numbers

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix} = \sum_{\text{all } m} (-1)^l \overline{V}\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \overline{V}\begin{pmatrix} j_1 & j_5 & j_6 \\ -m_1 & m_5 & -m_6 \end{pmatrix} \times \overline{V}\begin{pmatrix} j_4 & j_2 & j_6 \\ -m_4 & -m_2 & m_6 \end{pmatrix} \overline{V}\begin{pmatrix} j_4 & j_5 & j_3 \\ m_4 & -m_5 & -m_3 \end{pmatrix},$$

with

$$l = \sum_{i=1}^{6} (j_i - m_i).$$

This is called either a $\overline{W}$ or $6j$ coeffieicnt[5,6] and is the first nontrivial invariant under any change of basis. This may be extended to the invariant recoupling coefficients of four or more particles.[7-11] (e) If an operator satisfies the same commutation relations as a spherical harmonic Y_{lm}, it forms the mth component of a spherical tensor of rank l. We may separate out the radial and spherical parts, and by the Wigner–Eckart theorem[12,13]

$$\langle j_3 m_3 | T(k_1 q_1) | j_2 m_2\rangle = (-1)^{j_3 - m_3} \overline{V}\begin{pmatrix} j_3 & j_2 & k_1 \\ -m_3 & m_2 & q_1 \end{pmatrix} \times \langle j_3 || T(k_1) || j_2\rangle,$$

where the reduced matrix element $\langle j_3 || T(k_1) || j_2\rangle$ represents the radial contribution. By considering various products of tensor operators in coupled schemes, relations between the different reduced elements which arise may be obtained.[7]

The development of the Racah algebra for SU(2) has been largely based on the fact that this is a Lie group and so we may use the Lie algebra su(2) or the special function properties of the IR's to simplify the calculation of the $\overline{V}$ and related coefficients. Although this approach may be utilized for other Lie groups, our interest lies in another direction, namely, in the finite and compact groups. Although work is currently being done on developing a Lie algebra-type approach to finite groups,[14,15] at present we are restricted to group or group algebra methods.

Two approaches may be made to the problem. The first is to deal with the group in its own right and to develop the

[a]Present address: Department of Physics, Universiti Pertanian Malaysia, Serdang, Selangor, Malaysia.

 0022-2488/81/020233-12$1.00

algebra using only the properties of the group. Koster[16] has proven the Wigner–Eckart theorem using this method, and a large number of authors have determined the properties and values of $\overline{V}$, $3j$ (or $3jm$) and higher symbols for the point groups and arbitrary compact groups.[17–21] The other aproach is to treat the group as a subgroup of another group for which a Racah algebra has already been developed and use descent in symmetry.[22–32] Various authors have used a mixture of the two techniques.[33–40] Racah's lemma[4] shows that any of these methods give equivalent results. The papers by Derome and Sharp[17] and Derome[18] have given the conditions for the $3jm$ symbols to be multiplied by a simple phase factor only under permutations and complex conjugation, and symbols with this property may be found for all the double point groups.

Many of these authors have used the operation of complex conjugation K or of time reversal θ to relate the pairs of coefficients

$$\begin{pmatrix} \Gamma_1 & \Gamma_2 & \Gamma_3 \\ \alpha_1 & \alpha_2 & \alpha_3 \end{pmatrix} \text{ and } \begin{pmatrix} \Gamma_1{}^* & \Gamma_2{}^* & \Gamma_3{}^* \\ \alpha_1{}^* & \alpha_2{}^* & \alpha_3{}^* \end{pmatrix},$$

where the set of $|\Gamma\alpha\rangle$ form a basis for the representation D^{Γ} and the set of $|\Gamma^*\alpha^*\rangle$ for the representation D^{Γ^*}. These two operations are not completely equivalent for, whereas $K^2 = I$, $\theta^2 = \pm I$ according to whether we have an even or odd number of fermions. These operators are antilinear:

$$\theta(\alpha|f\rangle + \beta|g\rangle) = \alpha^*\theta|f\rangle + \beta^*\theta|g\rangle.$$

For nonmagnetic materials, or for paramagnetic or diamagnetic ions, time reversal is a symmetry operator of the system. Kramers[41] first demonstrated the importance of this operator with the Kramers degeneracy, but as it is antilinear it cannot be treated by the theory of representations but can only be dealt with by the theory of corepresentations developed by Wigner.[42] A large number of papers have been published concerning the corepresentations of linear/antilinear groups[43–48] but development of a Racah algebra has been slow compared to the linear groups. Methods for finding, and tables of, coupling coefficients have been published[44–46] and the Wigner–Eckart theorem has been proven,[48] but to our knowledge that is all.

In this paper we develop the Racah algebra for compact groups with time reversal symmetry. We take time reversal θ as a commuting operator of the group[49] and $\theta^2 = I$ for bosons or an even number of fermions, and $\theta^2 = -I$ for an odd number of fermions. These two properties allow us to simplify the development from that which would be needed for a general linear/antilinear group. In practice, we usually start from a compact group of linear operators H and extend it to G by adding in the antilinear operators $a = \theta u$, where u is linear. Since linear times antilinear is antilinear, and antilinear times antilinear is linear, the coset group G/H is isomorphic to C_2. As H is compact, from the homeomorphism $f_\theta(u) = f(\theta u)$, $G - H$ is compact and hence G is compact. We shall see later that all irreducible corepresentations of G are obtained from irreducible representations of H, and so the compactness of H guarantees a complete set of corepresentations of G. Finally, we shall often need to integrate over the group. G and H both possess invariant integrals, so we may write

$$\int_G f(x)\,dx = \int_H f(u)\,du + \int_{G-H} f(a)\,da.$$

By simple substitution we have in particular

$$|H| = \int_H 1\,du = \int_{G-H} 1\,da = \tfrac{1}{2}\int_G 1\,dx = \tfrac{1}{2}|G|.$$

For finite groups these integrals reduce to sums.

2. A COVARIANT NOTATION

Throughout the fields of representation and corepresentation theory there exists a wide variety of notations for the labeling of representation or corepresentation matrices and for the nj symbols, coupling coefficients, etc. Use of a naïve notation often hides difficulties, but a more sophisticated notation calls for more care in construction and manipulation. The papers by Derome and Sharp[17] and Derome[18] discuss the properties of the $3jm$ symbols of a compact linear group in a very general manner, and for this they used a covariant notation. In this paper we discuss the $3jm$ symbols[50] for a grey compact group, and to ensure that we do not "gloss over" difficulties we adopt a covariant notation here. The one we use here differs slightly from that of Derome and Sharp and is from spinor and rotor calculus.[51,52]

Let S be a complex vector space with basis $\{e_{(m)} : m = 1,2,\cdots\}$ and S^* be the conjugate space with basis $\{\bar{e}_{(\dot{m})} : \dot{m} = 1,2,\cdots\}$. If for a particular space S' the conjugate space S'^* contains elements not in S', we extend S' by these elements so that we may always take $S = S^*$. We may transform to new bases $\{e_{(m')} : m' = 1,2,\cdots\}$ and $\{\bar{e}_{(\dot{m}')} : \dot{m}' = 1,2,\cdots\}$ by

$$e_{(m')} = e_{(m)} P^m{}_{m'} \tag{2.1a}$$

and

$$\bar{e}_{\dot{m}'} = \bar{e}_{(\dot{m})} P^{\dot{m}}{}_{\dot{m}'}. \tag{2.1b}$$

The inverse transformations are

$$e_{(m)} = e_{(m')} P^{m'}{}_{m} \tag{2.1c}$$

and

$$\bar{e}_{(\dot{m})} = \bar{e}_{(\dot{m}')} P^{\dot{m}'}{}_{\dot{m}}, \tag{2.1d}$$

with the convention $(P^m{}_{m'})^{-1} = P^{m'}{}_{m}$. Throughout primes will always indicate a transformed basis. We let the transpose of $P^m{}_{m'}$ be $P_{m'}{}^m$ and the complex conjugate be $P^{\dot{m}}{}_{\dot{m}'}$. It follows that

$$P^m{}_{m'} P_{\dot{n}'}{}^{\dot{n}} e_{(m)} \bar{e}^T_{(\dot{n})} = e_{(m')} \bar{e}^T_{(\dot{n}')}. \tag{2.2}$$

Now $e_{(m)}\bar{e}^T_{(\dot{n})}$ is the metric in the old basis, and on the right-hand side we have the metric in the new basis. In group theory we usually consider only unitary transformations, i.e., those which preserve orthonormal bases. Thus unitary transforms are those P which satisfy

$$P^m{}_{m'} P_{\dot{n}'}{}^{\dot{n}} \delta_{\dot{n}m} = \delta_{\dot{n}'m'}, \tag{2.3a}$$

where

$$\delta_{m\dot{n}} = \delta_{m'\dot{n}'} = \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & \\ & & 1 & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & \cdot \end{pmatrix}.$$

They also satisfy

$$P^{m}{}_{m'} P_{\dot{n}'}{}^{\dot{n}} \delta^{m'\dot{n}'} = \delta^{m\dot{n}}. \tag{2.3b}$$

We now define the matrix of a linear operator $u : S \to S$ by

$$ue_{(m)} = e_{(n)} j(u)^{n}{}_{m}. \tag{2.4}$$

Note that we do not have primes here as we are not changing the basis. Under a change of basis,

$$j(u)^{n'}{}_{m'} = P^{n'}{}_{n} j(u)^{n}{}_{m} P^{m}{}_{m'}. \tag{2.5}$$

The matrix of an antilinear operator $a : S \to S^*$ is given by

$$ae_{(m)} = e_{(n)} j(a)^{n}{}_{\dot{m}}, \tag{2.6}$$

where we now have one dotted and one undotted variable. We may sum over covariant and contravariant indices; we make the further convention that we may only sum over a pair of dotted indices or a pair of undotted indices. With this the transformation of an antilinear operator is

$$j(a)^{n'}{}_{\dot{m}'} = P^{n'}{}_{n} j(a)^{n}{}_{\dot{m}} P^{\dot{m}}{}_{\dot{m}'} \tag{2.7}$$

and the multiplication rules are

$$j(u_1u_2)^{m}{}_{n} = j(u_1)^{m}{}_{m_1} j(u_2)^{m_1}{}_{n}, \tag{2.8a}$$

$$j(ua)^{m}{}_{\dot{n}} = j(u)^{m}{}_{m_1} j(a)^{m_1}{}_{\dot{n}}, \tag{2.8b}$$

$$j(au)^{m}{}_{\dot{n}} = j(a)^{m}{}_{\dot{m}_1} j(u)^{\dot{m}_1}{}_{\dot{n}}, \tag{2.8c}$$

$$j(a_1a_2)^{m}{}_{n} = j(a_1)^{m}{}_{\dot{m}_1} j(a_2)^{\dot{m}_1}{}_{n}. \tag{2.8d}$$

We see how this notation automatically keeps track of complex conjugates. For example, in ordinary matrix notation Eq. (2.8c) reduces to

$$j(au) = j(a)j(u)^*,$$

which is the multiplication rule for such operators.[42]

3. IRREDUCIBLE COREPRESENTATIONS

Given a group G with a subgroup H of index 2, a corepresentation of G is a set of matrices satisfying equations (2.8) with $u \in H$ and $a \in G - H$. Usually, we have a complex vector space which carries a representation of H, in which case we may calculate these matrices by Eqs. (2.4) and (2.6) but this is not necessary for the formal theory.

A corepresentation may be reduced if we can find a unitary matrix which block diagonalizes the corepresentation. As Maschke's theorem still holds,[43] we can reduce each corepresentation to a direct sum of irreducible corepresentations (ICR's). However, only a weak form of Schur's lemma holds which gives the following rather unpleasant orthogonality relations[47]:

$$\int_H j_1(u)^{m_1}{}_{n_1} j_2(u)^{m_2}{}_{\dot{n}_2}\, du = 0, \tag{3.1a}$$

$$\int_{G-H} j_1(a)^{m_1}{}_{\dot{n}_2} j_2(a)^{\dot{m}_2}{}_{n_1}\, da = 0, \tag{3.1b}$$

if $j_1 \neq j_2$ and

$$\int_H j(u)^{m_1}{}_{n_1} j(u)^{\dot{m}_2}{}_{\dot{n}_2}\, du + \int_{G-H} j(a)^{m_1}{}_{\dot{n}_2} j(a)^{\dot{m}_2}{}_{n_1}\, da = \frac{|G|}{[j]} \delta^{m_1\dot{m}_2} \delta_{n_1\dot{n}_2}, \tag{3.1c}$$

with $[j]$ equal to the dimension of j. Note how n_1 and $\dot{n}_2$ are interchanged in the last integral. One small point is necessary: in Eq. (3.1c) we have $\delta^{m_1\dot{m}_2}$, where both m_1 and $\dot{m}_2$ are inner indices. Since we are not multiplying matrices in the normal manner, this does not contradict the matrix convention of summing over an inner and outer index. Setting $m_1 = n_1$, $\dot{m}_2 = \dot{n}_2$ gives

$$\int_H |\chi_j(u)|^2\, du + \int_{G-H} \chi_j(a^2)\, da = |G|, \tag{3.2}$$

where $\chi_j(u)$ $[\chi_j(a^2)]$ is the character of $j(u)$ $[j(a^2)]$.

By restricting the operations of G to the linear subgroup H, each ICR j of G subduces to a representation k of H. We may classify each ICR of G by the representation of H in the following manner[43]:

Type (a): k is irreducible and equivalent to $\bar{k}$, where $\bar{k}(u) = k(a^{-1}ua)^*$ for arbitrary fixed a. The character test is

$$\int_H |\chi_j(u)|^2\, du = \int_{G-H} \chi_j(a^2)\, da = |G|/2. \tag{3.3}$$

Type (b): k reduces to $k_1 \oplus \bar{k}_1$ with $k_1 \equiv \bar{k}_1$. The character test is

$$\int_H |\chi_j(u)|^2\, du = 2|G|. \tag{3.4}$$

Type (c) k reduces to $k_1 \oplus \bar{k}_1$ with $k_1 \not\equiv \bar{k}_1$. Here

$$\int_H |\chi_j(u)|^2\, du = |G|. \tag{3.5}$$

For type (a) we may use the orthogonality relations for linear groups to simplify Eq. (3.1c) to

$$\int_H j(u)^{m_1}{}_{n_1} j(u)^{\dot{m}_2}{}_{\dot{n}_2}\, du = \int_{G-H} j(a)^{m_1}{}_{\dot{n}_2} j(a)^{\dot{m}_2}{}_{n_1}\, da = \frac{|G|}{2[j]} \delta^{m_1\dot{m}_2} \delta_{n_1\dot{n}_2}. \tag{3.6}$$

We shall use this frequently for $j = \mathbf{0}$, the identity ICR.

For the grey groups time reversal θ commutes with all elements of G[49] so that $\bar{k} = k^*$ and equivalence (nonequivalence) is the equivalence (nonequivalence) of complex conjugate represention. Thus, we may make a further classification according to the following: k equivalent to k^* and to a real representation (the first kind); k equivalent to k^* but not to a real representation (the second kind); and k not equivalent to k^* (the third kind). Further, θ^2 is either the identity or the negative identity, and we have this classification for the grey groups[49]:

$$\theta^2 = I:$$

(A) k is irreducible and of the first kind:

$$\int_H |\chi_j(u)|^2\, du = \int_H \chi_j(u^2)\, du = |G|/2, \tag{3.7}$$

(B) k reduces to $k_1 \oplus k_1$ with k_1 of the second kind:

$$\int_H |\chi_j(u)|^2\, du = 2|G| \text{ and } \int_H \chi_j(u^2)\, du = -|G|, \tag{3.8}$$

(C) k reduces to $k_1 \oplus k_1{}^*$ with k_1 of the third kind:

$$\int_H |\chi_j(u)|^2 du = |G| \text{ and } \int_H \chi_j(u^2) du = 0; \tag{3.9}$$

$\theta^2 = -I$:

(D) k is irreducible and of the second kind:

$$\int_H |\chi_j(u)|^2 du = |G|/2 \text{ and } \int_H \chi_j(u^2) du = -|G|/2, \tag{3.10}$$

(E) k reduces to $k_1 \oplus k_1$ with k_1 of the first kind:

$$\int_H |\chi_j(u)|^2 du = 2|G| \text{ and } \int_H \chi_j(u^2) du = |G|, \tag{3.11}$$

(F) k reduces to $k_1 \oplus k_1^*$ with k_1 of the third kind:

$$\int_H |\chi_j(u)|^2 du = |G| \text{ and } \int_H \chi_j(u^2) du = 0. \tag{3.12}$$

Conversely, if we start from the IR's of H, the ICR's of G are constructed exactly according to the above scheme. Thus, time reversal only fails to increase the degeneracy for types (A) and (D). We observe that these ICR's are equivalent to their complex conjugates from the character tests of Rudra.[45] Each ICR matrix may be chosen to be unitary in the normal matrix sense[43] which becomes in our notation

$$j(u)^{m_1}{}_{m_2} j(u)_{n_2}{}^{n_1} \delta_{n_1 m_1} = \delta_{n_2 m_2} \tag{3.13a}$$

and

$$j(a)^{m_1}{}_{m_2} j(a)_{n_2}{}^{n_1} \delta_{n_1 m_1} = \delta_{n_2 m_2}. \tag{3.13b}$$

TABLE I. Clebsch–Gordan coefficients d^3_{12} for ICR's in terms of the Clebsch–Gordan coefficients c^3_{12} of the linear subgroup.

j_1	j_2	j_3	d^3_{12}
(a)	(a)	(a)	c^3_{12}
(a)	(a)	(b)	$\frac{1}{2}c^3_{12}$
(a)	(a)	(c)	c^1_{12}
(a)	(b)	(a)	$2c^3_{12}$
(a)	(b)	(b)	c^3_{12}
(a)	(b)	(c)	$2c^3_{12}$
(a)	(c)	(a)	$c^3_{12} + c^3_{12*}$
(a)	(c)	(b)	$\frac{1}{2}c^3_{12} + \frac{1}{2}c^3_{12*}$
(a)	(c)	(c)	$c^3_{12} + c^3_{12*}$
(b)	(b)	(a)	$4c^3_{12}$
(b)	(b)	(b)	$2c^3_{12}$
(b)	(b)	(c)	$4c^3_{12}$
(b)	(c)	(a)	$2c^3_{12} + 2c^3_{12*}$
(b)	(c)	(b)	$c^3_{12} + c^3_{12*}$
(b)	(c)	(c)	$2c^3_{12} + 2c^3_{12*}$
(c)	(c)	(a)	$c^3_{12} + c^3_{12*} + c^3_{1*2} + c^3_{1*2*}$
(c)	(c)	(b)	$\frac{1}{2}c^3_{12} + \frac{1}{2}c^3_{12*} + \frac{1}{2}c^3_{1*2} + \frac{1}{2}c^3_{1*2*}$
(c)	(c)	(c)	$c^3_{12} + c^3_{12*} + c^3_{1*2} + c^3_{1*2*}$

TABLE II. The minimum nonzero multiplicity of the identity ICR in double and triple products. Multiplicity free couplings of IR's give the following multiplicities for the ICR's. If the IR coupling is not multiplicity free, Table I should be used.

j_1	j_2	j_3	Multiplicity
(a)	(a)	...	1
(b)	(b)	...	4
(c)	(c)	...	2
(a)	(a)	(a)	1
(a)	(a)	(b)	2
(a)	(a)	(c)	2
(a)	(b)	(b)	4
(a)	(b)	(c)	4
(a)	(c)	(c)	2
(b)	(b)	(b)	8
(b)	(b)	(c)	8
(b)	(c)	(c)	4
(c)	(c)	(c)	2

Rudra's proof that basis vectors for an ICR of type (b) are nonorthogonal[45] contains an error, for although his projection operator maps a basis vector $|\phi_{i0}\rangle$ onto a basis vector $|\phi_i\rangle$, it does not annihilate all other basis vectors. We shall later construct suitable orthogonal bases which preserve unitarity.

Character tests may be used to reduce direct products of ICR's, but Eqs. (3.1) show that this will not be particularly easy. If

$$j_1 \otimes j_2 = \sum_{j_3} d^{j_3}_{j_1 j_2}\, j_3,$$

Bradley and Davies[43] have related this to the corresponding reduction of representations in H:

$$k_1 \otimes k_2 = \sum_{k_3} C^{k_3}_{k_1 k_2} k_3.$$

In Table I we reproduce the results as given in their Table V. Here we have a very important departure from representation theory. Whereas

$$C^{j_3}_{j_1 j_2} = C^{j_3 *}_{j_1 j_2 *} = C^{0}_{j_1 j_2 j_3 *}, \text{ etc.},$$

this is not in general true for ICR's. For instance, as we shall see later, in the grey tetrahedral group, $d^{T}_{TE} = 1$, $d^{E}_{TT} = 2$, and $d^{0}_{TTE} = 2$. If the multiplicity in the double product is not the same as the multiplicity in the triple product, then we clearly have a fundamental difference in the two reductions, which will influence our development of the 3*jm* symbols. The reduction of the triple product is independent of the order

$$d^{0}_{j_1 j_2 j_3} = d^{0}_{j_2 j_1 j_3} = d^{0}_{j_1 j_3 j_2}, \text{ etc.},$$

and we shall base our 3*jm* symbols on this reduction. The multiplicity of **0** in double and triple products is given in Table II.

4. ICR's OF THE DOUBLE GREY GROUPS

In this section we give the ICR's of each of the double grey groups and give a notation similar to Mulliken's[53] which will be used throughout. For the single group the grey group is the direct product of the point group and C_2 (Dim-

TABLE III. Reduction of the direct products of grey T^*.

	A	E	T	E'	U'
A	A				
E	E	$2A+E$			
T	T	$2T$	$A+E+2T$		
E'	E'	U'	$E'+U'$	$A+T$	
U'	U'	$2E'+U'$	$2E'+2U'$	$E+2T$	$2A+E+4T$

mock and Wheeler[54]) but this is not true for the double group, and we shall abuse notation by referring to a grey group by the usual double point group notation of Griffith.[55]

(a) SU(2), 0^*, and K^*: Each IR of these groups induces an ICR of types (A) or (D) and hence the Mulliken notation may be used to label the ICR's. The basis vectors of the IR's for 0^* and K^* given by Griffith,[55] Golding,[22,23] and McLellan[56] are also basis vectors of the ICR's.

(b) *The tetrahedral double grey group* T^*: The pair of representations Γ_2 and Γ_3 [55] are of the third kind and hence induce the ICR E of type (C) and similarly E'' and E''' induce the ICR U' of type (F). The reduction of the direct products is given in Table III and the multiplicity of A in the triple product in Table IV. It may be seen that the multiplicity problem for these groups is much worse than in groups without time reversal.

(c) *The dihedral double group* D^*_n *with n odd* ($n = 2m+1$): The single group D_n has two one-dimensional IR's A_1 and A_2 and m two-dimensional IR's $E_1 \ldots E_m$ all of the first kind. Thus, they all induce ICR's of type (A) and the generating matrices are

$$A_1(C_n) = A_1(C_2) = A_1(\theta) = 1,$$
$$A_2(C_n) = -A_2(C_2) = A_2(\theta) = 1,$$
$$E_j(C_n) = \begin{bmatrix} \exp(ij\phi) & 0 \\ 0 & \exp(-ij\phi) \end{bmatrix},$$
$$E_j(C_2) = E_j(\theta) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

where $\phi = 2\pi/n$.

TABLE IV. Multiplicity of A in triple products in T^*.

Product	Multiplicity
AAA	1
EEA	2
EEE	1
TTA	1
TTE	2
TTT	2
$E'E'A$	1
$E'E'T$	1
$U'E'E$	1
$U'E'T$	2
$U'U'A$	2
$U'U'E$	1
$U'U'T$	4

The double group D^*_n has in addition m two-dimensional IR's $E'_{1/2} \ldots E'_{m-1/2}$ which induce ICR's of type (D) and a pair of IR's A_1 and B_1 which induce an ICR E' of type (F). Typical generators are

$$E'_j(C_n) = \begin{bmatrix} \exp(ij\phi) & 0 \\ 0 & \exp(-ij\phi) \end{bmatrix}, E'_j(\theta) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$
$$E'(C_n) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, E'(C_2) = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix},$$

and

$$E'(\theta) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

The multiplicity of A_1 in each triple product is given in Table V.

(d) D^*_n *with n even* ($n = 2m$)

The single group D_n has four one-dimensional IR's A_1, A_2, B_1, and B_2 and $(m-1)$ two-dimensional IR's $E_1 \ldots E_{(m-1)}$. The double group has the additional m two-dimensional representations $E'_{1/2} \ldots E'_{\{m-(1/2)\}}$. Each of these induces an ICR. The generators for B_1 and B_2 are

$$-B_1(C_n) = B_1(C_2) = B_1(\theta) = 1,$$
$$-B_2(C_n) = -B_2(C_2) = B_2(\theta) = 1.$$

The generators for the other ICR's follow $D^*_{(2m+1)}$. There is no point giving a table of triple products as the multiplicity of A_1 is always one or zero.

(e) *The cyclic double grey group* C^*_n: The cyclic group C^*_n is isomorphic to C_{2n} [49] and hence has $2n$ one-dimensional representations $\Gamma_{m/2}$ with $-n < m \leqslant n$, where

$$\Gamma_{m/2}(C_n) = \exp(im\phi/2),$$

with

$$\phi = 2\pi/n.$$

$A = \Gamma_0$ induces an ICR of type (A). For $1 \leqslant m \leqslant n-1$ the character is complex and hence $\Gamma_{m/2}$ and $\Gamma_{-m/2}$ induce a two-dimensional ICR $E_{m/2}$ or $E'_{m/2}$. If n is even, $\Gamma_{n/2}$ induces an ICR A_2, but if n is odd, $\Gamma_{n/2}$ is a spin representation of the first kind and $\Gamma_{n/2} \oplus \Gamma_{n/2}$ induces the ICR E'. The generators are

$$E_{m/2}(C_n) = \begin{bmatrix} \exp(im\phi/2) & 0 \\ 0 & \exp(-im\phi/2) \end{bmatrix},$$

TABLE V. Multiplicity of A_1 in triple products in D^*_n ($n = 2m+1$).

Product	Multiplicity	Product	Multiplicity
$A_1A_1A_1$	1	$E'_jE'_jA_1$	1
$A_2A_2A_1$	1	$E'_jE'_jA_2$	1
$E_jE_jA_1$	1	$E'_jE'_kE_{j-k}$	1
$E_jE_jA_2$	1	$E'_jE'_kE_{j+k}$	1
$E_jE_kE_{j-k}$	1	$E'E'_jE_{n/2-j}$	2
$E_jE_kE_{j+k}$	1	$E'E'_jE_{n/2+j}$	2
		$E'E'A_1$	2
		$E'E'A_2$	2

$$E_{m/2}(\theta)=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix},$$

$$E'_{m/2}(C_n)=\begin{bmatrix}\exp(im\phi/2) & 0\\ 0 & \exp(-im\phi/2)\end{bmatrix},$$

$$E'_{m/2}(\theta)=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix},$$

and for n even

$$A_2(C_n)=-1,\quad A_2(\theta)=1,$$

whereas for n odd

$$E'(C_n)=\begin{bmatrix}-1 & 0\\ 0 & -1\end{bmatrix},\quad E'(\theta)=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}.$$

The triple direct product is given in Table VI.

5. THE 3*jm* SYMBOL

In representation theory it matters little whether we reduce $j_1\otimes j_2$ to j_3^* or $j_1\otimes j_2\otimes j_3$ to the identity **0**. However, we have already noted that for ICR's the two multiplicities are not necessarily the same so that there is a fundamental difference between the two reductions. We also noted that the multiplicity of j_3^* in $j_1\otimes j_2$ need not be the same as of j_2^* in $j_1\otimes j_3$, etc., meaning that any attempt to base the 3*jm*symbols on the double product will impose very restrictive properties on permutations of the ICR's However, the multiplicity of **0** in $j_1\otimes j_2\otimes j_3$ is independent of the order of coupling and we base the 3*jm* symbol on this reduction. We shall return to the problem of the direct product of two ICR's later.

We consider the unitary transformation which reduces the triple product $j_1\otimes j_2\otimes j_3$ and define the 3*jm* symbol to be the part of this matrix which reduces the product to the identity **0**:

$$j_1(u)^{m_1}{}_{n_1}j_2(u)^{m_2}{}_{n_2}j_3(u)^{m_3}{}_{n_3}$$
$$=(j_1j_2j_3)^{m_1m_2m_3}{}_{r_1}\delta^{r_1}{}_{r_2}(j_1j_2j_3)^{r_2}{}_{n_1n_2n_3}\oplus\cdots \tag{5.1}$$

and

$$j_1(a)^{m_1}{}_{\dot n_1}j_2(a)^{m_2}{}_{\dot n_2}j_3(a)^{m_3}{}_{\dot n_3}$$
$$=(j_1j_2j_3)^{m_1m_2m_3}{}_{r_1}\delta^{r_1}{}_{\dot r_2}(j_1j_2j_3)^{\dot r_2}{}_{\dot n_1\dot n_2\dot n_3}\oplus\cdots, \tag{5.2}$$

where $\delta^{r_1}{}_{r_2}$ and $\delta^{r_1}{}_{\dot r_2}$ are the linear and antilinear matrices, respectively, of the identity ICR and which are numerically equal to the identity matrix of dimension equal to the multiplicity M of **0** in the triple product. We have no primes in these definitions as we have all matrices in "standard" forms given by Eqs. (2.4) and (2.6).

By integrating over the linear and antilinear operators, respectively, we find from the orthogonality relation (3.6):

$$(j_1j_2j_3)^{m_1m_2m_3}{}_{r_1}(j_1j_2j_3)^{r_1}{}_{n_1n_2n_3}$$
$$=\frac{2}{|G|}\int_H j_1(u)^{m_1}{}_{n_1}j_2(u)^{m_2}{}_{n_2}j_3(u)^{m_3}{}_{n_3}\,du \tag{5.3}$$

and

$$(j_1j_2j_3)^{m_1m_2m_3}{}_{r_1}(j_1j_2j_3)^{r_1}{}_{\dot n_1\dot n_2\dot n_3}$$
$$=\frac{2}{|G|}\int_{G-H} j_1(a)^{m_1}{}_{\dot n_1}j_2(a)^{m_2}{}_{\dot n_2}j_3(a)^{m_3}{}_{\dot n_3}\,da, \tag{5.4}$$

where

$$(j_1j_2j_3)^{r}{}_{\dot n_1\dot n_2\dot n_3}=\delta^{r}{}_{\dot r}(j_1j_2j_3)^{\dot r}{}_{\dot n_1\dot n_2\dot n_3}. \tag{5.5}$$

The unitary equation (2.3a) gives

$$(j_1j_2j_3)^{m_1m_2m_3}{}_{r_1}(j_1j_2j_3)_{\dot r_2}{}^{\dot n_1\dot n_2\dot n_3}\delta_{(\dot n_1\dot n_2\dot n_3)(m_1m_2m_3)}=\delta_{\dot r_2 r_1}. \tag{5.6}$$

We cannot use Eq. (2.3b) as we have here only the first M columns of the reducing matrix, not all of them.

We now show that any reduction yields a complete set of 3*jm* symbols in the sense that any two reductions are related by a unitary transformation in the multiplicity label independent of m_1, m_2, and m_3. Suppose we have two such reductions. Then, from Eq. (5.3),

$$(j_1j_2j_3)^{m_1m_2m_3}{}_{r_1}(j_1j_2j_3)^{r_1}{}_{n_1n_2n_3}$$
$$=[j_1j_2j_3]^{m_1m_2m_3}{}_{r_2}[j_1j_2j_3]^{r_2}{}_{n_1n_2n_3}. \tag{5.7}$$

Square brackets are used for the 3*jm* symbols of the second reduction. Multiplying throughout (and summing) by $(j_1j_2j_3)_{\dot r_3}{}^{\dot s_1\dot s_2\dot s_3}\delta_{(\dot s_1\dot s_2\dot s_3)(m_1m_2m_3)}$ gives

$$(j_1j_2j_3)^{r_1}{}_{n_1n_2n_3}\delta_{r_3r_1}=[j_1j_2j_3]^{m_1m_2m_3}{}_{r_2}(j_1j_2j_3)_{\dot r_3}{}^{\dot s_1\dot s_2\dot s_3}$$
$$\times\delta_{(\dot s_1\dot s_2\dot s_3)(m_1m_2m_3)}[j_1j_2j_3]^{r_2}{}_{n_1n_2n_3}.$$

Setting

$$U^{r_1}{}_{r_2}=\delta^{r_1\dot r_3}[j_1j_2j_3]^{m_1m_2m_3}{}_{r_2}(j_1j_2j_3)_{\dot r_3}{}^{\dot s_1\dot s_2\dot s_3}$$
$$\times\delta_{(\dot s_1\dot s_2\dot s_3)(m_1m_2m_3)} \tag{5.8}$$

gives

$$(j_1j_2j_3)^{r_1}{}_{n_1n_2n_3}=U^{r_1}{}_{r_2}[j_1j_2j_3]^{r_2}{}_{n_1n_2n_3} \tag{5.9}$$

independent of n_1, n_2, and n_3 as required. We may show it satisfies Eq. (2.3a) by using the unitarity of the 3*jm* symbols, and hence $U^{r_1}{}_{r_2}$ is unitary. If we turn to the antilinear equations, we find

$$U^{\dot r_1}{}_{\dot r_2}=\delta^{\dot r_1}{}_{r_1}\delta^{r_2}{}_{\dot r_2}U^{r_1}{}_{r_2} \tag{5.10}$$

and since δ is numerically the identity matrix, U is numerically real.

6. THE WIGNER TENSOR

The Wigner tensor[6], otherwise known as the 1*j* symbol[17] or the 1*jm* symbol,[20] plays a very important role not only in relating a matrix to its conjugate in ambivalent groups, but also in relating a 3*jm* symbol to a coupling coefficient. We shall find such a tensor similarly useful in corepresentation theory. It is straightforward to show that time re-

TABLE VI. Multiplicity of A_1 in triple products in C_n^*.

Product	Multiplicity	Product	Multiplicity
$A_1A_1A_1$	1	$E_kE_1'E_{k+1}$	2
$A_2A_2A_1$	1	$E_kE_1'E_{k-1}$	2
$E_kE_1E_{k+1}$	2	$E_kE_k'A_1$	2
$E_kE_1E_{k-1}$	2	$E_kE_{n/2-k}'A_2$	2
$E_kE_kA_1$	2		
$E_kE_{n/2-k}A_2$	2	$E'E_k'E_{n/2+k}$	2
		$E'E_k'E_{n/2-k}$	2
		$E'E'A_1$	4

versal may serve for this tensor using the commutativity of θ:

$$j(u)^{m_1}{}_{m_2} = j(\theta u \theta^{-1})^{m_1}{}_{m_2}$$
$$= j(\theta)^{m_1}{}_{\dot{n}_1} j(u)^{\dot{n}_1}{}_{\dot{n}_2} j(\theta^{-1})^{\dot{n}_2}{}_{m_2}. \quad (6.1)$$

Now

$$\delta^{m_1}{}_{m_2} = j(\theta\theta^{-1})^{m_1}{}_{m_2} = j(\theta)^{m_1}{}_{\dot{n}} j(\theta^{-1})^{\dot{n}}{}_{m_2},$$

so

$$j(\theta^{-1})^{\dot{n}}{}_{m_2} = \delta^{m_1}{}_{m_2} j(\theta)^{-1\dot{n}}{}_{m_1} \quad (6.2)$$

and

$$j(u)^{m_1}{}_{m_2} = j(\theta)^{m_1}{}_{\dot{n}_1} j(u)^{\dot{n}_1}{}_{\dot{n}_2} j(\theta)^{-1\dot{n}_2}{}_{m_2}. \quad (6.3)$$

Similarly,

$$j(a)^{m_1}{}_{\dot{m}_2} = j(\theta)^{m_1}{}_{\dot{n}_1} j(a)^{\dot{n}_1}{}_{n_2} j(\theta)^{-1n_2}{}_{\dot{m}_2}. \quad (6.4)$$

$j(\theta)$ is an operator from S to S^* and $j(\theta^{-1})$ is also, so that in matrix terms $j(\theta)^m{}_{\dot{n}} \to j(\theta)$ and $j(\theta^{-1})^m{}_{\dot{n}} \to j(\theta^{-1})$. However, $j(\theta)^{-1}$ is an operator from S^* to S and hence in matrix terms $j(\theta)^{-1\dot{m}}{}_n \to j(\theta)^{-1}$. Equations (6.3) and (6.4) are therefore

$$j(u) = j(\theta) j(u)^* j(\theta)^{-1}$$

and

$$j(a) = j(\theta) j(a)^* j(\theta)^{-1*},$$

respectively, in the usual matrix formation. For later use we shall use the abbreviations

$$j(\theta)^m{}_{\dot{n}} = \begin{pmatrix} m & \\ & \dot{n} \end{pmatrix} \text{ and } j(\theta)^{-1\dot{n}}{}_m = \begin{bmatrix} \dot{n} & \\ & m \end{bmatrix}, \quad (6.5)$$

giving for Eqs. (6.3) and (6.4)

$$j(u)^{m_1}{}_{m_2} = \begin{pmatrix} m_1 & \\ & \dot{n}_1 \end{pmatrix} j(u)^{\dot{n}_1}{}_{\dot{n}_2} \begin{bmatrix} \dot{n}_2 & \\ & m_2 \end{bmatrix} \quad (6.6)$$

and

$$j(a)^{m_1}{}_{\dot{m}_2} = \begin{pmatrix} m_1 & \\ & \dot{n}_1 \end{pmatrix} j(a)^{\dot{n}_1}{}_{n_2} \begin{bmatrix} n_2 & \\ & \dot{m}_2 \end{bmatrix}, \quad (6.7)$$

respectively. From the property $j(\theta^2) = (-1)^n$ for n integral,

$$\begin{pmatrix} m_1 & \\ & \dot{n}_1 \end{pmatrix} \begin{pmatrix} \dot{n}_1 & \\ & m_2 \end{pmatrix} = (-1)^n \delta^{m_1}{}_{m_2} \quad (6.8)$$

and

$$\begin{bmatrix} \dot{n}_1 & \\ & m_1 \end{bmatrix} \begin{bmatrix} m_1 & \\ & \dot{n}_2 \end{bmatrix} = (-1)^n \delta^{\dot{n}_1}{}_{\dot{n}_2}. \quad (6.9)$$

The transformation from an operator to its conjugate is not necessarily unique (for example, in a commutative group any antilinear operator will serve) but because of the central role played by time reversal, we feel justified in reserving the notation for this operator only.

We shall now investigate to what extent $j(\theta)$ may be cast into a simple form. Ideally, this would be diagonal but we shall see that this is not always possible. It is not necessary to use the covariant notation for this, as we are not considering components in any detail and we simplify notation by considering an antilinear operator T satisfying $TT^* = \pm I$ and $T\dagger = T^{-1}$. The invariant eigenvector equation for an antilinear operator is

$$Tv = \lambda \bar{v},$$

with $T^* \bar{v} = \lambda^* v$.

We first deal with the case $TT^* = I$. Choose any v in S and let

$$\bar{w} = Tv.$$

Then $Tw = \bar{v}$. If $v = \lambda w$, we have an eigenvector. If $v \neq w$, set $u = v + w$. Trivially,

$$Tu = \bar{u},$$

giving an eigenvector. If V is the subspace generated by the eigenvector, we turn to the orthogonal subspace $V^\perp$ and repeat the process. Thus, we can find an orthogonal eigenvector basis and diagonalize T. The eigenvalues are ± 1 from $TT^* = I$.

When $TT^* = -I$ we have a different situation, for we can no longer produce eigenvectors as above. However, as before, we let

$$Tv = \bar{w},$$

for which

$$Tw = -\bar{v}.$$

Again, if $w = \lambda v$, we have an eigenvector. If not, we search for vectors which preserve the above form and are also orthogonal. By setting

$$v_1 = v + aw \text{ and } v_2 = w - av,$$

we have $Tv_1 = \bar{v}_2$ and $Tv_2 = -\bar{v}_1$. We may normalize v so that it has modulus one, and from the unitarity of T, w also has modulus one. If v_1 and v_2 are to be orthogonal,

$$o = \langle v_1 | v_2 \rangle = \bar{a}\langle w|w\rangle - a\langle v|v\rangle + \langle v|w\rangle - a\bar{a}\langle w|v\rangle$$

or

$$a - \bar{a} = \langle v|w\rangle - a\bar{a}\langle w|v\rangle^*$$

If we let $a\bar{a} = 1$, this is

$$\mathrm{Im}(a) = \mathrm{Im}(\langle v|w\rangle)$$

since v, w are unit vectors. This equation can certainly be solved for a since $a\bar{a} = 1$, and we have orthogonal vectors. We continue as before by considering the orthogonal subspaces. Thus, we may transform T to

$$\begin{pmatrix} \lambda_1 & & & & & & & \\ & \lambda_2 & & & & & & \\ & & \ddots & & & & & \\ & & & \ddots & & & & \\ & & & & 0 & 1 & & \\ & & & & -1 & 0 & & \\ & & & & & & 0 & 1 \\ & & & & & & -1 & 0 \\ & & & & & & & & \ddots \end{pmatrix}, \quad (6.10)$$

where from $T^*Tv = -v$ each eigenvalue λ is $\pm i$.

The forms given here are not always the same as those used in practice. For example, in SU(2) with j half-odd integral, the Fano–Racah standardization[5] gives $j(\theta)$ as

$$\begin{pmatrix} 0 & & & 1 \\ & & -1 & \\ & 1 & & \\ -1 & & & \\ & & & 0 \end{pmatrix},$$

which may be obtained from the above by a real orthogonal transformation. It is convenient to have $j(\theta)$ as a matrix with only one entry in each row and column, and we allow for such variations by taking

$$j(\theta)^{m}{}_{\dot{n}} = \begin{pmatrix} m & \\ & \dot{n} \end{pmatrix}$$

to be nonzero for only one $\dot{n}$ for each m, and similarly for

$$j(\theta)^{-1\,\dot{n}}{}_{m} = \begin{bmatrix} \dot{n} & \\ & m \end{bmatrix}.$$

7. THE COUPLING COEFFICIENT

We have already discussed some of the problems associated with the reduction of the direct product and we shall give an example which apparently creates more difficulties but which in fact shows the way out. First, we give the definition. The coupling coefficient $\langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{rm_3}$ is given by

$$j_1(u)^{m_1}{}_{n_1} j_2(u)^{m_2}{}_{n_2} = \sum_{j_3} \langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{rm_3} j_3(u)^{m_3}{}_{n_3} \langle j_1 j_2 | j_3 \rangle^{rn_3}{}_{n_1 n_2} \tag{7.1}$$

and

$$j_1(a)^{m_1}{}_{\dot{n}_1} j_2(a)^{m_2}{}_{\dot{n}_2} = \sum_{j_3} \langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{rm_3} \times \delta^{r}{}_{\dot{r}} j_3(a)^{m_3}{}_{\dot{n}_3} \langle j_1 j_2 | j_3 \rangle^{\dot{r}\dot{n}_3}{}_{\dot{n}_1 \dot{n}_2}, \tag{7.2}$$

where as before we have no primes.

Consider the coupling in grey $C_3^*, E' \otimes A_1 = E'$. Under the generators C_3 and θ we need to find a unitary matrix P such that

$$P \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} P^{-1} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

and

$$P \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} P^{-1*} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

It is elementary to verify that any of the four matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \text{ or } \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

will serve. The multiplicity is only one and we see that we have, in contradistinction to the $3jm$ symbols, reductions which are *not* related by a unitary transformation in the multiplicity label (which here would be multiplication by an scalar). The saving point is that we have here four independent reductions which is exactly the multiplicity of A_1 in $E' \otimes E' \otimes A_1$, and is also the number of independent matrices which commute with E'. Thus, the problems of the coupling coefficient $\langle j_1 j_2 | j_3 \rangle$ divide into three: (a) the number of matrices which commute with j_3, which gives the number of independent sets of coupling coefficients; (b) the determination of the relation between each set and the $3jm$ symbols; and (c) the relation between all sets of coupling coefficients and the $3jm$ symbols.

The first of these is a form of Schur's lemma for grey groups:

Lemma: If the multiplicity of $\mathbf{0}$ in $j \otimes j$ is M, there are M linearly independent nonsingular matrices P commuting with j:

$$Pj(u) = j(u)\,P \text{ and } Pj(a) = j(a)\,P^*.$$

This is proved by considering each type of ICR in turn and applying Schur's lemma for linear groups. For type (a) the ICR j of G subduces to an IR k of the linear subgroup H, and hence $P = cI$, giving one independent matrix I. For type (b) the ICR j is of the form[43]

$$j(u) = \begin{pmatrix} k(u) & 0 \\ 0 & k(u) \end{pmatrix} \text{ and}$$

$$j(a) = \begin{pmatrix} 0 & \pm k(a\theta^{-1})\,P \\ -k(a\theta^{-1})\,P & 0 \end{pmatrix}.$$

The four matrices

$$\begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}, \begin{pmatrix} iI & 0 \\ 0 & -iI \end{pmatrix}, \begin{pmatrix} 0 & iI \\ iI & 0 \end{pmatrix}, \text{ and } \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

commute with these, and there are no more linearly independent ones. For type (c),

$$j(u) = \begin{pmatrix} k(u) & 0 \\ 0 & k^*(u) \end{pmatrix} \text{ and } j(a) = \begin{pmatrix} 0 & k(a\theta^{-1}) \\ k^*(a\theta^{-1}) & 0 \end{pmatrix},$$

with two commuting matrices

$$\begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \text{ and } \begin{pmatrix} iI & 0 \\ 0 & -iI \end{pmatrix}.$$

Comparison with the multiplicities in Table II gives the result.

Now these matrices P may be written in coupling coefficient notation as

$$P^{m_1}{}_{m_2} = \langle j\mathbf{0} | j \rangle^{m_1 \mathbf{0}}{}_{m_2},$$

giving that, if $\langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{rm_3}$ is a coupling coefficient, so is

$$\langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{rm_3} \langle j_3 \mathbf{0} | j_3 \rangle^{m_3 \mathbf{0}}{}_{m_4}.$$

Thus, if the multiplicity of $\mathbf{0}$ in $j_3 \otimes j_3$ is M_1, we have M_1 sets of coupling coefficients. Each set obeys the usual orthogonality relations, but as the example shows, different sets need not be orthogonal when summing over m_1 and m_2.

The second problem, that of the relation between a set of coupling coefficients and the $3jm$ symbols, may be solved by use of the orthogonality relation (3.1c). From Eqs. (7.1) and (7.2),

$$\langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{r_1 m_3} j_3(u)^{m_3}{}_{n_3} j_3(u)^{m_4}{}_{\dot{n}_4} = j_1(u)^{m_1}{}_{n_1} j_2(u)^{m_2}{}_{n_2} j_3(u)^{m_4}{}_{n_4} \langle j_1 j_2 | j_3 \rangle^{n_1 n_2}{}_{r_1 n_3} \tag{7.3}$$

and

$$\langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{r_1 m_3} j_3(a)^{m_3}{}_{\dot{n}_4} j_3(a)^{\dot{m}_4}{}_{n_3} = j_1(a)^{m_1}{}_{\dot{n}_1} j_2(a)^{m_2}{}_{\dot{n}_2} j_3(a)^{\dot{m}_4}{}_{n_3} \langle j_1 j_2 | j_3 \rangle^{\dot{n}_1 \dot{n}_2}{}_{r_1 n_4}. \tag{7.4}$$

Replacing the conjugated matrices on the right-hand sides of these equations by

$$j_3(u)^{m_4}{}_{\dot{n}_4} = \begin{bmatrix} m_4 & \\ & m_5 \end{bmatrix} j_3(u)^{m_5}{}_{\dot{n}_5} \begin{pmatrix} n_5 & \\ & \dot{n}_4 \end{pmatrix}$$

and

$$j_3(a)^{\dot{m}_4}{}_{n_3} = \begin{bmatrix} \dot{m}_4 & \\ & m_5 \end{bmatrix} j_3(a)^{m_5}{}_{\dot{n}_5} \begin{pmatrix} \dot{n}_5 & \\ & n_3 \end{pmatrix},$$

respectively, we may integrate over G and G-H to give

$$\langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{r_1 m_3} \int_H j_3(u)^{m_3}{}_{n_3} j_3(u)^{\dot{m}_4}{}_{\dot{n}_4} du$$
$$= \frac{|G|}{2} \begin{bmatrix} \dot{m}_4 & \\ & m_5 \end{bmatrix} (j_1 j_2 j_3)^{m_1 m_2 m_5}{}_{r_2}$$
$$\times \begin{pmatrix} n_5 & \\ & \dot{n}_4 \end{pmatrix} \langle j_1 j_2 | j_3 \rangle^{n_1 n_2}{}_{r_1 n_3} (j_1 j_2 j_3)^{r_2}{}_{n_1 n_2 n_5} \tag{7.5}$$

and

$$\langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{r_1 m_3} \int_{G-H} j_3(a)^{m_3}{}_{\dot{n}_3} j_3(a)^{\dot{m}_4}{}_{n_4} da$$
$$= \frac{|G|}{2} \begin{bmatrix} \dot{m}_4 & \\ & m_5 \end{bmatrix} (j_1 j_2 j_3)^{m_1 m_2 m_5}{}_{r_2}$$
$$\times \begin{pmatrix} \dot{n}_5 & \\ & n_3 \end{pmatrix} \langle j_1 j_2 | j_3 \rangle^{\dot{n}_1 \dot{n}_2}{}_{r_1 \dot{n}_4} (j_1 j_2 j_3)^{r_2}{}_{\dot{n}_1 \dot{n}_2 \dot{n}_5}. \tag{7.6}$$

Adding and using orthogonality gives

$$\langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{r_1 m_3}$$
$$= [j_3]^{1/2} U^{r_2}{}_{r_1} \begin{bmatrix} \dot{m}_4 & \\ & m_5 \end{bmatrix} \delta_{m_3 \dot{m}_4} (j_1 j_2 j_3)^{m_1 m_2 m_5}{}_{r_2}, \tag{7.7}$$

with $U^{r_2}{}_{r_1}$ as the real rectangular tensor

$$U^{r_2}{}_{r_1} = \delta^{\dot{n}_4 n_3} \frac{[j_3]^{1/2}}{2}$$
$$\times \left\{ \begin{pmatrix} n_5 & \\ & \dot{n}_4 \end{pmatrix} \langle j_1 j_2 | j_3 \rangle^{n_1 n_2}{}_{r_1 n_3} (j_1 j_2 j_3)^{r_2}{}_{n_1 n_2 n_5} \right.$$
$$\left. + \begin{pmatrix} \dot{n}_5 & \\ & n_3 \end{pmatrix} \langle j_1 j_2 | j_3 \rangle^{\dot{n}_1 \dot{n}_2}{}_{r_1 \dot{n}_4} (j_1 j_2 j_3)^{r_2}{}_{\dot{n}_1 \dot{n}_2 \dot{n}_5} \right\}. \tag{7.8}$$

For each r_1, we may perform an orthogonal transform in the $3jm$ multiplicity space to give

$$\langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{r_1 m_3}$$
$$= \pm [j_3]^{1/2} \delta_{m_3 \dot{m}_4} \begin{bmatrix} \dot{m}_4 & \\ & m_5 \end{bmatrix} (j_1 j_2 j_3)^{m_1 m_2 m_5}{}_{r_1}, \tag{7.9}$$

where the normalization of the coupling coefficients and the $3jm$ symbols has been used to give $|U| = 1$. These $3jm$ symbols possess the orthogonality property

$$(j_1 j_2 j_3)^{m_1 m_2 m_3}{}_{r_2 z} (j_1 j_2 j_3)^{r_2}{}_{m_1 m_2 m_4} = \delta^{r_2}{}_{r_1} \delta^{m_3}{}_{m_4} [j_3]^{-1}, \tag{7.10}$$

where, however, r_1 and r_2 are not free to vary over the $3jm$ multiplicity space, but only over the coupling coefficient multiplicity subspace.

If the multiplicity of j_3 in $j_1 \otimes j_2$ is M_2, and the multiplicity of $\mathbf{0}$ in $j_1 \otimes j_2 \otimes j_3$ is M_3, the coupling coefficients only span on M_2-dimensional subspace of the $3jm$ multiplicity space. However, if the multiplicity of $\mathbf{0}$ in $j_3 \otimes j_3$ is M_1, then as $M_3 = M_1 M_2$, all the sets of coefficients span this space. Thus, all the $3jm$ coefficients may be found from one set of coupling coefficients and the commuting matrices

$$P^{m_1}{}_{m_2} = \langle j\mathbf{0} | j \rangle^{m_1 \mathbf{0}}{}_{m_2}$$

given earlier.

The third problem may be tackled by considering

$$\langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{r_1 m_3} j_3(u)^{m_3}{}_{n_3} j_3(u)^{m_4}{}_{n_4}$$
$$= j_1(u)^{m_1}{}_{n_1} j_2(u)^{m_2}{}_{n_2} j_3(u)^{m_4}{}_{n_4} \langle j_1 j_2 | j_3 \rangle^{n_1 n_2}{}_{r_1 n_3}. \tag{7.11}$$

We further expand the left-hand side by

$$j_3(u)^{m_3}{}_{n_3} j_3(u)^{m_4}{}_{n_4}$$
$$= \sum_{j_4} \langle j_3 j_3 | j_4 \rangle^{m_3 m_4}{}_{r_2 m_5} j_4(u)^{m_5}{}_{n_5} \langle j_3 j_3 | j_4 \rangle^{r_2 n_5}{}_{n_3 n_4}$$

and use, this time, the orthogonality equation (3.6) to give

$$\langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{r_1 m_3} \langle j_3 j_3 | \mathbf{0} \rangle^{m_3 m_4}{}_{r_2 \mathbf{0}} \langle j_3 j_3 | \mathbf{0} \rangle^{r_2 \mathbf{0}}{}_{n_3 n_4}$$
$$= (j_1 j_2 j_3)^{m_1 m_2 m_4}{}_{r_3} (j_1 j_2 j_3)^{r_3}{}_{n_1 n_2 n_4} \langle j_1 j_2 | j_3 \rangle^{n_1 n_2}{}_{r_1 r_3}.$$

Setting $V^{r_3}{}_{r_1 r_2}$ as the square invertible tensor

$$V^{r_3}{}_{r_1 r_2} = (j_1 j_2 j_3)^{r_3}{}_{n_1 n_2 n_4} \langle j_1 j_2 | j_3 \rangle^{n_1 n_2}{}_{r_1 n_3} \langle j_3 j_3 | \mathbf{0} \rangle^{n_3 n_4}{}_{r_2 \mathbf{0}} \tag{7.12}$$

gives

$$\langle j_1 j_2 | j_3 \rangle^{m_1 m_2}{}_{r_1 m_3} \langle j_3 j_3 | \mathbf{0} \rangle^{m_3 m_4}{}_{r_2 \mathbf{0}} = V^{r_3}{}_{r_1 r_2} (j_1 j_2 j_3)^{m_1 m_2 m_4}{}_{r_3}. \tag{7.13}$$

The corresponding antilinear equations show that V is real.

In representation theory the coupling coefficient $\langle j_3 j_3 | \mathbf{0} \rangle^{m_3 m_4}{}_{r_2 \mathbf{0}}$ is of multiplicity one and is simply related to the Wigner tensor. Thus, this equation has an analog in representation theory, where it would be equivalent to Eq. (7.7), but here the multiplicity may be as high as four, and the two equations are distinct.

We have not completely answered the third problem, in that the left-hand side of Eq. (7.13) is not a coupling coefficient, but the coupling coefficients

$$\langle j_3 \mathbf{0} | j_3 \rangle^{m_3 \mathbf{0}}{}_{m_3} \text{ and } \langle j_3 j_3 | \mathbf{0} \rangle^{m_3 m_4}{}_{r_2 \mathbf{0}}$$

are given in terms of $3jm$ symbols by Eq. (7.9) which are related by the transformation properties discussed in the next section, and it is a straightforward matter to write the equation in terms of the full set of coupling coefficients.

We conclude this section with some brief comments on the Wigner tensor. Setting $j_1 = j_3$ and $j_2 = \mathbf{0}$ in Eq. (7.9) gives

$$\langle j_1 \mathbf{0} | j_1 \rangle^{m_1 \mathbf{0}}{}_{1 m_3}$$
$$= \pm [j_1]^{1/2} \delta_{m_3 m_4} \begin{bmatrix} m_4 & \\ & m_5 \end{bmatrix} (j_1 \mathbf{0} j_1)^{m_1 \mathbf{0} m_5}{}_1.$$

Trivially, a set of coupling coefficients is

$$\langle j_1 \mathbf{0} | j_1 \rangle^{m_1 \mathbf{0}}{}_{1 m_3} = \delta^{m_1}{}_{m_3}$$

so that

$$\begin{bmatrix} m_4 & \\ & m_5 \end{bmatrix} = \pm [j_1]^{1/2} \delta^{m_4 m_3} (j_1 \mathbf{0} j_1)^1{}_{m_3 \mathbf{0} m_5}$$

by Eq. (7.10). It must be stressed though that this equation should not be used to give the Winger tensor as happens in representation theory. There is an arbitrariness in the $3jm$ symbol which would make it hard to decide which to use, and besides, an standard form for the Wigner tensor has been given in the last section. Rather, this equation should be regarded as giving one of the $3jm$ symbols.

8. PROPERTIES OF THE 3*jm* SYMBOLS

With minor preparation, we may almost quote these from Derome and Sharp[17] and Derome[18]. Firstly,

$$(j_1j_2j_3)^{m_1m_2m_3}{}_{r_1}(j_1j_2j_3)^{r_1}{}_{n_1n_2n_3} = \frac{2}{|G|}\int_H j_1(u)^{m_1}{}_{n_1} j_2(u)^{m_2}{}_{n_2} j_3(u)^{m_3}{}_{n_3}\, du.$$

Using

$$j_1(u)^{m_1}{}_{n_1} = \begin{bmatrix} m_1 & \\ & \dot{m}_4 \end{bmatrix} j_1(u)^{\dot{m}_4}{}_{\dot{n}_4} \begin{bmatrix} \dot{n}_4 & \\ & n_1 \end{bmatrix}, \text{ etc.,}$$

this gives

$$(j_1j_2j_3)^{m_1m_2m_3}{}_{r_1}(j_1j_2j_3)^{r_1}{}_{n_1n_2n_3} = \begin{pmatrix} m_1 & \\ & \dot{m}_4 \end{pmatrix}\begin{pmatrix} m_2 & \\ & \dot{m}_5 \end{pmatrix}\begin{pmatrix} m_3 & \\ & \dot{m}_6 \end{pmatrix}(j_1j_2j_3)^{\dot{m}_4\dot{m}_5\dot{m}_6}{}_{\dot{r}_2} \times \begin{bmatrix} \dot{n}_4 & \\ & n_1 \end{bmatrix}\begin{bmatrix} \dot{n}_5 & \\ & n_2 \end{bmatrix}\begin{bmatrix} \dot{n}_6 & \\ & n_3 \end{bmatrix}(j_1j_2j_3)^{\dot{r}_2}{}_{\dot{n}_4\dot{n}_5\dot{n}_6}. \quad (8.1)$$

Taking $(j_1j_2j_3)^{r_1}{}_{n_1n_2n_3}$ to the other side and setting

$$A(123)^{\dot{r}_2}{}_{r_1} = \begin{bmatrix} \dot{n}_4 & \\ & n_1 \end{bmatrix}\begin{bmatrix} \dot{n}_5 & \\ & n_2 \end{bmatrix}\begin{bmatrix} \dot{n}_6 & \\ & n_3 \end{bmatrix} \times (j_1j_2j_3)^{\dot{r}_2}{}_{\dot{n}_4\dot{n}_5\dot{n}_6}(j_1j_2j_3)^{n_1n_2n_3}{}_{r_1} \quad (8.2)$$

gives

$$(j_1j_2j_3)^{m_1m_2m_3}{}_{r_1} = A(123)^{\dot{r}_2}{}_{r_1} \times \begin{pmatrix} m_1 & \\ & \dot{m}_4 \end{pmatrix}\begin{pmatrix} m_2 & \\ & \dot{m}_5 \end{pmatrix}\begin{pmatrix} m_3 & \\ & \dot{m}_6 \end{pmatrix} \times (j_1j_2j_3)^{\dot{m}_4\dot{m}_5\dot{m}_6}{}_{\dot{r}_2}. \quad (8.3)$$

The antilinear expression gives similarly

$$A(123)^{\dot{r}_2}{}_{r_1} = \delta^{\dot{r}_1}{}_{r_1}\delta^{\dot{r}_2}{}_{r_2}\begin{bmatrix} n_4 & \\ & \dot{n}_1 \end{bmatrix}\begin{bmatrix} n_5 & \\ & \dot{n}_2 \end{bmatrix}\begin{bmatrix} n_6 & \\ & \dot{n}_3 \end{bmatrix} \times (j_1j_2j_3)^{r_2}{}_{n_4n_5n_6}(j_1j_2j_3)^{\dot{n}_1\dot{n}_2\dot{n}_3}{}_{\dot{r}_1} \quad (8.4)$$

so that A is numerically real.

Let us now consider two different reductions of the triple product, one with j_1 and j_2 permuted. From Eq. (5.3)

$$(j_1j_2j_3)^{m_1m_2m_3}{}_{r_1}(j_1j_2j_3)^{r_1}{}_{n_1n_2n_3} = (j_2j_1j_3)^{m_2m_1m_3}{}_{r_2}(j_2j_1j_3)^{r_2}{}_{n_2n_1n_3}$$

so that

$$(j_1j_2j_3)^{m_1m_2m_3}{}_{r_1} = M(12,3)^{r_2}{}_{r_1}(j_2j_1j_3)^{m_2m_1m_3}{}_{r_2}, \quad (8.5)$$

with

$$M(12,3)^{r_2}{}_{r_1} = (j_2j_1j_3)^{r_2}{}_{n_2n_1n_3}(j_1j_2j_3)^{n_1n_2n_3}{}_{r_1}. \quad (8.6)$$

From Eq. (5.4),

$$M(12,3)^{r_2}{}_{r_1} = \delta^{r_2}{}_{r_2}(j_2j_1j_3)^{r_2}{}_{n_2n_1n_3}(j_1j_2j_3)^{n_1n_2n_3}{}_{r_1}$$

so that $M(12,3)$ is numerically real. This reality in fact forms the only difference between the analyses of Derome and Sharp[17] and Derome[18] and the corresponding analysis needed here. Since their results are not in any manner dependent on M being nonreal, we may use them without modification. We are most interested in the possibility of diagonalizing M over the multiplicity space, and from Derome[18] we find the following:

(a) If none of j_1, j_2, j_3 is equivalent, M may be diagonalized to the identity matrix,

(b) If exactly two are equivalent, then every transposition may be diagonalized to

$$\begin{pmatrix} 1_S & 0 \\ 0 & -1_A \end{pmatrix}$$

and every cyclic permutation leaves the 3*jm* symbol invariant,

(c) If $j_1 \equiv j_2 \equiv j_3$, M may be diagonalized if f

$$\int_H [\chi(u)]^3\, du = \int_H \chi(u^3)\, du.$$

This equation refers to the character of the ICR—not of the IR. If the ICR is of types (b) or (c), there is the possibility that the 3*jm* symbols of the grey group may be diagonalized even if the 3*jm* symbols of the linear subgroup may not be (and *vice versa*).

Finally, in this section, we give some simple conditions that ensure the reality of the 3*jm* symbols. So far we have only used the operator θ to generate the antilinear coset $G - H$ from H. However, any antilinear operator will serve for this.[43] Suppose, then, that we have an antilinear operator $\bar{\theta}$ which is numerically the identity matrix in all three ICR's j_1, j_2, and j_3. From Eq. (5.4)

$$(j_1j_2j_3)^{r}{}_{n_1n_2n_3} = \frac{2}{|G|}(j_1j_2j_3)^{r}{}_{m_1m_2m_3} \times \int_{G-H} j_1(a)^{m_1}{}_{\dot{n}_1} j_2(u)^{m_2}{}_{\dot{n}_2} j_3(a)^{m_3}{}_{\dot{n}_3}\, da.$$

We may replace $j_1(a)^{m_1}{}_{\dot{n}_1}$ by $j_1(u\bar{\theta})^{m_1}{}_{\dot{n}_1} = j_1(u)^{m_1}{}_{n_4}\delta^{n_4}{}_{\dot{n}_1}$, etc. to give

$$(j_1j_2j_3)^{r}{}_{n_1n_2n_3} = \frac{2}{|G|}\delta^{n_4}{}_{\dot{n}_1}\delta^{n_5}{}_{\dot{n}_2}\delta^{n_6}{}_{\dot{n}_3}(j_1j_2j_3)^{r}{}_{m_1m_2m_3} \times \int_H j_1(u)^{m_1}{}_{n_4} j_2(u)^{m_2}{}_{n_5} j_3(u)^{m_3}{}_{n_6}\, du. = \delta^{n_4}{}_{\dot{n}_1}\delta^{n_5}{}_{\dot{n}_2}\delta^{n_6}{}_{\dot{n}_3}(j_1j_2j_3)^{r}{}_{n_4n_5n_6}, \quad (8.7)$$

giving reality. In a similar manner, we may show the following possibilities:

(a) If $\bar{\theta} = I$ for all three, the 3*jm* symbols are all real.

(b) If $\bar{\theta} = -I$ for two of the ICR's and $+I$ for the third, the symbols are again real.

(c) If $\bar{\theta} = -I$ for all three, or $-I$ for one and $+I$ for the other two, the symbols are all imaginary.

There is no guarantee that an operator $\bar{\theta}$ exists satisfying any of these conditions, but if it does this is a very quick test for reality or nonreality. An important special case is grey SU(2) where in Fano–Racah standardization

$$\bar{\theta} = C_v^2\theta^{-1}$$

to give case (a).

9. THE WIGNER–ECKART THEOREM

Aviran and Zak[48] give a form of the Wigner–Eckart theorem for general linear/antilinear groups, but this is unsatisfactory in some respects. For an irreducible tensor $T(kq)$ their Eqs. (3) and (4) for the matrix elements become in our

notation

$$\langle j_1 m_1 | T(kq_1) | j_2 m_2 \rangle = \frac{2}{|G|} \int_H j_1(u)^{n_1}{}_{m_1} k(u)^{q_2}{}_{q_1} \times j_2(u)^{n_2}{}_{m_2} \langle j_1 n_1 | T(kq_2) | j_2 n_2 \rangle \, du \tag{9.1}$$

and

$$\langle j_1 m_1 | T(kq_1) | j_2 m_2 \rangle = \frac{2}{|G|} \int_{G-H} j_1(a)^{n_1}{}_{m_1} k(a)^{q_2}{}_{q_1} \times j_2(a)^{n_2}{}_{m_2} \langle j_1 n_1 | T(kq_2) | j_2 n_2 \rangle^* \, du. \tag{9.2}$$

They consider the reduction of the product $k \otimes j_2$ to j_1^*, but as we have already seen, this causes difficulties. It would also be equally valid to reduce $j_1^* \otimes k$ to j_2^* or $j_1^* \otimes j_2$ to k^* and examination of their results shows that this would give forms apparently dependent on the reduction used. These problems may all be avoided by reducing the triple product to the identity ICR. Use of the Wigner tensor for the grey groups will also yield a more familiar form. In Eq. (9.1) we have

$$\langle j_1 m_1 | T(kq_1) | j_2 m_2 \rangle = \frac{2}{|G|} \int_H \begin{pmatrix} \dot{n}_1 & \\ & n_3 \end{pmatrix} j_1(u)^{n_3}{}_{m_3} \begin{bmatrix} m_3 & \\ & \dot{m}_1 \end{bmatrix} \times k(u)^{q_2}{}_{q_1} j_2(u)^{n_2}{}_{m_2} \langle j_1 n_1 | T(kq_2) | j_2 n_2 \rangle \, du$$

$$= \begin{pmatrix} \dot{n}_1 & \\ & n_3 \end{pmatrix} (j_1 k j_2)^{n_3 q_2 n_2}{}_r (j_1 k j_2)^r{}_{m_3 q_1 m_2} \times \begin{bmatrix} m_3 & \\ & \dot{m}_1 \end{bmatrix} \langle j_1 n_1 | T(kq_2) | j_2 n_2 \rangle.$$

Setting the reduced matrix element

$$\langle j_1 \| T(k) \| j_2 \rangle_r = \begin{pmatrix} \dot{n}_1 & \\ & n_3 \end{pmatrix} (j_1 k j_2)^{n_3 q_2 n_2}{}_r \langle j_1 n_1 | T(kq_2) | j_2 n_2 \rangle \tag{9.3}$$

gives

$$\langle j_1 m_1 | T(kq_1) | j_2 m_2 \rangle = \langle j_1 \| T(k) \| j_2 \rangle_r \begin{bmatrix} m_3 & \\ & \dot{m}_1 \end{bmatrix} (j_1 k j_2)^r{}_{m_3 q_1 m_2}, \tag{9.4}$$

which is the desired result. The corresponding antilinear equation gives the reduced matrix element as numerically real.

10. RACAH'S LEMMA

Racah's lemma[4] has proved to be a fundamental result for any application of the descent in symmetry technique. Briefly, for IR's it relates the $3jm$ symbols of a group G to the $3jm$ symbols of a subgroup K by a factor or unitary transformation independent of the m values. To use the descent in symmetry technique in grey groups we shall find the lemma of equal importance. However, as the proofs we have sighted for linear groups directly use Schur's lemma,[4,40] and Schur's lemma only holds in a restricted sense for grey groups, we offer a proof based on the orthogonality relations. We may distinguish between two different types of subgroup of a grey group and the lemma will assume a different form for each. These are the following: a grey subgroup, consisting of both linear and antilinear operators, and a subgroup consisting of linear operators only.

We consider first a grey subgroup K of grey G. We label the ICR's of G by j as usual, and of K by k. Upon restriction to K, each ICR of G may be reduced to a direct sum of ICR's of K by some unitary transformation. Applying the inverse transformation to the ICR's of G, we obtain "symmetry adapted" ICR's of G—i.e., on restriction to K the matrices are already in block diagonal form:

$$j(u) = \begin{pmatrix} k(u) & & & 0 \\ & k'(u) & & \\ & & \ddots & \\ 0 & & & \ddots \end{pmatrix}$$

and similarly for the antilinear operators. We reduce the triple product in G in the usual way, i.e., if the multiplicity of $\mathbf{0}$ in $j_1 \otimes j_2 \otimes j_3$ is M_1, the first M_1 colums of the reduction matrix U are the $3jm$ symbols, with the other columns of U being formed from the other ICR's in this triple product. Upon restriction to K some of these may reduce to the identity ICR of K and we would have the form

$$\begin{pmatrix} I(u) & & & & & \\ & k_4(u) & & & & \\ & & k_5(u) & & & \\ & & & I(u) & & \\ & & & & \ddots & \\ & & & & & \ddots \end{pmatrix}.$$

If we sum over the unitary elements H' of K, we shall certainly get the $3jm$ symbols as in Eq. (5.3) but we shall also get the unitary elements from these other occurrences of the identity

$$(j_1 j_2 j_3)^{m_1 m_2 m_3}{}_{r_1} (j_1 j_2 j_3)^{r_1}{}_{n_1 n_2 n_3} + (j_1 j_2 j_3)^{m_1 m_2 m_3}{}_{r_2 j} (j_1 j_2 j_3)^{r_2 j}{}_{n_1 n_2 n_3} = \frac{2}{|K|} \int_{H'} j_1(u)^{m_1}{}_{n_1} j_2(u)^{m_2}{}_{n_2} j_3(u)^{m_3}{}_{n_3} \, du. \tag{10.1}$$

However, as U is unitary, we may pick out a particular $3jm$ symbol by orthogonality:

$$(j_1 j_2 j_3)^{m_1 m_2 m_3}{}_{r_1} = \frac{2}{|K|} (j_1 j_2 j_3)^{n_1 n_2 n_3}{}_{r_1} \times \int_{H'} j_1(u)^{m_1}{}_{n_1} j_2(u)^{m_2}{}_{n_2} j_3(u)^{m_3}{}_{n_3} \, du. \tag{10.2}$$

We let the reduction matrix of $(k_1 \oplus \cdots) \otimes (k_2 \oplus \cdots) \otimes (k_3 \oplus \cdots)$ be V, and if the multiplicity of $\mathbf{0}$ in $k_1 \otimes k_2 \otimes k_3$ is M_2, the first M_2 columns of V be the $3jm$ symbols $(k_1 k_2 k_3)^{m_1 m_2 m_3}{}_{r_2}$. Then for $m_1, n_1 \leqslant [k_1]$, $m_2, n_2 \leqslant [k_2]$, $m_3, n_3 \leqslant [k_3]$ we have

$$(k_1 k_2 k_3)^{m_1 m_2 m_3}{}_{r_2} (k_1 k_2 k_3)^{r_2}{}_{n_1 n_2 n_3} = \frac{2}{|K|} \int_{H'} k_1(u)^{m_1}{}_{n_1} k_2(u)^{m_2}{}_{n_2} k_3(u)^{m_3}{}_{n_3} \, du$$

$$= \frac{2}{|K|} \int_{H'} j_1(u)^{m_1}{}_{n_1} j_2(u)^{m_2}{}_{n_2} j_3(u)^{m_3}{}_{n_3}\, du. \qquad (10.3)$$

For $m_1 \leqslant [k_1]$ but $n_1 > [k_1]$, etc., we may take the $3jm$ symbols for K as zero and substitute into Eq. (10.2) to give

$$(j_1 j_2 j_3)^{m_1 m_2 m_3}{}_{r_1}$$
$$= (k_1 k_2 k_3)^{m_1 m_2 m_3}{}_{r_2} (k_1 k_2 k_3)^{r_2}{}_{n_1 n_2 n_3} (j_1 j_2 j_3)^{n_1 n_2 n_3}{}_{r_1}.$$

Defining the isoscalar by

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ k_1 & k_2 & k_3 \end{pmatrix}^{r_2}_{r_1} = (k_1 k_2 k_3)^{r_2}{}_{n_1 n_2 n_3} (j_1 j_2 j_3)^{n_1 n_2 n_3}{}_{r_1}, \qquad (10.4)$$

gives us

$$(j_1 j_2 j_3)^{m_1 m_2 m_3}{}_{r_1} = \begin{pmatrix} j_1 & j_2 & j_3 \\ k_1 & k_2 & k_3 \end{pmatrix}^{r_2}_{r_1} (k_1 k_2 k_3)^{m_1 m_2 m_3}{}_{r_2}, \qquad (10.5)$$

$$(10.6)$$

As we might expect, the antilinear equations show the isoscalar is numerically real.

For a linear subgroup the analysis proceeds as above, but of course, as we have no antilinear operators, we have no reality condition. This is to be expected as we have a free choice of phase for $3jm$ symbols of a linear group which we do not have for a grey group.

The properties of the isoscalar under permutations follow from the properties of the two sets of $3jm$ symbols in Eq. (10.4). If $M(12,3)$ is the transposition matrix in G, and $N(12,3)$ the transposition matrix in K, from Eq. (8.5)

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ k_1 & k_2 & k_3 \end{pmatrix}^{r_3}_{r_1}$$
$$= M(12,3)^{r_3}{}_{r_4} N^{-1}(12,3)^{r_2}{}_{r_4} \begin{pmatrix} j_2 & j_1 & j_3 \\ k_2 & k_1 & k_3 \end{pmatrix}^{r_4}_{r_3}. \qquad (10.7)$$

ACKNOWLEDGMENTS

We would like to thank Dr. G. Sneddon of the Papua New Guinea University of Technology for showing us the covariant notation and for discussions involving both the best form of it for our purposes and the properties of antilinear operators. An earlier version of this paper contained many flaws, and we are very grateful to the referee for pointing these out.

[1]G. Racah, Phys. Rev. **61**, 186 (1942).
[2]G. Racah, Phys. Rev. **62**, 438 (1942).
[3]G. Racah, Phys. Rev. **63**, 367 (1943).
[4]G. Racah, Phys. Rev. **76**, 1352 (1949).
[5]U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1959).
[6]E.P. Wigner, "On the Matrices which Reduce the Kronecker Products of SR Groups," privately circulated but later printed in *Quantum Theory of Angular Momentum*, edited by L.C. Biedenharn and H. Van Dam (Academic, New York 1965).
[7]A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, N. J., 1960).
[8]H.A. Jahn and J. Hope, Phys. Rev. **93**, 318 (1954).
[9]R.J. Ord-Smith, Phys. Rev. **94**, 1227 (1954).
[10]J. P. Elliot and B. H. Flowers, Proc. R. Soc. London A **229**, 536 (1955).
[11]F.R. Innes and C.W. Ufford, Phys. Rev. **111**, 194 (1958).
[12]E.P. Wigner, *Gruppentheorie* (Viewig, Brannschweig 1931).
[13]C. Eckart, Revs. Mod. Phys. **2**, 305 (1931).
[14]A. Gamba, J. Math. Phys. **10**, 872 (1969).
[15]T. Lulek, Acta. Phys. Polon. A **52**, 189 (1977).
[16]G.F. Koster, Phys. Rev. **109**, 227 (1958).
[17]J.-R. Derome and W.T. Sharp, J. Math. Phys. **10**, 1584 (1965).
[18]J.-R. Derome, J. Math. Phys. **7**, 612 (1966).
[19]J.S. Griffith, *The Irreducible Tensor Method for Molecular Symmetry Groups* (Prentice-Hall, Englewood Cliffs, N. J., (1962).
[20]P.H. Butler and B.G. Wybourne, Int. J. Quantum Chem. **10**, 581 (1976).
[21]P.A. Dobosh, Phys. Rev. A **5**, 2376 (1972).
[22]R.M. Golding, Mol. Phys. **21**, 157 (1971).
[23]R.M. Golding, Mol. Phys. **26**, 661 (1973).
[24]R.M. Golding and J.D. Newmarch, Mol. Phys. **33**, 1301 (1977).
[25]J. Moret-Bailly and L. Gautier, J. Mol. Spectrosc. **15**, 355 (1965).
[26]F. Michelot and J. Moret-Bailley, J. Phys. (Paris) **36**, 451 (1975).
[27]M. Kibler, Int. J. Quantum Chem. **3**, 795 (1969).
[28]M.R. Kibler, J. Math. Phys. **17**, 855 (1976).
[29]M.R. Kibler and P.A.M. Guichon, Int. J. Quantum Chem. **10**, 87 (1976).
[30]E. König and S. Kremer, Theor. Chim. Acta **32**, 27 (1973).
[31]E. König and S. Kremer, Z. Natursforsch. Teil A **29**, 1179 (1974).
[32]T. Buch, J. Math. Phys. **13**, 1892 (1972).
[33]S.E. Harnung and C.E. Schäffer, Struct. Bonding **12**, 257 (1972).
[34]S.E. Harnung, Mol. Phys. **26**, 473 (1973).
[35]B.R. Hollebone, A.B.P. Lever, and J.C. Donini, Mol. Phys. **22**, 155 (1971).
[36]B.R. Hollebone and J.C. Donini, Theor. Chim. Acta **37**, 233 (1975).
[37]M.R. Kibler, J. Phys. A **10**, 2041 (1977).
[38]G. Fieck, J. Phys. B **10**, L1 (1977).
[39]P.H. Butler and B.G. Wybourne, Int. J. Quantum Chem. **10**, 615 (1976).
[40]P.H. Butler, "Properties and Application of Point Group Coupling Coefficients," in *Recent Advances in Group Theory and their Application to Spectroscopy*, edited by J.C. Donini (Plenum, New York, 1979).
[41]H.A. Kramers, Proc. R. Soc. Amsterdam **33**, 959 (1930).
[42]E.P. Wigner, *Group Theory and its Applications to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).
[43]C.J. Bradley and B.L. Davies, Rev. Mod. Phys. **40**, 359 (1968). This paper contains extensive reference to work published prior to 1968.
[44]A. Aviran and D.B.Litvin, J. Math. Phys. **14**, 1491 (1973).
[45]P. Rudra, J. Math. Phys. **15**, 2031 (1974).
[46]I. Säkata, J. Math. Phys. **15**, 1710 (1974).
[47]J.O. Dimmock, J. Math. Phys. **4**, 1307 (1963).
[48]A. Aviran and J. Zak, J. Math. Phys. **9**, 2138 (1968).
[49]L. Jansen and M. Boon, *Theory of Finite Groups. Applications in Physics* (North-Holland, Amsterdam, 1967).
[50]We use "$3jm$" rather than "$3j$" in accordance with the notation of Butler and Wybourne[20] as $3j$ is better reserved for the permutation matrix, independent of m, of the $3jm$ symbols.
[51]H.A. Buchdahl, J. Aust. Math. Soc. **6**, 402 (1966).
[52]H.A. Buchdahl, Lecture notes ANU, Canberra, Australia (1971).
[53]R.S. Mulliken, Phys. Rev. **43**, 279 (1933).
[54]J.O. Dimmock and R.G. Wheeler, *The Mathematics of Physics and Chemistry* (Van Nostrand, New York, 1964).
[55]J.S. Griffith, *The Theory of Transition Metal Ions* (Cambridge U.P, Cambridge, England, 1971).
[56]A.G. McLellan, Nucl. Phys. **13**, 104 (1959).

Representations of Osp(2, 1) and the metaplectic representation

J. W. B. Hughes
Department of Applied Mathematics, Queen Mary College, Mile End Road, London E1 4NS, United Kingdom

(Received 28 November 1979; accepted for publication 14 January 1980)

Shift operator techniques are used to treat the irreducible representations of the superalgebra Osp(2, 1). Apart from obtaining the well known gradestar dispin representations which arise when the even part is the compact SU(2) algebra, the case when the star conditions on the even part are those satisfied by the noncompact SU(1, 1) algebra is also treated. In this case no gradestar representations arise, and the star representations are found to consist of the direct sum of two discrete series representations of SU(1, 1). One of these representations can be realized in terms of functions of a single complex variable, and turns out to be a simple example of a metaplectic representation.

PACS numbers: 02.20.Qs

1. INTRODUCTION

The superalgebra Osp(2,1), which is the subject of this paper, has been considered by several authors.[1–4] A basis for the even part of this algebra consists of the SU(2) [or Sl(2)] generators $l_0, l_\pm$, and the odd part has basis elements $q_{\pm 1/2}$, which form a two-dimensional tensor representation of the SU(2) algebra. The commutation and anticommutation relations satisfied by these elements are:

$$[l_0,l_\pm] = \pm l_\pm, \quad [l_+,l_-] = 2l_0,$$
$$[l_0,q_{\pm 1/2}] = \pm \tfrac{1}{2} q_{\pm 1/2}, \quad [l_\pm,q_{\mp 1/2}] = q_{\pm 1/2}, \quad (1.1)$$
$$\{q_{\pm 1/2},q_{\pm 1/2}\} = \pm l_\pm, \quad \{q_{1/2},q_{-1/2}\} = -l_0.$$

In this paper we classify the representations of Osp(2,1) using SU(2) shift operators developed by Hughes and Yadegar,[5] restricting our considerations, however, to those representations which are star or gradestar.[2] The case where the star (or Hermiticity) conditions on the even part are those appropriate for the SU(2) Lie algebra, namely $l_0^\dagger = l_0$, $l_\pm^\dagger = -l_\mp$, is dealt with first, and we merely rederive the well known dispin representations obtained by other authors. These turn out to be all gradestar, as was shown by Nahm, Rittenberg, and Scheunert.[3]

We then consider the case where the even part satisfies the star conditions $l_0^\dagger = l_0$, $l_\pm^\dagger = -l_\mp$, in other words we look at representations of Osp(2,1) which subduce to the infinite dimensional Hermitian representations of the noncompact SU(1,1), or Sl(2), Lie algebra. Although the shift operator techniques of Hughes and Yadegar have been used for the analysis of representations of many different Lie algebras with respect to an SU(2) or O(3) subalgebra (see for instance Hughes and Backhouse[6] and references therein), this is the first time they have been used for algebras containing the noncompact SU(1,1) subalgebra. Also, the author believes this to be the first case where the infinite dimensional irreducible representations of a simple superalgebra have been classified. The term "Osp(2,1)" is usually used to denote the case where the even part is SU(2), so here we are extending the use of the term to cover the case where the even part of the superalgebra is SU(1,1).

In this case we find that no gradestar representations can occur, but we get two classes of star representations, depending on whether we take $q_{\pm 1/2}^\dagger = q_{\mp 1/2}$ or $q_{\pm 1/2}^\dagger = -q_{\mp 1/2}$. Both star conditions give rise to "dispin" representations, although again the term "dispin" is used rather loosely to include the case where the SU(1,1) Casimir L^2 has eigenvalues $l(l+1)$ where l is nonpositive half-integral. We find that, for $q_{\pm 1/2}^\dagger = q_{\mp 1/2}$, the irreducible star representations of Osp(2,1) consist of the direct sum of two SU(1,1) discrete series D^- representations,[7] whereas if $q_{\pm 1/2}^\dagger = -q_{\mp 1/2}$, we get two D^+ representations of SU(1,1). Since the two cases give rise to mutually contragredient sets of star representations, only the latter case is considered in detail.

One of these representations, namely that for which l has the values $-\frac{3}{4}$ and $-\frac{1}{4}$ turns out to be a particular example of a metaplectic representation, considered in more generality by Sternberg and Wolf.[8] These authors show that every symplectic group Sp($2m$) has a special two-valued representation called the metaplectic representation, which is a reducible representation with two irreducible components. For the particular case where $m = 1$, Sp(2)$\simeq$Sl(2)$\simeq$SU(1,1) and the metaplectic representation is, at the Lie algebra level, just the above reducible "dispin" representation of SU(1,1).

Sternberg and Wolf also show that in the general case Sp($2m$) can be embedded in a larger algebra Sp($2m$) + $\mathbb{R}^{2m}$ which admits of a Hermitian structure with respect to which it becomes the orthosymplectic superalgebra Osp($2m$,1). The metaplectic representation of Sp($2m$) extends to an irreducible representation of Osp($2m$,1), which can be realized in terms of the space $\mathscr{H}$ of all holomorphic functions $f:\mathbb{C}^m \to \mathbb{C}$ such that $\int |f(z)|^2 \exp(-|z|^2)\, d\lambda(z) < \infty$, where λ is the Lebesgue measure on $\mathbb{C}^m$. On restriction of the metaplectic representation to Sp($2m$), the two irreducible components are supported by the subspaces $\mathscr{H}^\pm$ of $\mathscr{H}$, where $\mathscr{H}^+$ and $\mathscr{H}^-$ are the closed spans of the set of functions $z^n \equiv (z_1^{n_1},\ldots,z_m^{n_m})$, $n_\alpha \in \mathbb{Z}$, with $|n| = \Sigma n_\alpha$ even and odd, respectively. The results of this paper show that for the case where $m = 1$, the metaplectic representation of Osp(2,1) is generic in the sense that every irreducible star representation of Osp(2,1) shares this property of having precisely two irredu-

 0022-2488/81/020245-06$1.00

cible components on restriction to $Sp(2) \simeq SU(1,1)$.

Also, for the case where $m = 1$, the central extension of $Sp(2) + \mathbb{R}^2$, which we shall refer to in a following paper as $C(Sl(2)_A T^2)$, also has an irreducible representation which on restriction to Sl(2) yields the metaplectic representation. We shall show there that for this Lie algebra, the metaplectic representation is not generic, but in fact pathological, being the only "dispin" irreducible representation it possesses, and moreover the only irreducible representation it has in common with Osp(2,1).

In Sec. 2, we summarise the algebraic properties of Osp(2,1), and write down the shift operators which will be used in the analysis of its representations. In Sec. 3 we rederive the well-known representations of the compact version, and in Sec. 4 we classify the infinite dimensional representations of its noncompact version. Finally, in Sec. 5 we consider the metaplectic representation in more detail.

2. THE SUPERALGEBRA Osp(2,1)

The defining relations of Osp(2,1) have already been given in Eqs. (1.1). Osp(2,1) possesses a single invariant

$$K_2 = L^2 + \tfrac{1}{2}(q_{1/2}q_{-1/2} - q_{-1/2}q_{1/2}), \tag{2.1}$$

where L^2 is the SU(2) Casimir operator, $L^2 = l_- l_+ + l_0^2 + l_0$. Now using Eqs. (1.1), one may show that

$$(q_{1/2}q_{-1/2} - q_{-1/2}q_{1/2})^2 = L^2 + (q_{1/2}q_{-1/2} - q_{-1/2}q_{1/2}),$$

so one can derive the following relation satisfied by K_2:

$$K_2^2 = (L^2 + \tfrac{1}{4})(2K_2 - L^2). \tag{2.2}$$

Given an irreducible representation (I.R.) of Osp(2,1) for which K_2 has the eigenvalue α, then acting on a state $|\, l,m\rangle$ of the I.R. with both sides of Eq. (2.2), one obtains

$$\alpha = \tfrac{1}{2}l(2l+1), \quad \text{or} \quad \tfrac{1}{2}(l+1)(2l+1). \tag{2.3}$$

The fact that α has the same value for all the states of the I.R. therefore shows that the range of permissible l values is severely restricted by Eq. (2.3); in fact only two distinct l values can arise, as we shall see shortly.

The shift operators we shall need in order to analyse the I.R. of Osp(2,1) are particular cases of more general operators derived by Hughes and Yadegar.[5] These shift the eigenvalues l and m of R and l_0 by $\pm\frac{1}{2}$, where $R(R+1) \equiv L^2$, when acting to the right on eigenstates of these operators, and are given by

$$\begin{aligned} O^{1/2,1/2} &= q_{1/2}(R + l_0 + 1) + q_{-1/2}\, l_+ , \\ O^{-1/2,-1/2} &= -q_{-1/2}(R + l_0) + q_{1/2} l_- . \end{aligned} \tag{2.4}$$

The fact that there is only one pair of such operators means that no l-degenracies can occur within the I.R. of this algebra.

It will often be more convenient to use the normalized operators whose action on eigenstates $|l,m\rangle$ are related to those of the above operators by

$$\begin{aligned} A_l^{1/2} &= (l+m+1)^{-1/2} O_{l,m}^{1/2,1/2}, \\ A_l^{-1/2} &= (l+m)^{-1/2} O_{l,m}^{-1/2,-1/2}. \end{aligned} \tag{2.5}$$

The following L^2-commuting operators form the basis of the following analysis:

$$A_{l-1/2}^{1/2} A_l^{-1/2} = -l(2K_2 - (2l+1)(l+1)), \tag{2.6}$$

$$A_{l+1/2}^{-1/2} A_l^{1/2} = (l+1)(2K_2 - l(2l+1)). \tag{2.7}$$

These two operators contain the whole structure of the irreducible representations in that minimum and maximum l values are determined entirely by the vanishing of their eigenvalues. In addition to (2.6), (2.7), we also have the identities

$$A_{l+1/2}^{1/2} A_l^{1/2} = A_{l-1/2}^{-1/2} A_l^{-1/2} = 0. \tag{2.8}$$

These show that it is impossible to change the l values of states within an I.R. of Osp(2,1) by more than $\frac{1}{2}$ so that only two distinct l values occur, i.e., that the I.R. are "dispin". Hence l must have a minimum value, $\underline{l} = j$, say, for which $A_{j-1/2}^{1/2} A_j^{-1/2} = 0$.

Using Eq. (2.6), this gives

$$j(2K_2 - (2j+1)(j+1)) = 0, \tag{2.9}$$

so that we have two possibilities: A:$K_2 = \frac{1}{2}(2j+1)(j+1)$ and B:$j = 0$.

A. Using $K_2 = \frac{1}{2}(2j+1)(j+1)$ in (2.6) and (2.7) gives

$$A_{j+1/2}^{-1/2} A_j^{1/2} \,|\, j\rangle = (j+1)(2j+1)\,|\, j\rangle, \tag{2.10}$$

$$A_j^{1/2} A_{j+1/2}^{-1/2} \,|\, j+\tfrac{1}{2}\rangle = (j+1)(2j+1)\,|\, j+\tfrac{1}{2}\rangle, \tag{2.11}$$

and

$$A_{j+1}^{-1/2} A_{j+1/2}^{1/2} \,|\, j+\tfrac{1}{2}\rangle = 0. \tag{2.12}$$

Thus l has the two values j and $(j+\frac{1}{2})$.

B. $j = 0$. In this case we obtain from Eqs. (2.4) and (2.5)

$$A_{1/2}^{-1/2} A_0^{1/2} \,|\, 0\rangle = 2K_2 \,|\, 0\rangle,$$

$$A_0^{1/2} A_{1/2}^{-1/2} \,|\, \tfrac{1}{2}\rangle = -\tfrac{1}{2}(2K_2 - 3)\,|\, \tfrac{1}{2}\rangle$$

and $A_1^{-1/2} A_{1/2}^{1/2} \,|\, \frac{1}{2}\rangle = 3(2K_2 - 1)\,|\, \frac{1}{2}\rangle$. However, using Eqs. (2.3), we see that either $\langle 0|K_2|0\rangle = 0$ or $\frac{1}{2}$, and either $\langle \frac{1}{2}\,|K_2|\,\frac{1}{2}\rangle = \frac{1}{2}$ or $\frac{3}{2}$. Since K_2 has the same value for both states, we see that $K_2 = \frac{1}{2}$, in which case $A_1^{-1/2} A_{1/2}^{1/2} \,|\, \frac{1}{2}\rangle = 0$, as we should expect since the I.R. must be dispin. However, $j = 0$, $K_2 = \frac{1}{2}$ is just a special case of A, so case B gives nothing new.

The actual permissible values of j depend on the star or gradestar conditions imposed on Osp(2,1), and this we consider in the following sections.

3. FINITE DIMENSIONAL REPRESENTATIONS OF Osp(2,1)

We consider in this section those I.R. of Osp(2,1) satisfying star or gradestar conditions which on the even part of the algebra reduce to the usual Hermiticity conditions for SU(2), i.e., we require $l_0^\dagger = l_0$ and $l_\pm^\dagger = l_\mp$. First consider star (Hermiticity) conditions on the odd part of Osp(2,1), i.e., if B is an odd operator and $\langle x \mid y\rangle$ is a positive definite bilinear form on the representation space, then $\langle B^\dagger x \mid y\rangle = \langle x \mid By\rangle$, and $(B^\dagger)^\dagger = B$. Now it is well known[9] that when, as in this case, one has an even dimensional tensor representation of SU(2), it is not possible to close the tensor representation with respect to the star operation, i.e., for our case, no complex, a, b, c, d can be found such that

$$q_{1/2}^\dagger = aq_{1/2} + bq_{-1/2} \cdot q_{-1/2}^\dagger = cq_{1/2} + dq_{-1/2} ,$$

is compatible with the SU(2) Hermiticity conditions and the commutation relations in Eq. (1.1). Hence no star representations of Osp(2,1) can occur when the even part is taken to be SU(2).

One can, however, have gradestar representations. Let again $\langle x \mid y\rangle$ be a positive definite bilinear form on the representation space and A an element of Osp(2,1). Then its gradestar adjoint A is defined by[2]

$$\langle A^{+}x \mid y\rangle = (-1)^{\alpha\xi}\langle x \mid Ay\rangle , \tag{3.1}$$

so that

$$(AB)^{+} = (-1)^{\alpha\beta}B^{+}A^{+}, \quad (A^{+})^{+} = (-1)^{\alpha}A, \tag{3.2}$$

where α, β, ξ, are the degrees of A, B and $|x\rangle$, respectively, i.e., $\alpha = 1$ if A is odd, $\alpha = 0$ if A is even, etc. In this section we still require $l_0^{+} = l_0$, $l_{\pm}^{+} = l_{\mp}$; one can show that the odd part can be closed in selfconsistent manner with respect to the gradestar operation if and only if

$$q_{1/2}^{+} = bq_{-1/2}, \quad q_{-1/2}^{+} = -bq_{-1/2},$$

where b is either 1 or -1.

We now investigate what restrictions are imposed on the I.R. by these star conditions. First of all, the conditions on the even operators require that l, and therefore j, be nonnegative half-integral, i.e., $j = 0, \frac{1}{2}, 1, \ldots$. [Actually, $j = -\frac{3}{2}$, -2, $-\frac{5}{2}, \ldots$ are also permissible but equivalent to the above due to the invariance of $l^2 = l(l+1)$ under $l \to -(l+1)$.] To proceed further we need the following Hermiticity properties of the shift operators, which can easily be worked out using results of Hughes and Yadegar[5]:

$$(O^{1/2,1/2})^{\pm}(2R+1) = -b(O^{-1/2,-1/2})(2R), \tag{3.3}$$

$$(O^{-1/2,-1/2})^{\pm}(2R+1) = b(O^{1/2,1/2})(2R+2). \tag{3.4}$$

From these equations it is easy to show that

$$\langle l \pm \tfrac{1}{2} \mid A_{l}^{\pm 1/2} \mid l\rangle = \mp b\frac{(2l+1)}{(2l+1\pm 1)} \times \langle l \pm \tfrac{1}{2} \mid (A_{l\pm 1/2}^{\mp 1/2})^{\pm} \mid l\rangle, \tag{3.5}$$

and

$$\langle l \mid A_{l\pm 1/2}^{\mp 1/2} A_{l}^{\pm 1/2} \mid l\rangle = \pm b(-1)^{\xi}\frac{(2l+1\pm 1)}{(2l+1)} \mid \langle l \pm \tfrac{1}{2} \mid A_{l}^{\pm 1/2} \mid l\rangle \mid^2, \tag{3.6}$$

where ξ is the degree of the states $|l \pm \frac{1}{2}\rangle$.

Now recall that l has only two values within an I.R., namely $l = j$ or $(j + \frac{1}{2})$, where j is a nonnegative half-integer. Let the $l = j$ state have degree α; thus using Eqs. (2.10) and (3.6) we obtain

$$\mid \langle j + \tfrac{1}{2} \mid A_{j}^{1/2} \mid j\rangle \mid^2 = b(-1)^{\alpha+1}(2j+1)^2/2. \tag{3.7}$$

This shows that $b(-1)^{\alpha+1} = 1$, so if $b = 1$, the $l = j$ state must be odd in order to preserve the star conditions; if $b = -1$, then the $l = j$ state must be even.

Using (3.7), it is now easy to show that, with a suitable choice of relative phase for the states,

$$q_{1/2} \mid j,m\rangle = \left[\frac{(j+m+1)}{2}\right]^{1/2} \mid j+\tfrac{1}{2}, m+\tfrac{1}{2}\rangle,$$

$$q_{-1/2} \mid j,m\rangle = \left[\frac{(j-m+1)}{2}\right]^{1/2} \mid j+\tfrac{1}{2}, m-\tfrac{1}{2}\rangle, \tag{3.8}$$

$$q_{1/2} \mid j+\tfrac{1}{2}, m\rangle = \left[\frac{(j-m+(1/2))}{2}\right]^{-1/2} \mid j, m+\tfrac{1}{2}\rangle,$$

$$q_{-1/2} \mid j+\tfrac{1}{2}, m\rangle = -\left[\frac{(j+m+(1/2))}{2}\right]^{1/2} \mid j, m-\tfrac{1}{2}\rangle.$$

The results obtained here are in complete agreement with those obtained by Nahm, Rittenberg, and Scheunert.[3]

4. INFINITE DIMENSIONAL REPRESENTATIONS OF Osp(2,1)

We now consider I.R. of Osp(2,1) for which the star (or gradestar) conditions on the even part of the algebra are those appropriate to the noncompact SU(1,1) [or Sl(2,R), or Sp(2)] algebra, i.e., we require $l_0^{\dagger} = l_0$, $l_{\pm}^{\dagger} = -l_{\mp}$. The situation here is the reverse of that of the last section, in that now no gradestar representations can arise; it is a straightforward matter to check this and so we omit the details here. On the other hand, the star conditions $q_{\pm 1/2}^{\dagger} = bq_{\mp 1/2}$ are easily seen to be compatible with the commutation relations and $l^{\dagger} = -l_{\mp}$ provided $b = \pm 1$. We shall give detailed analysis for the case $b = -1$, and just state the results for $b = +1$.

For the case where $q_{\pm 1/2}^{\dagger} = -q_{\mp 1/2}$, one obtains the following analogs of Eqs. (3.3) and (3.4):

$$(O^{1/2,1/2})^{\dagger}(2R+1) = (O^{-1/2,-1/2})(2R), \tag{4.1}$$

$$O^{-1/2,-1/2})^{\dagger}(2R+1) = (O^{1/2,1/2})(2R+2). \tag{4.2}$$

From these one obtains

$$\langle l \pm \tfrac{1}{2}, m \pm \tfrac{1}{2} \mid O_{l,m}^{\pm 1/2,\pm 1/2} \mid l,m\rangle = \frac{(2l+1)}{(2l+1\pm 1)}\langle l \pm \tfrac{1}{2}, m \pm \tfrac{1}{2} \mid (O_{l\pm 1/2,m\pm 1/2}^{\mp 1/2,\mp 1/2})^{\dagger} \mid l,m\rangle \tag{4.3}$$

and

$$\langle l,m \mid O_{l\pm\frac{1}{2},m\pm\frac{1}{2}}^{\mp 1/2,\mp 1/2} O_{l,m}^{\pm 1/2,\pm 1/2} \mid l,m\rangle = \frac{(2l+1\pm 1)}{(2l+1)} \mid \langle l \pm \tfrac{1}{2}, m \pm \tfrac{1}{2} \mid O_{l,m}^{\pm 1/2,\pm 1/2} \mid l,m\rangle \mid^2. \tag{4.4}$$

The reason that, unlike in the last section, we give relations for the $O^{\pm 1/2,\pm 1/2}$ rather than the normalized $A^{\pm 1/2}$ here is because since we are now dealing with the SU(1,1) subalgebra, whose I.R.'s are somewhat more complicated than those of SU(2), we need to take more careful account of the internal structure of these I.R.

From (4.3) and (4.4), we see that $O^{-1/2,-1/2}O^{1/2,1/2}$ is positive for $l < -1$ and for $l > -\frac{1}{2}$, but negative for $-1 < l < -\frac{1}{2}$. On the other hand, $O^{1/2,1/2}O^{-1/2,-1/2}$ is positive for $l < -\frac{1}{2}$ and for $l > 0$, but negative for $-\frac{1}{2} < l < 0$.

Before proceeding, we give a brief summary of the I.R. of SU(1,1). These fall into four classes (apart from the trivial representation): the principal series D^{P} for which $l = -\frac{1}{2} + i\rho$, where ρ is an arbitrary real number and where m takes on an infinity of values differing from one another by integral amounts; the supplementary series D^{S} for which

$-1<l<0$ and m again has an infinity of values; the positive discrete series D^+ for which l is an arbitrary real number and m has a minimum but no maximum value, and finally the negative discrete series D^- for which again l is an arbitrary real number but now m has a maximum but no minimum value. Note that in the case of $D^\pm$, since the eigenvalue $l(l+1)$ of L^2 is unchanged by the replacement $l\to -(l+1)$, there is symmetry about $l=-\frac{1}{2}$, i.e., for every I.R. corresponding to an l value $>-\frac{1}{2}$, there is an equivalent I.R. for which $l<-\frac{1}{2}$. One could therefore restrict l to either of these ranges with no loss of generality, but usually one takes $l<-\frac{1}{2}$ to distinquish these I.R. from the I.R. of SU(2) for which one conventionally takes $l\geqslant 0$. Here we shall actually find it more convenient to take $l<0$ rather than $l<-\frac{1}{2}$ for reasons that will become apparent.

To return to the I.R. of Osp(2,1), recall that l has two values, j and $j+\frac{1}{2}$. Since if $j=-\frac{1}{2}+i\rho$ then $j+\frac{1}{2}=i\rho$, which is not a permissible value of l for an I.R. of SU(1,1), i.e., which violates the star conditions on $l_\pm$, we see that the principal series D^P of SU(1,1) cannot occur. To see which of D^S or $D^\pm$ occur, we use the positivity and negativity conditions for the $O^{\mp 1/2,\mp 1/2}O^{\pm 1/2,\pm 1/2}$ stated above.

Now we saw in Sec. 2 that

$$\begin{aligned}\langle j,m|O^{-1/2,-1/2}O^{1/2,1/2}|j,m\rangle &= \langle j+\tfrac{1}{2},m+\tfrac{1}{2}|O^{1/2,1/2}O^{-1/2,-1/2}|j+\tfrac{1}{2},m+\tfrac{1}{2}\rangle \\ &= (j+m+1)(j+1)(2j+1). \end{aligned} \tag{4.5}$$

Using (4.5) together with the positivity and negativity conditions on the operator, we see that the star conditions are always obeyed if $(j+m+1)\geqslant 0$ and never obeyed if $(j+m+1)<0$. Hence for the $l=j$ state we must have $m\geqslant -(j+1)$, so m has a minimum value and hence the I.R. of SU(1,1) which occur are the positive discrete D^+ representations.

Choosing the relative phases of the states appropriately we obtain from Eqs. (4.4) and (4.5):

$$O^{1/2,1/2}|j,m\rangle = (2j+1)\left[\frac{(j+m+1)}{2}\right]^{1/2}|j+\tfrac{1}{2},m+\tfrac{1}{2}\rangle, \tag{4.6}$$

$$O^{-1/2,-1/2}|j+\tfrac{1}{2},m+\tfrac{1}{2}\rangle = (j+1)[2(j+m+1)]^{1/2}|j,m\rangle\,. \tag{4.7}$$

From these one obtains the following actions of $q_{\pm 1/2}$, together with those of $l_\pm$:

$$q_{1/2}|j,m\rangle = \left[\frac{(j+m+1)}{2}\right]^{1/2}|j+\tfrac{1}{2},m+\tfrac{1}{2}\rangle, \tag{4.8}$$

$$q_{-1/2}|j,m\rangle = \left[\frac{-(j-m+1)}{2}\right]^{1/2}|j+\tfrac{1}{2},m-\tfrac{1}{2}\rangle, \tag{4.9}$$

$$q_{1/2}|j+\tfrac{1}{2},m+\tfrac{1}{2}\rangle = \left[\frac{-(j-m)}{2}\right]^{1/2}|j,m+1\rangle, \tag{4.10}$$

$$q_{-1/2}|j+\tfrac{1}{2},m+\tfrac{1}{2}\rangle = -\left[\frac{(j+m+1)}{2}\right]^{1/2}|j,m\rangle, \tag{4.11}$$

$$l_+|l,m\rangle = [-(l-m)(l+m+1)]^{1/2}|l,m+1\rangle, \tag{4.12}$$

$$l_-|l,m\rangle = -[-(l+m)(l-m+1)]^{1/2}|l,m-1\rangle, \tag{4.13}$$

where in the last two equations $l=j$ or $j+\frac{1}{2}$.

The question as to what is the minimum m value for the two D^+ occuring in the above $l=j, j+\frac{1}{2}$ I.R. of Osp(2,1) needs to be answered with care, and we need to consider again the above-mentioned invariance of the set of I.R. of SU(1,1) under the replacement $l\to -(l+1)$. When $l\geqslant 0$, in order to satisfy the SU(1,1) star conditions $l_-l_+ = (l-m)(l+m+1)\leqslant 0$ and $l_+l_- = (l+m)(l-m+1)\leqslant 0$, we see from Eq. (4.13) that m has the minimum value $\underline{m}=(l+1)$. Similarly if $l\leqslant 1$, the star conditions require $\underline{m}=-l$, and for any D^+ corresponding to an $l>0$, the replacement $l\to -(l+1)$ yields an equivalent D^+ corresponding to an $l\leqslant -1$.

When $-1<l<0$, the situation is more complicated. If $-\frac{1}{2}<l<0$, then we still get a D^+ with $\underline{m}=(l+1)$, but now, as can easily be checked, the star conditions are also satisfied if $\underline{m}=-l$. Similarly, if $-1<l<-\frac{1}{2}$, we still get a D^+ with $\underline{m}=-l$, which is equivalent to a D^+ with $-\frac{1}{2}<l<0$ and $\underline{m}=(l+1)$, but we can now also have a D^+ with $\underline{m}=(l+1)$ which is equivalent to a D^+ with $-\frac{1}{2}<l<0$ and $\underline{m}=-l$. Hence we get two, not just one, inequivalent D^+ for any l where $-1<l<0$.

Thus, for the I.R. $l=j, j+\frac{1}{2}$ of Osp(2,1), we see that if $j\leqslant -1$, then there is a D^+ with $l=j$ and $\underline{m}=-j$ coupled by the $q_{\pm 1/2}$ to a D^+ with $l=j+\frac{1}{2}$ and $\underline{m}=-(j+\frac{1}{2})$. However, if $-1<j<-\frac{1}{2}$, we can either have an I.R. of Osp(2,1) containing a D^+ with $l=j$ and $\underline{m}=-j$ coupled to a D^+ with $l=j+\frac{1}{2}$ and $\underline{m}=-(j+\frac{1}{2})$, or we can have an I.R. of Osp(2,1) containing a D^+ with $l=j$ and $\underline{m}=(j+1)$ coupled to a D^+ with $l=j+\frac{1}{2}$ and $\underline{m}=(j+\frac{3}{2})$; thus for any $j\leqslant -1$ there is just one I.R. of Osp(2,1), but for a j with $-1<j<-\frac{1}{2}$, there are two inequivalent I.R.'s of Osp(2,1).

There is just one exception to this, nemaly the case when $j=-\frac{3}{4}$, so $l=-\frac{3}{4}, -\frac{1}{4}$. In this case $l(l+1)=-\frac{3}{16}$ for both l values since they are symmetrically spaced either side of $-\frac{1}{2}$, so the representation of Osp(2,1) where $l=-\frac{3}{4}$, $\underline{m}=\frac{3}{4}$ and $l=-\frac{1}{4}$, $\underline{m}=\frac{1}{4}$ is in fact equivalent to the representation where $l=-\frac{3}{4}$, $\underline{m}=\frac{1}{4}$ and $l=-\frac{1}{4}$, $\underline{m}=\frac{3}{4}$. This is the metaplectic representation which we shall consider in more detail in the following section.

Finally in the section we summarise the I.R. obtained for the star condition $q^\dagger_{\pm 1/2}=q_{\mp 1/2}$. In this case the star condition obtained for $O^{\mp 1/2,\mp 1/2}O^{\pm 1/2,\pm 1/2}$ require that any $|j,m\rangle$ state of the I.R. must have $m\leqslant -(j+1)$, so here m has a maximum value and the negative discrete representations D^- of SU(1,1) occur. The action of $q_{\pm 1/2}$ on $|j,m\rangle$ and $|j+\frac{1}{2},m+\frac{1}{2}\rangle$ for this case are

$$q_{1/2}|j,m\rangle = \left[\frac{-(j+m+1)}{2}\right]^{1/2}|j+\tfrac{1}{2},m+\tfrac{1}{2}\rangle, \tag{4.14}$$

$$q_{-1/2}|j,m\rangle = \left[\frac{(j-m+1)}{2}\right]^{1/2}|j+\tfrac{1}{2},m-\tfrac{1}{2}\rangle, \tag{4.15}$$

$$q_{1/2}|j+\tfrac{1}{2},m+\tfrac{1}{2}\rangle = \left[\frac{(j-m)}{2}\right]^{1/2}|j,m+1\rangle, \tag{4.16}$$

$$q_{-1/2}\,|j+\tfrac{1}{2},m+\tfrac{1}{2}\rangle = \left[-\frac{(j+m+1)}{2}\right]^{1/2}|j,m\rangle, \tag{4.17}$$

and the actions of $l_{\pm}$ are as given in Eqs. (4.12) and (4.13).

If $j \leqslant -1$, then there is a unique I.R. of Osp(2.1) with a D^- with $l=j$ and $\bar{m}=j$ coupled by the $q_{\pm 1/2}$ to a D^- with $l=j+\frac{1}{2}$ and $\bar{m}=j+\frac{1}{2}$. If $-1<j<-\frac{1}{2}$, there is in addition an I.R. of Osp(2,1) containing a D^- with $l=j$ and $\bar{m}=-(j+1)$ coupled to a D^- with $l=j+\frac{1}{2}$ and $\bar{m}=-(j+\frac{3}{2})$, except for the case where $j=-\frac{3}{4}$ when the two I.R.'s of Osp(2,1) so obtained are in fact equivalent, this being the counterpart for $q^{\dagger}_{\pm 1/2}=q_{\mp 1/2}$ of the metaplectic representation discussed above and in the following section.

5. THE METAPLECTIC REPRESENTATION

In this section we shall consider in more detail the $j=-\frac{3}{4}$ I.R. of Osp(2,1), which we refer to as the metaplectic representation since on restriction to SU(1,1) it yields the metaplectic representation of that algebra.[8] The states of the representation are $|l,m\rangle = |-\frac{3}{4},\frac{3}{4}+n\rangle, |-\frac{1}{4},\frac{1}{4}+n\rangle$ where $n=0,1,2,\ldots$, and the actions of the elements of Osp(2,1) on these states are given by

$$q_{1/2}\,|-\tfrac{3}{4},\tfrac{3}{4}+n\rangle = \left[\frac{n+1}{2}\right]^{1/2}|-\tfrac{1}{4},\tfrac{1}{4}+(n+1)\rangle, \tag{5.1}$$

$$q_{-1/2}\,|-\tfrac{3}{4},\tfrac{3}{4}+n\rangle = -\left[\frac{2n+1}{4}\right]^{1/2}|-\tfrac{1}{4},\tfrac{1}{4}+n\rangle, \tag{5.2}$$

$$q_{1/2}\,|-\tfrac{1}{4},\tfrac{1}{4}+n\rangle = \left[\frac{2n+1}{4}\right]^{1/2}|-\tfrac{3}{4},\tfrac{3}{4}+n\rangle, \tag{5.3}$$

$$q_{-1/2}\,|-\tfrac{1}{4},\tfrac{1}{4}+n\rangle = -\left[\frac{n}{2}\right]^{1/2}|-\tfrac{3}{4},\tfrac{3}{4}+(n-1)\rangle, \tag{5.4}$$

$$l_{+}\,|-\tfrac{3}{4},\tfrac{3}{4}+n\rangle = \left[\frac{(n+1)(2n+3)}{2}\right]^{1/2}|-\tfrac{3}{4},\tfrac{3}{4}+(n+1)\rangle, \tag{5.5}$$

$$l_{-}\,|-\tfrac{3}{4},\tfrac{3}{4}+n\rangle = -\left[\frac{n(2n+1)}{2}\right]^{1/2}|-\tfrac{3}{4},\tfrac{3}{4}+(n-1)\rangle, \tag{5.6}$$

$$l_{+}\,|-\tfrac{1}{4},\tfrac{1}{4}+n\rangle = \left[\frac{(n+1)(2n+1)}{2}\right]^{1/2}|-\tfrac{1}{4},\tfrac{1}{4}+(n+1)\rangle, \tag{5.7}$$

$$l_{-}\,|-\tfrac{1}{4},\tfrac{1}{4}+n\rangle = -\left[\frac{n(2n-1)}{2}\right]^{1/2}|-\tfrac{1}{4},\tfrac{1}{4}+(n-1)\rangle. \tag{5.8}$$

The metaplectic representation of SU(1,1), has been considered, explicitly or implicitly, by many authors,[10–12] the most extensive work being due to Sternberg and Wolf.[8] Let z be a complex variable; then the above operators can be realized as

$$l_0=\tfrac{1}{4}z(d/dz)+\tfrac{1}{4}I, \quad l_{+}=(i/2)z^2,$$
$$l_{-}=(i/2)(d^2/dz^2), \tag{5.9}$$
$$q_{1/2}=\tfrac{1}{2}e^{-3i\pi/4}z, \quad q_{-1/2}=\tfrac{1}{2}e^{-i\pi/4}(d/dz). \tag{5.10}$$

A simple calculation verifies that $L^2=-\frac{3}{16}, K^2=-\frac{1}{16}$ [incidentally this is the minimum possible value of K_2 for any star representation of Osp(2,1)], so only the metaplectic representation can be realized in this way.

The states of the representation can be realized in terms of powers of z as follows:

$$|-\tfrac{3}{4},\tfrac{3}{4}+n\rangle = \frac{i^n z^{2n+1}}{[2\pi(2n+1)!]^{1/2}}, \tag{5.11}$$

$$|-\tfrac{1}{4},\tfrac{1}{4}+n\rangle = \frac{i^{(n+3/2)}z^{2n}}{[2\pi(2n)!]^{1/2}}, \tag{5.12}$$

where the constants of proportionality have been chosen to exactly correspond to the actions of the operators as given by Eq. (5.1)–(5.8).

In this realization, the star conditions $l_0^{\dagger}=l_0$, $l_{\pm}^{\dagger}=-l_{\pm}$, $q^{\dagger}_{\pm 1/2}=-q_{\mp 1/2}$ reduce to the Fock condition $(d/dz)^{\dagger}=z$. The inner product on the ring of polynomials in z for which the Fock condition is satisfied, and with respect to which the states $|-\frac{3}{4},\frac{3}{4}+n\rangle, |-\frac{1}{4},\frac{1}{4}+n\rangle$, $n=0,1,2,\ldots$, are orthonormal, is given by

$$\left(\sum_{s=0}^{\infty}a_s z^s, \sum_{t=0}^{\infty}b_t z^t\right) = 2\pi\sum_{s=0}^{\infty}s!\,\bar{a}_s b_s. \tag{5.13}$$

This inner product can be expressed as an integral in more ways than one. Sternberg and Wolf[8] realize the metaplectic representation in terms of the space $\mathscr{H}$ of all holomorphic $f:\mathbb{C}\to\mathbb{C}$ such that $\int |f(z)|^2 \exp(-|z|^2)\,d\lambda(z)<\infty$, where λ is Lebesque measure on $\mathbb{C}$. Then the inner product which corresponds to (5.13) can be expressed as

$$(f,g)=2\int f(z)\overline{g(z)}\exp(-|z|^2)\,d\lambda(z). \tag{5.14}$$

An alternative method of expressing the inner product is as follows. Let Γ be the unit circle with origin as center in $\mathbb{C}$, and for any $f:\Gamma\to\Gamma$, denote by L_f the Laplace transform of f. Denote by $\mathscr{H}'$ the space of all f such that $\oint_{\Gamma}\overline{L_f(\bar{z})}f(z)\,dz<\infty$. Then the inner product on $\mathscr{H}'$ which corresponds to (5.13) is

$$(f,g)=-i\oint_{\Gamma}\overline{L_f(\bar{z})}\,g(z)\,dz. \tag{5.15}$$

Clearly in this definition of $\mathscr{H}'$ and the inner product, Γ could be replaced by any smooth contour in $\mathbb{C}$ which encircles the origin and does not intersect itself.

We see that the two component I.R.'s of SU(1,1) contained in the metaplectic representation of Osp(2,1) are supported by the subspaces of $\mathscr{H}$ consisting of, respectively, even and odd functions of z. This is just a particular example of the more general result of Sternberg and Wolf[8] for the metaplectic representation of Sp(n) quoted in the introduction to this paper.

In a following paper the star representations of the central extension of the semi-direct product Lie algebra $SU(1,1)_A T^2$ will be classified. There we shall find one I.R. which can also be realized in terms of the metaplectic representation.

ACKNOWLEDGMENTS

The author would like to thank Dr. R.B. Jones of the Physics Department at Queen Mary College, and Dr. N.B.

Backhouse of the Department of Applied Mathematics and Theoretical Physics at Liverpool University for helpful discussions.

[1]A. Pais and V. Rittenberg, J. Math. Phys. **16**, 2062 (1975).
[2]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146 (1977).
[3]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977).
[4]L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975).
[5]J.W.B. Hughes and J. Yadegar, J. Math. Phys. **19**, 2068 (1978).
[6]N.B. Backhouse and J.W.B. Hughes, J. Math. Phys. **20**, 1545 (1979).
[7]A.O. Barut and C. Fronsdal, Proc. R. Soc. A **289**, 532 (1965).
[8]S. Sternberg and J.A. Wolf, Trans. Am. Math. Soc. **238**, 1 (1978).
[9]A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, New Jersey, 1957).
[10]V. Bargmann, *Analytic Methods in Mathematical Physics* (Gordon and Breach, London, 1970), pp. 27–63.
[11]C. Itzykson, Commun. Math. Phys. **4**, 92 (1967).
[12]D. Shale, Trans. Am. Math. Soc. **103**, 149 (1962).

Subgroups of Lie groups and separation of variables [a)]

W. Miller, Jr.
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

J. Patera and P. Winternitz
Centre de recherches mathématiques, Université de Montréal, Montréal, Québec, Canada

(Received 4 December 1978; accepted for publication 11 September 1979)

Separable systems of coordinates for the Helmholtz equation $\Delta_d \Psi = E\Psi$ in pseudo-Riemannian spaces of dimension d have previously been characterized algebraically in terms of sets of commuting second order symmetry operators for the operator Δ_d. They have also been characterized geometrically by the form that the metric $ds^2 = g_{ik}(x)dx^i dx^k$ can take. We complement these characterizations by a group theoretical one in which the second order operators are related to continuous and discrete subgroups of G, the symmetry group of Δ_d. For $d = 3$ we study all separable coordinates that can be characterized in terms of the Lie algebra L of G and show that they are of eight types, seven of which are related to the subgroup structure of G. Our method clearly generalizes to the case $d > 3$. Although each separable system corresponds to a pair of commuting symmetry operators, there do exist pairs of commuting symmetries S_1, S_2 that are not associated with separable coordinates. For subgroup related operators we show in detail just which symmetries S_1, S_2 fail to define separation and why this failure occurs.

PACS numbers: 02.20.Qs

I. INTRODUCTION

The purpose of this article is to investigate the relationship between separation of variables in the Helmholtz equation for a pseudo-Riemannian space and the subgroup structure of the invariance Lie group of the equation. The article thus brings together the results of three different research programs that have been actively pursued during the past few years. These are (i) a systematic algebraic approach to the separation of variables in p.d.e.[1-18]; (ii) the classification of Lie subgroups of Lie groups[1,19-28]; (iii) applications of discrete subgroups of Lie groups.[29-33]

Historically, the approach to separation of variables has been in terms of Riemannian and differential geometry.[34-38] In the algebraic approach[1-18] for a d-dimensional manifold the Helmholtz equation

$$\Delta_d \Psi = E\Psi(x) \tag{1.1}$$

is considered, where $x = (x_1, x_2, \ldots, x_d)$ is a local coordinate system and Δ_d is the Laplace–Beltrami operator on the manifold. It is assumed that Eq. (1.1) has the Lie symmetry group G. Its Lie algebra L consists of first order linear operators X satisfying $[\Delta_d, X] = 0$, and we choose a basis $\{X_1, \ldots X_n\}$ for L. Separable coordinates for Eq. (1.1) are associated with $(d-1)$-tuplets of commuting second-order symmetry operators $\{S_\alpha\}$ for Δ_d. A classification of the sets of operators $\{S_\alpha\}$ into orbits under the action of G provides a classification of separable systems of coordinates. The separable functions

$$\Psi(x) = \prod_{i=1}^{d} f_i(\xi_i) \tag{1.2}$$

are the common eigenfunctions of the operators Δ_d and S_α $(1 \leqslant \alpha \leqslant d-1)$.

There are some puzzling aspects to the algebraic approach. First of all, while there is a mechanical procedure for computing the symmetries $\{S_\alpha\}$ from a separable system of coordinates, the precise relationship between the $\{S_\alpha\}$ and the subgroup structure of G has remained unclear. Furthermore, there exist commuting symmetries $\{S_\alpha\}$ that do not correspond to any separable coordinates at all! The discovery of practical criteria to determine precisely which commuting symmetries lead to variable separation remains one of the most important problems in this theory. Here we show for $d = 3$ the relation between the subgroup structure of G and the coordinate systems yielding separation of variables for the Helmholtz equation on the manifold. (This analysis clearly generalizes to the case $d > 3$.) Furthermore, for subgroup related operators $\{S_\alpha\}$ we show in detail which symmetries fail to define variable separation and why this failure occurs.

Section 2 is devoted to the general theory. We show that separable coordinates fall into different classes, depending on how many of the operators in the set $\{S_\alpha\}$ are squares of the linear operators X (these correspond to Abelian subgroups of G), how many are invariant operators of nonAbelian Lie subgroups, and how many are invariants of discrete subgroups. In Sec. 3 we treat three-dimensional manifolds of constant curvature in some detail.

2. GENERAL THEORY

Let Δ_d be the Laplace–Beltrami operator on a d-dimensional pseudo-Riemannian manifold with metric $ds^2 = \sum_{i,j=1}^{d} g_{ij} dx^i dx^j$, i.e.,

$$\Delta_d \Psi = \sum_{ij} g^{-1/2} \partial_i (g^{1/2} g^{ij} \partial_j \psi), \tag{2.1}$$

[a)]Supported in part by the National Research Council of Canada, the Ministère de l'Education du Gouvernement du Québec and the National Science Foundation.

 0022-2488/81/020251-10$1.00

where $g = \det(g_{ij})$, $\partial_i = \partial/\partial x^i$, and $\Sigma_j g^{ij} g_{jk} = \delta^i_k$. The Helmholtz equation for this manifold is

$$\Delta_d \Psi = E\Psi, \tag{2.2}$$

where E is a nonzero constant. In Refs. 17 and 18 the possible coordinate systems that permit separation of variables for the Helmholtz equation have been classified in the cases $d = 2,3,4$. The classification of separable types is closely related to the symmetry algebra of Eq. (2.2). A first order symmetry operator X for Eq. (2.2) is an operator

$$X = \sum_{i=1}^{d} \xi_i(x^j)\partial_i, \tag{2.3}$$

such that $[X,\Delta_d] = 0$, where $[\cdot,\cdot]$ is the usual commutator of differential operators. (This is equivalent to the assertion that $\{\xi_i\}$ is a Killing vector.) The set of all first order symmetries of Eq. (2.2) forms a Lie algebra L with $\dim L \leqslant d(d+1)/2$. If $\{x^1,\ldots,x^d\}$ is a separable system for Eq. (2.2), we say the variable x^1 is *ignorable* provided $X = \partial_1 \in L$, i.e., provided the tensor g_{ij} in these coordinates is independent of x^1.

In this paper we restrict ourselves to the case $d = 3$. For $d = 3$ each separable system $\{x^1,x^2,x^3\}$ is characterized by a pair of second order differential operators $\{S_1,S_2\}$ such that

$$[S_1,S_2] = 0, \quad [S_j,\Delta_3] = 0, \quad j = 1,2. \tag{2.4}$$

Here the corresponding separable solutions $\psi = A(x^1)B(x^2)C(x^3)$ of Eq. (2.3) have the characterization

$$S_j\psi = \lambda_j\psi, \quad j = 1,2, \tag{2.5}$$

where the eigenvalues λ_j are separation constants.

As shown in Ref. 17 the separable systems are of eight distinct types: (I) Three ignorable variables:

$$ds^2 = (dx^1)^2 + (dx^2)^2 + \epsilon(dx^3)^2, \epsilon = \pm 1,$$
$$S_1 = \partial_1^2, S_2 = \partial_2^2. \tag{2.6}$$

Here, L contains a three-dimensional Abelian subalgebra generated by $L_j = \partial_j, j = 1,2,3$, and the manifold is flat. Note that the operator $S_3 = \partial_3^2$ is automatically diagonalized in this case. (II) Two ignorable variables:

$$ds^2 = \sum_{i,j=1}^{3} g_{ij}(x^3)dx^i dx^j, \quad S_1 = \partial_1^2, S_2 = \partial_2^2. \tag{2.7}$$

Here, L contains a two-dimensional Abelian subalgebra A generated by $L_j = \partial_j, j = 1,2$. The coordinates may be nonorthogonal. The subalgebra A must be maximal Abelian since otherwise the system would be type I. (III) One ignorable variable: This case splits into four subtypes, for each of which L contains the operator $L_1 = \partial_1$, and we have $S_1 = \partial_1^2$: (III_1) Centralizer coordinates (orthogonal):

$$ds^2 = (dx^1)^2 + (\sigma_2(x^2) + \sigma_3(x^3))[(dx^2)^2 + \epsilon(dx^3)^2], \tag{2.8}$$
$$S_2 = \frac{1}{\sigma_2 + \sigma_3}[\sigma_3\partial_2^2 - \epsilon\sigma_2\partial_3^2].$$

(III_2) Centralizer coordinates (nonorthogonal):

$$ds^2 = \sigma_2[\sigma_3(dx^2)^2 + 2dx^1dx^2 + (dx^3)^2], \tag{2.9}$$
$$S_2 = \partial_3^2 - \sigma_3\partial_1^2.$$

(III_3) Subgroup coordinates:

$$ds^2 = \sigma_3(dx^3)^2 + \sigma_3\sigma_2[(dx^1)^2 + \epsilon(dx^2)^2], \tag{2.10}$$
$$S_2 = \frac{1}{\sigma_2}(\partial_1^2 + \epsilon\partial_2^2).$$

(III_4) Generic type III coordinates:

$$ds^2 = (\sigma_2 + \sigma_3)[dx^2)^2 + \epsilon_1(dx^3)^2] + \epsilon_2\sigma_2\sigma_3(dx^1)^2,$$
$$S_2 = \epsilon_2\left(\frac{1}{\sigma_3} - \frac{1}{\sigma_2}\right)\partial_1^2 + \frac{1}{\sigma_2 + \sigma_3}(\sigma_3\partial_2^2 - \epsilon_1\sigma_2\partial_3^2) \tag{2.11}$$
$$+ \frac{1/2}{(\sigma_2 + \sigma_3)}\left(\frac{\sigma_3\sigma_2'}{\sigma_2}\partial_2 - \frac{\epsilon_1\sigma_2\sigma_3'}{\sigma_3}\partial_3\right), \quad \epsilon_j = \pm 1.$$

(IV) No ignorable variables: Here there are two subtypes: (IV_1) We have

$$ds^2 = \sigma_1^2(dx^1)^2 + \sigma_1(\sigma_2 + \sigma_3)[(dx^2)^2 + \epsilon(dx^3)^2], \tag{2.12}$$
$$S_1 = \frac{1}{\sigma_2 + \sigma_3}(\partial_2^2 + \epsilon\partial_3^2), S_2 = \frac{1}{\sigma_2 + \sigma_3}(\sigma_3\partial_2^2 - \epsilon\sigma_2\partial_3^2).$$

(IV_2) Generic coordinates:

$$ds^2 = (\sigma_1 - \sigma_2)(\sigma_1 - \sigma_3)(dx^1)^2 + (\sigma_2 - \sigma_1)(\sigma_2 - \sigma_3)(dx^2)^2$$
$$+ (\sigma_3 - \sigma_1)(\sigma_3 - \sigma_2)(dx^3)^2, \tag{2.13}$$
$$S_1 = \frac{\sigma_2 + \sigma_3}{(\sigma_1 - \sigma_2)(\sigma_3 - \sigma_1)}\partial_1^2 + \frac{\sigma_3 + \sigma_1}{(\sigma_2 - \sigma_3)(\sigma_1 - \sigma_2)}\partial_2^2$$
$$+ \frac{\epsilon(\sigma_1 + \sigma_2)}{(\sigma_3 - \sigma_1)(\sigma_2 - \sigma_3)}\partial_3^2,$$
$$S_2 = \frac{\sigma_2\sigma_3}{(\sigma_1 - \sigma_2)(\sigma_3 - \sigma_1)}\partial_1^2 + \frac{\sigma_3\sigma_1}{(\sigma_2 - \sigma_3)(\sigma_1 - \sigma_2)}\partial_2^2$$
$$+ \frac{\epsilon(\sigma_1\sigma_2)}{(\sigma_3 - \sigma_1)(\sigma_2 - \sigma_3)}\partial_3^2.$$

In all of the above expressions $\sigma_j = \sigma_j(x^j)$. We refer to systems III_4 and IV_2 as "generic" since all other systems of types III and IV are degenerate cases of these two. It is only for Minkowski space $E_{2,1}$ that all eight separable types actually occur. As shown in Ref. 17, types I, III_1, and III_2 do not appear for space of nonzero constant curvature.

In this paper we are concerned with a purely group theoretic characterization of the various separation types. To successfully characterize a separable system $\{x^j\}$ for Eq. (2.2) in terms of the symmetry algebra L it is necessary that the defining operators S_1,S_2 for the system belong to the enveloping algebra of L. If this is so, we say that the coordinates $\{x^j\}$ are of *class I*; otherwise they are of *class II*. Reference 17 contains a derivation of all class I coordinates for all types except IV_2.

We now describe a general group theoretic procedure for characterizing all class I coordinates associated with the Helmholtz equation on a three-dimensional Riemannian manifold with symmetry algebra L. The validity of this procedure will be demonstrated using the results of Ref. 17 but will also be illustrated by examples in 3. The procedure is as follows:

First we determine if L contains a maximal Abelian subalgebra of dimension 3. This will be the case if and only if the manifold is flat and corresponds to type I (Cartesian) coordinates. Then we find the (conjugacy classes of) maxi-

mal Abelian subalgebras of dimension 2. Each such subalgebra determines a type II system.

Next we determine the conjugacy classes of one dimensional subalgebras of L. Let X be a representative from such a class and let cent(X) be the centralizer of X in L. There are four possibilities:

Type III_1: cent(X) = $\{X\} \oplus$ (cent(X)/$\{X\}$), cent (X) non-Abelian: (2.14)
Let L_X = (cent(X)/$\{X\}$) and decompose the space of second order elements in the enveloping algebra of L_X into orbits under the action of the normalizer Nor(X) of X in G. Every type III_1 system with $S_1 = X^2$ has the property that S_2 is a representative from one of these orbits. Two representatives from the same orbit correspond to equivalent coordinates.

Type III_2: cent(X) $\neq \{X\} \oplus$ (cent(X)/$\{X\}$), cent(X) non-Abelian: (2.15)
Decompose the space of second order elements in the enveloping algebra of cent(X) into orbits under the action of Nor(X). Every type III_2 system with $S_1 = X^2$ has the property that S_2 is a representative from one of these orbits. Class I coordinates of this type arise only for flat space.

Type III_3: Subgroup type coordinates: (2.16)
Given the one-dimensional subalgebra X, find all subalgebras A of L such that (1) $A \supset X$ (properly), (2) A is non-Abelian, and (3) A has a second order Casimir operator S_2, not equal to Δ_3 or to a linear combination of Δ_3 and the square of an element of L. Every type III_3 system is of the form $S_1 = X^2$, S_2.

Type III_4: Generic type III coordinates: (2.17)
Let X be as above and determine the space S of all second order elements Y in the enveloping algebra of L such that $[X,Y] = 0$. Decompose S into orbits under the adjoint action of Nor(X) and let S_2 be a representative from such an orbit. Every type III_4 system is of the form $S_1 = X^2, S_2$ such that this commuting pair has not already been included under types I–III_3 listed above.

The remaining two types characterize all pairs S_1, S_2 for which neither operator is a perfect square:

Type IV_1: Semisubgroup coordinates: (2.18)
Consider the three-dimensional subalgebras A of L with properties (2) and (3) discussed above in III_3. Take S_1 to be the Casimir operator of such an A and S_2 to be a second order element in the envoloping algebra of A. (Operators S_2 and S_2' are considered equivalent if they lie on the same orbit under the adjoint action of the maximal group of symmetries whose Lie algebra is A.) Every type IV_1 system is of the form S_1, S_2.

Type IV_2: Generic coordinates: (2.19)
This is the generic case. Here S_1, S_2 are simply a pair of commuting second order symmetries in the enveloping algebra of L, classified into orbits under the action of the symmetry group G, and such that this pair has not already been included under types I–IV_1 above.

For types I, II, and III_3 both operators S_1 and S_2 are invariants of Lie subgroups of G. For III_1, III_2, III_4, and IV_1 only S_1 has this property; for IV_2 neither of the operators is directly related to a Lie subgroup. The group G also contains discrete subgroups and is itself not necessarily connected. We shall see below that those operators S_1 that are not invariants of Lie groups can be characterized by the fact that they occur as invariants of discrete subgroups of G.

Now we demonstrate the validity of our group theoretic classification of defining operators for class I separable coordinates on a three-dimensional Riemannian manifold. First we note that every orbit of two-dimensional vector spaces, each space composed of mutually commuting second order elements in the enveloping algebra of L, belongs to exactly one of the eight classes listed above. Thus, it will be sufficient for us to show that the defining operators S_1, S_2 corresponding to a class I separable system of a given type (2.6)–(2.18) themselves have the group theoretic characterization for the corresponding type listed above. For this we draw on the results of Ref. 17.

The group theoretic characterization of types I and II is obvious.

(III_1) Centralizer coordinates (orthogonal): It follows from the results of Sec. 5 in Ref. 17 that the separable system (2.8) is class I precisely when

$$ds^2 = (dx^1)^2 + d\omega^2(x^2,x^3),$$

where $d\omega^2$ is the metric for a two dimensional Riemannian space of constant curvature [with Lie algebra L' isomorphic to one of e(3), e(2,1), o(4), o(3,1), o(2,2)] and S_2 a second order element in the enveloping algebra of L' which is not a square. Here $L \supseteq \{X\} \oplus L'$, where $X = \partial_1$, so the pair S_1, S_2 is of the form (2.14).

(III_2) Centralizer coordinates (nonorthogonal): According to Ref. 17, coordinates (2.9) are class I only for flat space and the possibilities are listed in Sec. 4 of that paper. One can directly verify that in each case the operators S_1, S_2 are of the form (2.15).

(III_3) Subgroup coordinates: In Ref. 17 it is shown that coordinates (2.10) are class I precisely when

$$ds^2 = \sigma_3(dx^3)^2 + \sigma_3 d\omega^2(x^1,x^2),$$

where $d\omega^2$ is the metric for a two dimensional space of constant curvature, $X = \partial_1$ is a Lie symmetry of $d\omega^2$, and S_2 is the Laplace–Beltrami operator for this two-dimensional space. With $X = \partial_1, S_1 = X^2$ it follows that S_1, S_2 is of the form (2.16).

(III_4) Generic type III coordinates: According to Ref. 17 coordinates (2.11) are class I if and only if the manifold is a space of constant curvature. These coordinates cannot be type III_3 because, as is straightforward to verify for spaces of constant curvature, the subalgebras A in the definition of type III_3 must have Casimir operators that are Laplace–Beltrami operators on two-dimensional manifolds. The operator S_1 [Eq. (2.11)] is clearly not a Laplace–Beltrami operator. The coordinates cannot be type III_2 because among the symmetry algebras for spaces of constant curvature only e(2,1) contains an element X such that cent(X) $\neq \{X\} \oplus L_X$ and cent(X) is non-Abelian. For this case all corresponding orbits of operators S_2 in the enveloping algebra of cent(X) were computed in Ref. 17 and the coordinates were shown to be of the form (2.9). If the coordinates (2.11) were type III_1, then the manifold would be flat, because among the symme-

try algebras for constant curvature spaces, only e(3) and e(2,1) contain elements X such that $\mathrm{cent}(X) = \{X\} \oplus L_X$ with $\mathrm{cent}(X)$ non-Abelian. These cases are classified in the following section and shown to correspond to coordinates (2.8). Thus, class I coordinates (2.11) correspond to operators of the form (2.17).

(IV_1) Semisubgroup coordinates: It is shown in Ref. 17 that coordinates (2.12) are class I provided

$$ds^2 = \sigma_1^2(dx^1)^2 + \sigma_1 d\omega^2(x^2,x^3),$$

where $d\omega^2$ is the metric for a two-dimensional subspace of constant curvature. It is clear from Eq. (2.12) that S_1 is the Laplace–Beltrami operator on this subspace; hence, the Casimir operator for the symmetry algebra L' of the subspace, where $L' \subseteq L$. Since S_2 is defined on the subspace and commutes with S_1, it must be expressable in terms of second order elements in the enveloping algebra of L'. Thus, operators S_1,S_2 are of the form (2.18).

(IV_2) Generic coordinates: Class I coordinates (2.13) cannot be of operator types I–III since we can see by inspection that one cannot construct from a linear combination of S_1 and S_2 an operator which is a perfect square of a Lie symmetry. The operators cannot be of type IV_1 because the only possible choices for the algebra A are e(2), e(1,1), o(3), o(2,1) acting as transitive symmetry algebras on a two-dimensional submanifold. It follows in these cases that the Casimir operator of A is the Laplace–Beltrami operator on the submanifold, and hence that S_1,S_2 can be written in the form (2.12) for appropriate coordinates. Since a set of orthogonal separable coordinates is uniquely determined by its defining operators S_1,S_2 (see Ref. 34), these coordinates must be of the form (2.12), a contradiction. Hence, class I coordinates (2.13) correspond to operators (2.19).

The above results hold for all Riemannian manifolds admitting class I separable coordinates, and there are an infinite number of such manifolds. However, of special interest are the manifolds of constant curvature, since they have the property that all separable coordinates are class I. In the following section we shall study the symmetry algebra L of each of the three-dimensional constant curvature spaces to see in detail how the subalgebra structure of L corresponds to the separable coordinates I–IV_1. We provide a complete orbit analysis for all pairs of commuting operators that correspond to proper subalgebras of L, i.e., for all operator types except IV_2. In a number of cases we will uncover orbits of type III_4 operators that do not correspond to variable separation.

3. THREE-DIMENSIONAL SPACES OF CONSTANT CURVATURE

In this section we illustrate the general theory by considering all spaces of constant curvature.

A. Group E(3)

The algebra e(3) of the group E(3) is generated by the infinitesimal rotations L_i and translations P_i, satisfying the commutation relations

$$[L_i,L_k] = \epsilon_{ikl}L_l,\quad [L_i,P_k] = \epsilon_{ikl}P_l,\quad [P_i,P_k] = 0. \tag{3.1}$$

It has two Casimir operators, namely,

$$\Delta = \mathbf{P}^2 = P_1^2 + P_2^2 + P_3^2 \text{ and}$$
$$\Delta' = \mathbf{L}\cdot\mathbf{P} = L_1P_1 + L_2P_2 + L_3P_3. \tag{3.2}$$

For the representations considered here we have $\Delta' = 0$ (a space of scalar functions in Euclidean space).

The subalgebras of e(3) have been classified into orbits under the action of E(3), e.g., in Ref. 19, where the results are presented in a diagram. Let us use this classification to investigate different types of separable coordinates for the equation $\Delta\Psi = E\Psi$, following Sec. 2.

I. *Three ignorable variables*

The algebra e(3) has precisely one class of maximal Abelian subalgebras (MASA) of dimension 3 represented by $\{P_1,P_2,P_3\}$. This provides *Cartesian coordinates* for which

$$S_i = P_i^2,\quad i = 1,2,3;\ \Delta = S_1 + S_2 + S_3. \tag{3.3}$$

II. *Two ignorable variables*

The algebra e(3) has precisely one class of MASA of dimension 2, represented by $\{L_3,P_3\}$. This provides *cylindrical coordinates*, for which

$$S_1 = L_3^2,\quad S_2 = P_3^2. \tag{3.4}$$

III. *One ignorable variable*

To find coordinates of type III_1 and III_2 we must consider separately a representative X of each class of one-dimensional (Abelian nonmaximal) subalgebras and find its centralizer cent$X$ in e(3). We are only interested in non-Abelian centralizers. The only type of element of e(3) having a non-Abelian centralizer can be represented by P_3, where

$$\mathrm{cent}(P_3) = \{P_3\} \oplus \{L_3,P_1,P_2\}, \tag{3.5}$$

i.e., centP_3 splits into a direct sum of P_3 and $\{\mathrm{cent}(P_3)\}/\{P_3\}$. Hence, no III_2 type coordinates exist in this case. Type III_1 coordinates (orthogonal centralizer type coordinates) are obtained by putting

$$S_1 = P_3^2, \tag{3.6}$$

$$S_2 = aL_3^2 + b(L_3P_1 + P_1L_3) + c(L_3P_2 + P_2L_3) + d(P_1^2 - P_2^2) + 2e(P_1P_2) + f(P_1^2 + P_2^2), \tag{3.7}$$

i.e., S_2 is the most general symmetric second order operator in the enveloping algebra of $e(2) = \{L_3,P_1,P_2\}$. We must now classify the operators (3.7) into orbits under Nor(P_3), i.e., the normalizer of P_3 in E(3). This is a well-known problem.[1,15] These orbits can be represented by

$$P_1^2, L_3^2,\ L_3^2 + a(P_1^2 - P_2^2), (a > 0),\text{ and } L_3P_2 + P_2L_3. \tag{3.8}$$

The first two operators should be omitted, since they are squares of generators and lead back to the case I or II. The last two operators provide type III_1 coordinates, namely, *elliptic cylindrical* and *parabolic cylindrical* coordinates, respectively.

Type III_3 coordinates (subgroup type) are obtained by taking a representative X of each orbit of generators of e(3) and finding all proper subalgebras of e(3) that properly contain X, are non-Abelian, and have a second order Casimir operator, not equal to $\Delta = \mathbf{P}^2$ or to a linear combination of Δ

and the square of a generator. The only such chain of subalgebras is

$$e(3)\supset o(3)\supset o(2),$$

and we have

$$S_1 = L_3^2,\ S_2 = \mathbf{L}^2, \tag{3.9}$$

i.e., S_2 is the Casimir operator of o(3), providing *spherical* coordinates.

Type III_4 coordinates are obtained by running through all representative generators X, and for each X finding the most general second order operator S_2 in the enveloping algebra of e(3) satisfying $[X,S_2]=0$. We find a representative of each orbit and eliminate representatives already encountered, i.e., corresponding to squares of generators, members of the enveloping algebra of cent(X), or Casimir operators of subalgebras. Let us examine each case separately.

(i) $X = L$, nor $L_3 = \{L_3,P_3\}$, and

$$S_2 = a\mathbf{L}^2 + b(L_1P_2 + P_2L_1 - L_2P_1 - P_1L_2) + c(P_1^2 + P_2^2) + dL_3P_3 \tag{3.10}$$

[we have dropped the Casimir operator of e(3) from Eq. (3.10)]. Separable coordinates (u,v,Φ) of this type satisfy

$$L_3 \equiv \frac{\partial}{\partial\Phi} = x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x} = \frac{\partial x}{\partial\Phi}\frac{\partial}{\partial x} + \frac{\partial y}{\partial\Phi}\frac{\partial}{\partial y} + \frac{\partial z}{\partial\Phi}\frac{\partial}{\partial z}. \tag{3.11}$$

Hence, we have

$$\frac{\partial x}{\partial\Phi} = -y,\quad \frac{\partial y}{\partial\Phi} = x,\quad \frac{\partial z}{\partial\Phi} = 0. \tag{3.12}$$

The relations (3.12) imply

$$x = f(u,v)\cos\Phi,\ y = f(u,v)\sin\Phi,\ z = h(u,v). \tag{3.13}$$

The operators

$$S_1 = L_3^2 \text{ and } S_2$$

in Eq. (2.11) are invariant under the reflection $\Phi\to -\Phi$ (i.e., $y\to -y$). Since L_3P_3 does not have this invariance, property, we must put $d=0$ in Eq. (3.10), i.e., operator (3.10) with $d\neq 0$ does not correspond to variable separation. We can now use the translation $\exp\alpha P_3$, belonging to the normalizer of L_3 in E(3), to simplify S_2. For $a\neq 0$ we can reduce Eq. (3.10) to

$$S_2 = \mathbf{L}^2 + c(P_1^2 + P_2^2), \tag{3.14}$$

For $c>0$ and $c<0$ this corresponds to oblate and prolate spheriodal coordinates, respectively. If $a=0$, $b\neq 0$, we can reduce S_2 to

$$S_2 = L_1P_2 + P_2L_1 - L_2P_1 - P_1L_2, \tag{3.15}$$

corresponding to *parabolic* coordinates.

If $a=b=0$ we return to type II coordinates.

(ii) $X = P_3$; nor $P_3 = \{L_3,P_1,P_2,P_3\}$: We have

$$S_2 = aL_3^2 + b(L_3P_1 + P_1L_3) + c(L_3P_2 + P_2L_3) + dL_3P_3 + c_{ik}P_iP_k. \tag{3.16}$$

The coordinates (u,v,x_3) satisfy

$$P_3 \equiv \frac{\partial}{\partial x_3} = \frac{\partial x}{\partial x_3}\frac{\partial}{\partial x} + \frac{\partial y}{\partial x_3}\frac{\partial}{\partial y} + \frac{\partial z}{\partial x_3}\frac{\partial}{\partial z}.$$

Hence,

$$x = x(u,v),\quad y = y(u,v),\quad z = x_3,$$

and L_3P_3 changes sign under the reflection $z\to -z$. Hence, $d=0$ in Eq. (3.16) in order to yield variable separation. Similarly, $c_{i3} = c_{3i} = 0$.

If $a\neq 0$, we use the normalizer of P_3 to reduce S_2 to $S_2 = L_3^2 + c(P_1^2 - P_2^2)$, corresponding to II or III_1 type coordinates. If $a=0, b^2 + c^2\neq 0$, we obtain $S_2 = L_3P_1 + P_1L_3$, corresponding to the type III_1. If $a=b=c=0$, we obtain type I coordinates.

(iii) $X = L_3 + aP_3$; nor $(L_3 + aP_3) = \{L_3,P_3\}$: A straightforward computation shows that in this case S_2 satisfying $[X,S_2]=0$ can be reduced to

$$S_2 = P_3(bL_3 + cP_3). \tag{3.17}$$

Since L_3 and P_3 commute, a diagonalization of P_3 and L_3 separately is equivalent to a diagonalization of any polynomials in L_3 and P_3. We thus reobtain case II.

(IV) No ignorable variables: Neither of the operators S_1 or S_2 is the square of a generator of e(3).

Type (IV_1): We return to the non-Abelian subalgebras of e(3) discussed above in III_3. We take S_1 to be the Casimir operator of such a subalgebra and S_2 some second order element of the enveloping algebra of the corresponding subalgebra. These operators S_2 must be classified into orbits under the group A whose Lie algebra is A. Only one such case occurs for e(3), namely

$$S_1 = \mathbf{L}^2,\ S_2 = L_1^2 + rL_2^2,\quad 0<r<1, \tag{3.18}$$

corresponding to *spheroconical* coordinates (S_2 is not allowed to be the square of a generator).

Type (IV_2): Here S_1 and S_2 are simply commuting second order operators in the enveloping algebra of e(3). Neither of them is the square of a generator nor a Casimir operator of any Lie algebra. This is the generic case with the lowest symmetry. The remaining coordinates *ellipsoidal* and *paraboloidal* are of this type.

This completes the list of all 11 types of separable coordinates in Euclidean 3-space.

Finally, let us discuss the question of discrete symmetries that further characterize some of the coordinate systems. Indeed, for coordinates of the type III_1, III_4, and IV_1 only one of the diagonal operators is characterized by the fact that it is an invariant operator of a one or higher dimensional Lie algebra. For coordinates of the type IV_2 neither S_1 nor S_2 has this property. These operators will, in general, be invariants of certain discrete subgroups of E(3). No operator of the type

$$S = a_{ik}L_iL_k + b_{ik}P_iP_k + c_{ik}(L_iP_k + P_kL_i) \tag{3.19}$$

is left invariant by discrete translations (unless $a_{ik} = c_{ik} = 0$ and we have continuous translational invariance). We can hence restrict ourselves to point groups and indeed to groups of reflections in planes through the origin. Let us use X,Y, and Z to denote a reflection of the coordinate x,y, and z, respectively, and I_2n $(A_1,\ldots,A_n)$ to denote the Abelian group of order 2^n generated by $A_1,\ldots,A_n$. By inspection we see that the operators S_i not related to Lie subgroups have the following invariance groups:

$$
\begin{aligned}
&L_3^2 + a(P_1^2 - P_2^2) && : I_4(X,Y)\,,\\
&L_3P_2 + P_2L_3 && : I_2(Y)\,,\\
&L_1P_2 + P_2L_1 - L_2P_1 - P_1L_2 && : I_4(X,Y)\,,\\
&\mathbf{L}^2 \pm a(P_1^2 + P_2^2) && : I_4(Z,XY)\,,\\
&\left.\begin{aligned} &L_2^2 + aL_1^2 + bP_3^2\\ &\mathbf{L}^2 + bP_1^2 + aP_2^2 + (a+b)P_3^2 \end{aligned}\right\} && : I_8(X,Y,Z)\,,\\
&\left.\begin{aligned} &L_3^2 - c^2P_3^2 + c(L_2P_1 + P_1L_2 + L_1P_2 + P_2L_1)\\ &L_2P_1 + P_1L_2 - L_1P_2 - P_2L_1 + c(P_2^2 - P_1^2) \end{aligned}\right\} && : I_4(X,Y)\,.
\end{aligned}
\tag{3.20}
$$

Thus, the operators S_1,S_2 for each of the 11 separable coordinate systems can be viewed as corresponding to a certain subgroup reduction of E(3) and both Lie subgroups and discrete subgroups figure in the reductions. The subgroups will determine the symmetry properties of the separated solutions of the Helmholtz equations. In particular, the discrete subgroups are often important in physical applications, especially in the context of "symmetry adapted basis functions" in molecular physics and general many body theories.[5,29-33]

The results of this section are summarized in Table I. We do not spell out the explicit form of the coordinates. The ones used are listed, for example, in Ref. 38.

B. The group O(4)

Separable systems of coordinates in s_3, the unit sphere, were first obtained by Eisenhart[34] and studied from the algebraic point of view in Ref. 13. Let us now classify them from the subgroup point of view. The continuous subgroups of O(4) are listed, for example, in Ref. 20 (they were first obtained by Goursat[39]).

Using the isomorphism $o(4) \sim o(3) \oplus o(3)$ we write the algebra o(4) as $\{A_i,B_i, i = 1,2,3,\}$, satisfying

$$
[A_i,A_k] = \epsilon_{ikl}A_l,\quad [B_i,B_k] = \epsilon_{ikl}B_l,\quad [A_i,B_k] = 0\,. \tag{3.21}
$$

The algebra o(4) has precisely one MASA [up to conjugacy under O(4)], namely, $\{A_3,B_3\}$. Hence, no class I systems exist and just one class II system. The one-dimensional subalgebras are $A_3, A_3 + xB_3 (0 < x < 1)$, and $A_3 + B_3$. No type III_1 or III_2 coordinates exist on s_3; III_2 is excluded because cent(A_3) is a direct sum and III_1 is not realized because the operators $(A_3^2, B_1^2 + k^2B_2^2)$ would correspond to separation on $s_2 \otimes s_2$ rather than s_3 (the III_1 and III_2 type coordinates only exist on flat three-dimensional manifolds). The only non-Abelian subalgebra of o(4) with a second order Casimir operator that is not a Casimir operator of o(4) is $\{A_1 + B_1, A_2 + B_2, A_3 + B_3\}$. This provides III_3 coordinates for

$$
\begin{aligned}
&S_1 = (A_1 + B_1)^2 + (A_2 + B_2)^2 + (A_3 + B_3)^2,\\
&S_3 = (A_3 + B_3)^2\,,
\end{aligned}
\tag{3.22}
$$

and type IV_1 coordinates for

$$
\begin{aligned}
&S_1 = (A_1 + B_1)^2 + (A_2 + B_2)^2 + (A_3 + B_3)^2\,,\\
&S_3 = (A_1 + B_1)^2 + k^2(A_2 + B_2)^2 \quad (0 < k^2 < 1)\,.
\end{aligned}
\tag{3.23}
$$

Type III_4 coordinates are obtained from $A_3 + B_3$ only. The operator S_2 commuting with $A_3 + B_3$ can be reduced to

$$
S_2 = A_1B_1 + A_2B_2 + \alpha A_3B_3,\quad \alpha > 0, \alpha \neq 1
$$

and we distinguish between $0 < \alpha < 1$ and $1 < \alpha \quad \infty$.

Type IV_1 coordinates were discussed above and type IV_2 also occurs.[13]

Notice that only pairs of operators S_1,S_2 that are invariant under parity Π, i.e.,

$$
\Pi: (x_1,x_2,x_3,x_4) \rightarrow (-x_1,-x_2,-x_3,x_4)\,, \tag{3.24}
$$

lead to separable coordinates on s_3, as was shown in Ref. 13. Here,

$$
x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1\,.
$$

TABLE I. Separable coordinate systems for E(3).

Type	Coordinates	Diagonal operators		Subgroup chain
I	Cartesian	P_1,P_2,P_3		T(3)
II	Cylindrical	L_3,P_3		$O(2) \otimes T(1)$
III_1	Elliptic cylindrical	$P_3,L_3^2 + a(P_1^2 - P_2^2)$	$a > 0$	$E(2) \otimes T(1) \supset I_4(X,Y) \otimes T(1)$
	Parabolic cylindrical	$P_3,L_3P_2 + P_2L_3$		$E(2) \otimes T(1) \supset I_2(Y) \otimes T(1)$
III_3	Spherical	$L_3,L_1^2 + L_2^2 + L_3^2$		$O(3) \supset O(2)$
III_4	Parabolic	$L_3,L_1P_2 + P_2L_1 - L_2P_1 - P_1L_2$		$O(2) \otimes I_4(X,Y)$
	Oblate spheroidal	$L_3,L_1^2 + L_2^2 + L_3^2 + a(P_1^2 + P_2^2)$	$a > 0$	$O(2) \otimes I_4(Z,XY)$
	Prolate spheroidal	$L_3,L_1^2 + L_2^2 + L_3^2 - a(P_1^2 + P_2^2)$	$a > 0$	$O(2) \otimes I_4(Z,XY)$
IV_1	Spheroconical	$L_1^2 + L_2^2 + L_3^2, L_1^2 + rL_2^2$	$0 < r < 1$	$O(3) \supset I_8(X,Y,Z)$
IV_2	Ellipsoidal	$L_2^2 + aL_1^2 + bP_3^2$,	$a > b > 0$	
		$L_1^2 + L_2^2 + L_3^2 + bP_1^2 + aP_2^2 + (b+a)P_3^2$		$I_8(X,Y,Z)$
	Paraboloidal	$L_3^2 - c^2P_3^2 + c(L_2P_1 + P_1L_2 + L_1P_2 + P_2L_1)$,		
		$L_2P_1 + P_1L_2 - L_1P_2 - P_2L_1 + c(P_2^2 - P_1^2)$	$c > 0$	$I_4(X,Y)$

TABLE II. Separable coordinate systems on s_3.

Type	Coordinates	Operators	Subgroup reduction
II	Cylindrical	L_3^2, J_3^2	$O(2)\times O(2)$
III_3	Spherical	$L_1^2+L_2^2+L_3^2, L_3^2$	$O(3)\supset O(2)$
III_4	Elliptic cylindrical I and II	$L_1^2+L_2^2+L_3^2+a(L_3^2-J_3^2), L_3^2 \quad (a\neq 0)$	$O(2)\times I_8(X_1,X_2,X_3,X_4)$
IV_1	Spheroelliptic	$L_1^2+L_2^2+L_3^2, L_1^2+rL_2^2 \quad (0<r<1)$	$O(3)\supset D_2$
IV_2	Ellipsoidal	$L_1^2-J_1^2+\frac{1-a+b}{a+b-1}(L_2^2-J_2^2)+\frac{1+a-b}{a+b-1}(L_3^2-J_3^2)$, $L_1^2+J_1^2+\frac{b-a-1}{a+b-1}(L_2^2-J_2^2)+\frac{a-b}{b(a-1)}(L_3^2+J_3^2)$ $+\frac{a(b-1)(a-b-1)}{b(a-1)(a+b-1)}(L_3^2-J_3^2) \quad (1<b<a)$	D_2

The results of this paragraph are summarized in Table II, together with the discrete subgroup properties of each system. Again I_{2n} $(A_1,\ldots,A_n)$ will be a group of reflections in hyperplanes through the origin with, for example, X reflecting the Cartesian coordinate x only. We write the invariant operators S_1 and S_2 in terms of $L_i = A_i + B_i$ and $J_i = A_i - B_i$, rather than A_i and B_i directly (the J_i do not constitute a subalgebra).

C. The group O(3,1)

The subalgebras of o(3,1) have been classified[1] under the action of O(3,1) and the results are reproduced in, for example, Ref. 20.

The algebra o(3,1) is generated by the rotations L_i and boosts K_i, satisfying

$$[L_i,L_j]=\epsilon_{ijk}L_k,\ [L_i,K_j]=\epsilon_{ijk}K_k,\ [K_i,K_j]=-\epsilon_{ijk}L_k\ . \tag{3.25}$$

The Casimir operators are $\Delta = \mathbf{L}^2 - \mathbf{K}^2$ and $\Delta' = \mathbf{L}\cdot\mathbf{K}$ (we have $\Delta' = 0$). All separable coordinates for O(3,1) hyperboloids were obtained by Olevskiĭ[36]; the pairs of commuting operators S_1 and S_2 corresponding to these 34 coordinate systems are also known.[14]

The algebra o(3,1) has two MASA. Both are two-dimensional, namely,

$$\{L_3,K_3\} \text{ and } \{L_2+K_1, L_1-K_2\}\ .$$

We hence have no type I coordinates and two type II coordinate systems.

The one-dimensional subalgebras are $\{L_3\}$, $\{K_3\}$, $\{L_2+K_1\}$, and $\{L_3+aK_3; a>0\}$. None of these have non-Abelian centralizers, so we obtain no III_1 or III_2 type coordinate systems. Subgroup type coordinates III_3 are obtained from the subgroups O(3), O(2,1), and E(2). The corresponding pairs of operators are

$$(\mathbf{L}^2,L_3^2),(K_1^2+K_2^2-L_3^2,L_3^2),(K_1^2+K_2^2-L_3^2,K_1^2)\ , \tag{3.26}$$

$$(K_1^2+K_2^2-L_3^2,(K_1+L_3)^2) \text{ and } ((K_1+L_2)^2+(L_1-K_2)^2,L_3^2)\ .$$

Now let us consider III_4 type coordinates:

(i) $S_1 = L_3^2$, $\mathrm{nor}(L_3) = \{L_3,K_3\}$: The most general second order operator S_2 commuting with L_3 can, after linear combinations with Δ,Δ', and L_3^2 have been accounted for, be written as

$$S_2 = a(K_1^2+K_2^2+L_1^2+L_2^2)+b(K_2L_1+L_1K_2 - K_1L_2-L_2K_1)+cK_3^2+dL_3K_3\ . \tag{3.27}$$

The transformation $\exp\alpha K_3$ induces a hyperbolic rotation between the first two terms. Hence, if $|a|<|b|$, we can transform a into zero; if $|a|>|b|$, we can transform b into zero; and if $|a|=|b|$, the first two terms reduce to $(L_1+K_2)^2+(L_2-K_1)^2$. In these coordinates we have $L_3 = \partial/\partial\phi$ and the term L_3K_3 will be odd under the transformation $\phi\to-\phi$ which should leave S_2 invariant. Hence, $d=0$. If $|a|=|b|$, K_3 can be used to scale the value of a (and b) with respect to c. Using $\exp\alpha K_3$, parity, and linear combinations with Δ we can finally reduce S_2 to one of the forms:

$$K_1^2+K_2^2+aK_3^2, L_1K_2+K_2L_1-L_2K_1-K_1L_2+aK_3^2 \quad (a\geqslant 0) \tag{3.28}$$

$$(L_1+K_2)^2+(L_2-K_1)^2+\epsilon K_3^2 \quad (\epsilon=\pm 1)\ .$$

In the first case we distinguish between the regions $0<a<1, 1<a<\infty$, and $-\infty<a<0$.

(ii) $S_1 = K_3^2$, $\mathrm{nor}(K_3) = \{L_3,K_3\}$: Imposing $[K_3^2,S_2]=0$ and using linear combinations with Δ,Δ', and K_3^2 we have

$$S_2 = a(L_1^2-K_2^2-L_2^2+K_1^2)+b(L_1L_2+L_2L_1+K_1K_2 + K_2K_1)+c(L_1^2-K_2^2+L_2^2-K_1^2)+dL_3K_3\ . \tag{3.29}$$

In these coordinates we have $K_3 = \partial/\partial\beta$ and L_3K_3 changes sign for $\beta\to-\beta$. Hence, $d=0$. The operator $\exp\alpha L_3$ will rotate between the first two terms. Hence, we can always rotate b in zero (the case $a^2+b^2=0$ would lead back to type II coordinates). We thus obtain

$$S_2 = K_1^2-L_2^2+a(L_1^2-K_2^2), \quad 0<|a|<1 \tag{3.30}$$

and we distinguish between $0<a<1$ and $-1<a<0$.

(iii) $S_1 = (K_1+L_2)^2$,
$\mathrm{nor}(K_1+L_2) = \{K_3, K_1+L_2, K_2-L_1\}$: (3.31)

The operator S_2 satisfying $[K_1+L_2,S_2]=0$ can be written as

$$S_2 = a(K_1^2+K_3^2-L_2^2)+b[(K_1+L_2)L_3+L_3(K_1+L_2) + K_3(K_2-L_1)+(K_2-L_1)K_3] + c[(K_2-L_1)^2+(K_1+L_2)^2] + d(K_1+L_2)(K_2-L_1)\ . \tag{3.32}$$

The separable coordinates (u,v,t) will be such that $K_1 + L_2 = \partial/\partial_t$. The term $(K_1 + L_2)(K_2 - L_1)$ will be odd under the reflection $t \to -t$; hence, $d = 0$. If $a \neq 0$, we put $a = 1$ and use $\exp\alpha(K_2 - L_1)$ to transform $b \to 0$. Further $\exp\beta K_3$ will scale c with respect to a. We obtain

$$S_2 = K_1^2 + K_3^2 - L_2^2 + \epsilon[(K_2 - L_1)^2 + (K_1 + L_2)^2], \quad \epsilon = \pm 1. \tag{3.33}$$

($c = 0$ is excluded, since it would lead to type III_3.) If $a = 0$, $b \neq 0$, we put $b = 1$ and use $\exp\alpha(K_2 - L_1)$ to transform $c \to 0$. Finally $a = b = 0$ is excluded, since it would lead to type II.

(iv) $S_1 = (L_3 + aK_3)^2$, $\mathrm{nor}(L_3 + aK_3) = \{L_3, K_3\}$: (3.34)

The most general second-order operator commuting with $L_3 + aK_3$ $(a \neq 0)$ can be reduced to, for example, $K_3(bL_3 + cK_3)$ and hence leads back to type II coordinates.

Type IV_1 coordinates are obtained similarly as type III_3 ones. Indeed, we consider the subgroups O(3),O(2,1), and E(2) of O(3,1) and take S_1 as the corresponding Casimir operator. The operator S_2 will then be a second-order operator in the enveloping algebra of o(3), o(2,1), or e(2), respectively. These operators must be classified into orbits under O(3), O(2,1), or E(2), as the case may be, and orbits corresponding to squares of generators must be excluded. For O(3), O(2,1), and E(2) we obtain one, six, and two orbits, respectively.[15]

Finally, we are left with the generic case IV_2. The operators S_1 and S_2 are such that neither of them is the square of a generator or a Casimir operator of a subgroup of O(3,1) [nor is it conjugate under O(3,1) to such operators].

A further subclassification is obtained by considering discrete subgroups of O(3,1) leaving the individual pairs of operators invariant we omit all details here but summarize the results in Table III, where we give the invariant operators, the subgroup reductions, and identify the coordinate system by the number it carries in Refs. 14 and 36.

D. The group O(2,2)

We shall consider this case in somewhat less detail than the previous ones. Separable systems of coordinates on the hyperboloid $x_1^2 + x_2^2 - x_3^2 - x_4^2 = 1$ were discussed in Ref. 16. The subalgebras of o(2,2) were classified in Ref. 20 and a

TABLE III. Diagonal operators and corresponding subgroup chains for separable coordinate systems on the O(3,1) hyperboloid.

Type	Diagonal operators		Subgroup chain
II	L_3^2, K_3^2		$O(2) \otimes O(1,1)$
	$(L_1 + K_2)^2, (L_2 - K_1)^2$		$T(2)$
III_3	$L_1^2 + L_2^2 + L_3^2, L_3^2$		$O(3) \supset O(2)$
	$K_1^2 + K_2^2 - L_3^2, L_3^2$		$O(2,1) \supset O(1,1)$
	$K_1^2 + K_2^2 - L_3^2, K_1^2$		$O(2,1) \supset O(1,1)$
	$K_1^2 + K_2^2 - L_3^2, (K_1 + L_3)^2$		$O(2,1) \supset T(1)$
	$(L_1 - K_2)^2 + (L_2 + K_1)^2, L_3^2$		$E(2) \supset O(2)$
III_4	$L_3^2, K_1^2 + K_2^2 + aK_3^2$	$0 < a < 1, 1 < a < \infty$ or $-\infty < a < 0$	$O(2) \times I_4(Z,T)$
	$L_3^2, L_1K_2 + K_2L_1 - L_2K_1 - K_1L_2 + aK_3^2$	$a \geqslant 0$	$O(2) \times I_4(X,Y)$
	$L_3^2, (L_1 + K_2)^2 + (L_2 - K_1)^2 + \epsilon K_3^2$	$\epsilon = \pm 1$	$O(2) \times I_4(X,Y)$
	$K_3^2, K_1^2 - L_2^2 + a(L_1^2 - K_2^2)$	$0 < a < 1$ or $-1 < a < 0$	$O(1,1) \times I_{16}(X,Y,Z,T)$
	$(K_1 + L_2)^2, K_1^2 + K_3^2 - L_2^2 + \epsilon[(K_2 - L_1)^2 + (K_1 + L_2)^2]$	$\epsilon = \pm 1$	$T(1) \times I_4(X,Y)$
	$(K_1 + L_2)^2, (K_1 + L_2)L_3 + L_3(K_1 + L_2) + K_3(K_2 - L_1) + (K_2 - L_1)K_3$		$T(1) \times I_2(X)$
IV_1	$L_1^2 + L_2^2 + L_3^2, L_1^2 + aL_2^2$	$0 < a < 1$	$O(3) \supset I_{16}(X,Y,Z,T)$
	$K_1^2 + K_2^2 - L_3^2, L_3^2 - aK_2^2$	$a < -1$ or $0 < a$	$O(2,1) \supset I_{16}(X,Y,Z,T)$
	$K_1^2 + K_2^2 - L_3^2, K_1^2 + a(K_2L_3 + L_3K_2)$	$0 < a$	$O(2,1) \supset I_4(Y,Z)$
	$K_1^2 + K_2^2 - L_3^2, L_3^2 + (L_3K_2 + K_2L_3)$		$O(2,1) \supset I_4(Y,Z)$
	$K_1^2 + K_2^2 - L_3^2, K_2^2 + (L_3K_2 + K_2L_3)$		$O(2,1) \supset I_4(Y,Z)$
	$K_1^2 + K_2^2 - L_3^2, K_1K_2 + K_2K_1 + K_2L_3 + L_3K_2$		$O(2,1) \supset I_2(Z)$
	$(K_1 + L_2)^2(K_2 - L_1)^2, L_3^2 + (K_1 + L_2)^2$		$E(2) \supset I_4(X,Y)$
	$(K_1 + L_2)^2 + (K_2 - L_1)^2, L_3(K_1 + L_2) + (L_2 + K_1)L_3$		$E(2) \supset I_2(X)$
IV_2	$M_1^2 + bM_2^2 + aM_3^2 - (a+b)K_1^2 - (a+1)K_2^2 - (b+1)K_3^2, abK_1^2 + aK_2^2 + bK_3^2,$	$1 < b < a$	$I_{16}(X,Y,Z,T)$
	$M_1^2 - aK_2^2 - bK_3^2 - (a+b)K_1^2 + (a+1)M_3^2 + (b+1)M_2^2,$		$I_{16}(X,Y,Z,T)$
	$abK_1^2 - aM_3^2 - bM_2^2,$	$1 < b < a$	
	$2aM_1^2 - (a+1)(K_3^2 - M_2^2) - a(K_2^2 - M_3^2) - b(K_2M_3 + M_3K_2 - M_2K_3 - K_3M_2),$		$I_4(Y,Z)$
	$(a^2 + b^2)M_1^2 - a(K_3^2 - M_2^2) + b(K_3M_2 + M_2K_3),$	$a, b \in \mathbb{R}$	
	$(K_2 + M_3)^2 + (K_3 + M_2)^2 + (a+1)K_1^2 + K_3^2 - M_2^2 + a(M_3^2 - K_2^2),$		$I_4(Y,Z)$
	$(K_3 + M_2)^2 - a(K_2 + M_3)^2 + aK_1^2,$	$1 < a$	
	$(K_2 + M_3)^2 + (K_3 + M_2)^2 - (a+1)K_1^2 - M_2^2 + K_3^2 - a(K_2^2 - M_3^2),$		$I_4(Y,Z)$
	$(K_3 + M_2)^2 - a(K_2 + M_3)^2 - aK_1^2$	$1 < a$	
	$(K_2 + M_3)^2 - (K_3 + M_2)^2 - (a-1)K_1^2 - M_3^2 + K_2^2 - a(M_2^2 - K_3^2),$		$I_4(Y,Z)$
	$(K_2 + M_3)^2 - a(K_3 + M_2)^2 - aK_1^2$	$0 < a$	
	$M_2^2 - K_3^2 - M_1^2 - (M_2 - K_3)^2 - M_1(M_2 - K_3) - (M_2 - K_3)M_1,$		$I_4(Y,Z)$
	$(M_2 - K_3)^2 - K_1(K_2 - M_3) - (K_2 - M_3)K_1$		

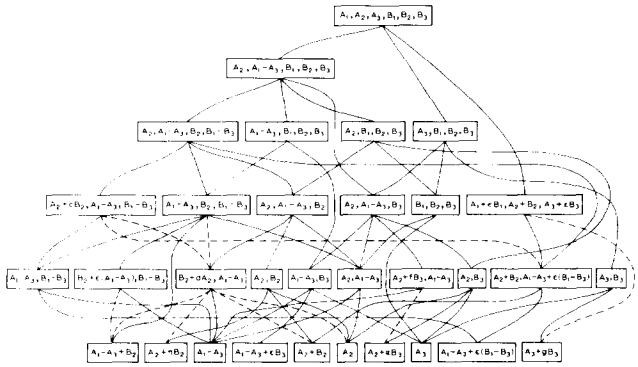


FIG. 1. The O(2,2) conjugacy classes of subalgebras of $o(2,2)$. The parameters take the following real values:
$\epsilon = \pm 1, 0 < |c| \leqslant 1, d \geqslant 0, e > 0, f \neq 0, 0 < |g| \leqslant 1, 0 < h < 1$. A line connects each algebra with its maximal subalgebras. A solid (broken) line indicates an inclusion for all (some) values of the parameters involved.

diagram of them is given in Fig. 1. We use the isomorphism $o(2,2) \sim o(2,1) \oplus o(2,1)$ and write the algebra o(2,2) in the form $\{A_i, B_i\}$:

$$[A_1,A_2] = -A_3,\ [A_3,A_1] = A_2,\ [A_2,A_3] = A_1,$$
$$[B_1,B_2] = -B_3,\ [B_3,B_1] = B_2,\ [B_2,B_3] = B_1, \quad (3.35)$$
$$[A_i,B_k] = 0,\quad i,k = 1,2,3$$

(A_3 and B_3 are the compact elements).

Let us discuss the individual classes of coordinates.

Type I: The algebra o(2,2) has no MASA of dimension 3, and hence this class does not occur.

Type II: There exist six different MASA of dimension 2, each corresponding to a different system of coordinates. Systems I1, I2, and I3 of Ref. 16 are orthogonal and correspond to the subalgebras $\{A_2,B_3\}$, $\{A_2,B_2\}$, and $\{A_1 - A_3, B_1 - B_3\}$, respectively. Systems I4, I5, and I6 are nonorthogonal and correspond to the subalgebras $\{A_3,B_1 - B_3\}$, $\{A_2,B_1 - B_3\}$, and $\{A_3,B_2\}$, respectively.

Type III_1 and III_2: These do not occur since the centralizers of all one-dimensional subalgebras are either Abelian, or reductive of the type

$$\{A_3\} \oplus \{B_1,B_2,B_3\},\ \{A_2\} \oplus \{B_1,B_2,B_3\},$$
$$\text{or } \{A_1 - A_3\} \oplus \{B_1,B_2,B_3\}.$$

These do not lead to separable coordinate systems on the considered hyperboloid [they would on the direct product of two O(2,1) hyperboloids].

Type III_3: The algebra o(2,2) has three non-Abelian subalgebras with second order Casimir operators distinct from the Casimir operators of o(2,2). These are

(i) $e(1,1)$:$\{A_2 - B_2, A_1 - A_3, B_1 - B_3\}$

[here $S_1 = (A_2 - B_2)^2, S_2 = (A_1 - A_3)(B_1 - B_3)$ leads to one coordinate system],

(ii) $o(2,1)$:$\{A_1 + B_1, A_2 + B_2, A_3 + B_3\}$,

(iii) $o(2,1)$:$\{A_1 - B_1, A_2 + B_2, A_3 - B_3\}$.

Each of the o(2,1) subalgebras leads to three different subgroup type coordinate systems.

Type III_4: The one-dimensional subalgebras providing III_4 type coordinates are $\{A_3 + B_3\}$, $\{A_3 - B_3\}$, $\{A_2 + B_2\}$, $\{A_1 - A_3 + B_1 - B_3\}$, and $\{A_1 - A_3 - B_1 + B_3\}$, leading to two, two, nine, three, and three systems, respectively.

Type IV_1: The same subalgebras e(1,1) and o(2,1) as in case III_3 lead to these "semisubgroup" type coordinates, of which there exist $8 + 6 + 6 = 20$.

Type IV_2: The remaining generic case leads to 22 more coordinate systems.[16]

Thus, altogether 74 separable coordinate systems exist. Of these exactly six are nonorthogonal. We shall not discuss their discrete symmetries here.

E. The group E(2,1)

Separation of variables in three-dimensional Minkowski space has not been investigated with the same amount of detail as in the other three-dimensional spaces of constant curvature. The coordinate systems can however be extracted from Refs. 17 and 14. The subgroup structure of E(2,1) on the other hand is known.[21] We write the algebra e(2,1) in the form $\{K_1,K_2,L_3,P_0,P_1P_2\}$:

$$[K_1,K_2] = -L_3, [L_3,K_1] = K_2, [L_3,K_2] = -K_1,$$
$$[K_i,P_0] = P_i, [K_i,P_k] = \delta_{ik}P_0, [L_3,P_0] = 0, \quad (3.36)$$
$$[L_3,P_1] = P_2, [L_3,P_2] = -P_1, [P_\mu,P_\nu] = 0,$$
$$(i,k = 1,2; \mu,\nu = 0,1,2).$$

Type I: There is one three-dimensional MASA: $\{P_0,P_1,P_2\}$ corresponding to Cartesian coordinates.

Type II: There are four different MASA of dimension 2. Two of them $\{K_1,P_2\}$ and $\{L_3,P_0\}$ correspond to orthogonal coordinates, and two others correspond to nonorthogonal ones. These are

$S_1 = (P_0 - P_2)^2$, $S_2 = (L_3 + K_1)^2$:
$$x = x_2x_3,\ y = x_1 - \tfrac{1}{2}x_2^2x_3,\ t = -x_1 + x_3 + \tfrac{1}{2}x_2^2x_3,$$

and

$S_1 = (P_0 - P_2)^2$, $S_2 = (L_3 + K_1 + P_0 + P_2)^2$:
$$x = x_2(x_2 + x_3),\ y = x_1 + x_2 - x_2^2\left(\frac{x_2}{3} + \frac{x_3}{2}\right),$$
$$t = -x_1 + x_2 + x_3 + x_2^2\left(\frac{x_2}{3} + \frac{x_3}{2}\right).$$

Type III_1: Among the nine types of one-dimensional subalgebras of e(2,1) precisely three algebras have non-Abelian centralizers, two of which are direct sums. These are as follows: (i) $\{P_1\}$ with $\text{cent}(P_1) = P_1 \oplus \{K_2,P_0,P_2\}$: Hence, $S_1 = P_1^2$ and S_2 is an element of the enveloping algebra of e(1,1), not equal to the Casimir operator, nor to the square of a generator. This leads to eight orthogonal coordinate systems. (ii) $\{P_0\}$ with $\text{cent}(P_0) = P_0 \oplus \{L_3,P_1,P_2\}$: Hence, $S_1 = P_0^2$ and S_2 is either $L_3P_1 + P_1L_3$ or $L_3^2 + a(P_1^2 - P_2^2)$ with $a > 0$ (two orthogonal systems).

Type III_2: The only element of e(2,1) that has a nonseparable centralizer is $(P_0 - P_2)$ with $\text{cent}(P_0 - P_2) = \{L_3 - K_1, P_0 + P_2, P_1, P_0 - P_2\}$ (this is a nilpotent algebra). In this case we have nor $(P_0 - P_2) = \{K_2, L_3 - K_1, P_0 + P_2, P_1, P_0 - P_2\}$. The choice $S_1 = (P_0 - P_2)^2$ and S_2 a member of the enveloping algebra of $\text{cent}(P_0 - P_2)$ (not equal to a square of a generator, nor to

a Casimir operator) leads to three nonorthogonal coordinate systems. These are as follows:

(i) $S_2 = P_1(L_3 - K_1) + (L_3 - K_1)P_1$,

$x = x_3\sqrt{x_2}$, $y - t = x_1 - \frac{1}{4}x_3^2$,

$y + t = 2x_2$;

(ii) $S_2 = (L_3 - K_1)^2 + 4P_1^2$,

$x = x_3\sqrt{1 + x_2^2}$, $y - t = x_1 - \frac{1}{2}x_3^2 x_2$,

$y + t = 2x_2$;

(iii) $S_2 = (L_3 - K_1)^2 + 8aP_1(P_1 - P_2)$, $a > 0$,

$$x = x_2 x_3 + \frac{a}{x_2}, \quad y - t = x_1 - \tfrac{1}{2}x_2 x_3^2 + a\frac{x_3}{x_2} + \frac{a^2}{6(x_2)^3},$$

$y + t = 2x_2$.

Type III_3: Subgroup type coordinates in this case only originate from the O(2,1) subgroup. We obtain three coordinate systems, corresponding to $S_1 = L_3^2, K_2^2$, or $(L_3 - K_1)^2$ and $S_2 = K_1^2 + K_2^2 - L_3^2$.

Type III_4: Taking $S_1 = L_3^2, K_2^2$, or $(L_3 - K_2)^2$ we obtain 10 orthogonal coordinate systems.

Type IV_1: Semisubgroup type coordinates again originate from O(2,1) only and six types of them exist.

Type IV_2: The generic class here consists of 22 types of coordinates.

The total is 54 orthogonal coordinate systems, and five nonorthogonal ones. We shall not go into the problem of discrete symmetries here.

[1]P. Winternitz and I. Friš, Yad. Fiz. **1**, 889 (1965) [Sov. J. Nucl. Phys. **1**, 636 (1965)].
[2]A. Makarov, Ya. A. Smorodinsky, Kh. Valiev, and P. Winternitz, Nuovo Cimento A **52** 1061 (1967).
[3]P. Winternitz, I. Lukăc, and Ya. A. Smorodinsky, Yad. Fiz. **7**, 192 (1968) [Sov. J. Nucl. Phys. **7**, 139 (1968)].
[4]N. Macfadyen and P. Winternitz, J. Math. Phys. **12**, 281 (1971).
[5]J. Patera and P. Winternitz, J. Math. Phys. **14**, 1130 (1973).
[6]W. Miller, Jr., SIAM (Soc. Ind. Appl. Math.) J. Math. Anal. **5**, 626, 822 (1974).
[7]E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **15**, 1263 (1974).
[8]E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **15**, 1025, 1728 (1974); **16**, 2507 (1975); **17**, 331, 356, 369 (1976).
[9]C.P. Boyer, E.G. Kalnins, and W. Miller, Jr., J. Math. Phys. **16**, 499, 512 (1975).
[10]P. Havas, J. Math. Phys. **16**, 1461, 2476 (1975).
[11]E.G. Kalnins, SIAM (Soc. Ind. Appl. Math.) J. Math. Anal. **6**, 340 (1975).
[12]C.P. Boyer, E.G. Kalnins, and W. Miller, Jr., Nagoya Math. J. **60**, 35 (1976).
[13]E.G. Kalnins, W. Miller, Jr., and P. Winternitz, SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **30**, 630 (1976).
[14]E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **18**, 1 (1977).
[15]W. Miller, Jr., *Symmetry and Separation of Variables* (Addison-Wesley, Reading, Mass., 1977).
[16]E.G. Kalnins and W. Miller, Jr., Proc. R. Soc. Edinburgh, Sect. A **79**, 227 (1977).
[17]E.G. Kalnins and W. Miller, Jr., "Separable Coordinates for Three-Dimensional Complex Riemanian Spaces," J. Diff. Geom. (to be published).
[18]C.P. Boyer, E.G. Kalnins, and W. Miller, Jr., Commun. Math. Phys. **59**, 285 (1978).
[19]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **15**, 1378, 1932 (1974); **16**, 1597, 1615 (1975).
[20]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 717 (1976).
[21]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, Can. J. Phys. **54**, 950 (1976).
[22]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 977, 986 (1976).
[23]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **18**, 2259 (1977).
[24]G. Burdet, J. Patera, M. Perrin, and P. Winternitz, Ann. Sci. Math. Québec **2**, 81 (1978); J. Math. Phys. **19**, 1758 (1978).
[25]C.P. Boyer, R.T. Sharp, and P. Winternitz, J. Math. Phys. **17**, 1439 (1976).
[26]J. Beckers, J. Patera, M. Perroud, and P. Winternitz, J. Math. Phys. **18**, 72 (1977).
[27]H. Zassenhaus, preprint CRM-764, Centre de Recherches Mathématiques, Université de Montréal, 1978 (to be published).
[28]J. Patera, P. Winternitz, and H. Zassenhaus, preprint CRM-814, Centre de Recherches, Mathématiques, Université de Montréal, 1978 (to be published).
[29]B. Meyer, Can. J. Math. **6**, 135 (1954).
[30]J. Patera and P. Winternitz, J. Chem. Phys. **65**, 2725 (1976).
[31]J. Patera, R.T. Sharp, and P. Winternitz, J. Math. Phys. **19**, 2362 (1978); P. Desmier and R.T. Sharp, J. Math. Phys. **20**, 74 (1979).
[32]M.R. Kibler and P.A.M. Guichon, Int. J. Quantum Chem. **10**, 87 (1976).
[33]R.P. Bickerstaff and B.G. Wybourne, J. Phys. A **9**, 1051 (1976); M.V. Jarić and J.L. Birman, J. Math. Phys. **18**, 1456, 1459, 2085, (1977).
[34]L.P. Eisenhart, Ann. Math. **35**, 284 (1934).
[35]P. Stäckel, Math. Ann. **49**, 145 (1897).
[36]P.M. Olevskiĭ, Mat. Sb. **27**, 379 (1950) (In Russian).
[37]P. Moon and D.E. Spencer, *Field Theory Handbook* (Springer, Berlin, 1961).
[38]Ph. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Part 1.
[39]E. Goursat, Ann. Sci. Ecole Supér. **6**, 9 (1889).

Generating functions for SU(2) plethysms with fixed exchange symmetry [a)]

J. Patera
Centre de recherche de mathématiques appliquées, Université de Montréal, Montréal, Québec, Canada
R. T. Sharp
Physics Department, McGill University, Montréal, Québec, Canada

(Received 13 May 1980; accepted for publication 7 August 1980)

A number of fixed-plethysm generating functions are given for SU(2); a fixed -plethysm generating function gives the content of the component with definite exchange symmetry of the direct product of a given number of copies of an irreducible representation R_l, with l running through all values. New symmetries are found relating antisymmetric to symmetric products, and relating plethysms in which the number of factors is interchanged with the factor representation label l. Expressions for two-box plethysm generating functions for SU(3) and for fixed-plethysm generating functions for SU(2) based on reducible representations are also given.

PACS numbers: 02.20Qs

1. INTRODUCTION

For over a century there have existed examples of a certain type of generating function[1] related to the representation theory of the group SU(2).

Consider the tensor product of p copies of a representation R_l of SU(2)

$$R_l \otimes \cdots \otimes R_l = \bigoplus_{[\lambda],s} n_{ls}^{[\lambda]} [\lambda] \times R_s . \tag{1.1}$$

The sum is over representations s of SU(2) and those representations $[\lambda]$ of the permutation group S_p whose Young tableaus have no columns exceeding $l+1$ in length. The dimension of the SU(2) representation R_l (or R_s) is $l+1$ (or $s+1$); thus l (or s) is twice the angular momentum associated with the representation l (or s). The partial sum $\oplus_s n_{ls}^{[\lambda]} R_s$ in (1.1), to be denoted by $R_l^{[\lambda]}$, is the plethysm of exchange symmetry $[\lambda]$ based on the representation R_l; the coefficient $n_{ls}^{[\lambda]}$ is the multiplicity of the representation R_s in the plethysm. The S_p representation $[\lambda]$ is more explicitly written $[1^{\lambda_1}, 2^{\lambda_2}, \ldots, r^{\lambda_r}]$ where the non-negative integer λ_i is the number of rows of length i in its Young tableau. The λ_i satisfy

$$\sum_{i=1}^{r} i\lambda_i = p . \tag{1.2}$$

Cayley, Sylvester,[1] and collaborators long ago calculated a rational form of the generating function

$$F_l(P,S) = \sum_{p,s=0}^{\infty} n_{ls}^{[\lambda]} P^p S^s , \tag{1.3}$$

for the symmetric pelthysm $R_l^{[p]}$, $0 \leqslant l \leqslant 12$.

Analogous generating functions have been given[2-4] for groups (including finite groups) other than SU(2) and for more general plethysms. We describe such generating functions as being of the Cayley–Sylvester, or fixed-l, type. For practical reasons, R_l is usually restricted to representations of fairly low dimension.

Multiplicities $n_{ls}^{[\lambda]}$ of the symmetric part or n_{ls} of the full tensor product were studied previously (cf. Ref. 5 and references therein).

Recently[4] examples have been given of a new type of generating function

$$\phi_{[\lambda]}(L,S) = \sum_{l,s} n_{ls}^{[\lambda]} L^l S^s . \tag{1.4}$$

We describe them as being of fixed-plethysm, or fixed-symmetry, type. The Young tableau $[\lambda]$, for practical reasons, is restricted to have a fairly small number p of boxes. This paper is concerned with fixed-symmetry generating functions.

In Sec. 2 we give explicit expressions, for a number of fixed-symmetry generating functions; their interpretation is explained.

Section 3 points out some remarkable symmetries connecting apparently unrelated plethysms and their generating functions.

Section 4 explains how the generating functions of Sec. 2 are derived.

Section 5 contains a brief discussion of possible generating functions of more general type. Fixed-symmetry generating functions of type [2] and $[1^2]$ are given for SU(3) representations. Generating functions for plethysms [1], [2], $[1^2]$, [3], $[1^3]$, [2,1] based on reducible SU(2) representations are also shown.

2. SOME FIXED-SYMMETRY GENERATING FUNCTIONS

Fixed-symmetry generating functions of symmetric type $[p]$, $0 \leqslant p \leqslant 5$, are[4,6]

$$\phi_{[0]}(L,S) = (1-L)^{-1} , \tag{2.1}$$

$$\phi_{[1]}(L,S) = (1-LS)^{-1} , \tag{2.2}$$

$$\phi_{[2]}(L,S) = [(1-L^2)(1-LS^2)]^{-1} , \tag{2.3}$$

[a)]Supported in part by the Natural Sciences and Engineering Research Council of Canada and by the Ministère de l'Education du Québec.

 0022-2488/81/020261-06$1.00

$$\phi_{[3]}(L,S) = (1+L^3S^3)[(1-L^4)(1-L^2S^2)(1-LS^3)]^{-1}, \tag{2.4}$$

$$\phi_{[4]}(L,S) = (1+L^3S^6)[(1-L^2)(1-L^3)(1-L^2S^4)(1-LS^4)]^{-1}, \tag{2.5}$$

$$\begin{aligned}\phi_{[5]}(L,S) = {} & [(1-L^4)(1-L^8)(1-LS^5)(1-L^2S^2)]^{-1} \\ & \times[(1-L^2S^6)^{-1}\{1+L^3S^9+L^4S^6+L^5(S^7+S^9)+L^6(S^4+S^{10}) \\ & +L^7(S^5+S^7+L^8(S^2+S^8)+L^{11}S^{11}\}+(1-L^{12})^{-1}\{L^3(S^3+S^5) \\ & +L^4S^4+L^5(S+S^3)+L^6S^2+L^7S+L^8S^4+L^9(S^3+S^5)+L^{10}(S^2+S^4) \\ & +L^{11}(S+S^3)+L^{12}+L^{13}S+L^{14}S^4+L^{16}S^2+L^{18}\}]\,. \end{aligned} \tag{2.6}$$

Fixed-symmetry generating functions $\phi_{[1^p]}$ of antisymmetric type $[1^p]$, $1\leqslant p$, are related to the fixed-symmetry generating functions of symmetric type through

$$\phi_{[1^p]}(L,S) = L^{p-1}\phi_{[p]}(L,S)\,. \tag{2.7}$$

For mixed symmetry [2,1] the generating function is

$$\phi_{[2,1]} = \frac{LS}{(1-L^2)(1-LS)(1-LS^3)}\,. \tag{2.8}$$

For the three mixed symmetries with four boxes the generating functions are

$$\phi_{[2,1^2]}(L,S) = \frac{L^2S^2}{(1-L)(1-L^2)(1-LS^2)(1-LS^4)} = L\phi_{[3,1]}(L,S)\,, \tag{2.9}$$

$$\phi_{[2^2]}(L,S) = \frac{L+L^4S^6}{(1-L^2S^4)(1-LS^4)(1-L)(1-L^3)}\,. \tag{2.10}$$

For the rectangular Young tableaus with six boxes we find

$$\begin{aligned}\phi_{[3^2]}(L,S) = {} & \frac{1}{(1-L^2)^2(1-L^6)(1-L^2S^8)}\left(\frac{L+L^5S^4+L^5S^{10}+L^7S^6+L^8S^4+L^{10}+L^{10}S^6+L^{14}S^{10}}{(1-L^2S^4)(1-L^4)}\right. \\ & +\frac{L^2S^2+L^3S^6+L^4S^4+L^5S^8+L^5S^{10}+L^5S^{12}+L^6S^6+L^7S^6+L^7S^8+L^7S^{10}+L^8S^8+L^9S^{12}}{(1-LS^6)(1-LS^2)} \\ & +\frac{L^2S^6+L^4S^6+L^4S^8+L^4S^{12}+L^6S^4+L^6S^6+2L^6S^8+L^6S^{10}+L^8S^4+L^8S^{10}+L^{10}S^{10}}{(1-LS^6)(1-L^2S^4)} \\ & \left.+\frac{L^4S^2+L^4S^6+L^5S^2+L^5S^6+2L^6S^2+L^7S^2+L^7S^6+L^8S^6+L^9S^2+L^9S^6+L^{10}S^6}{(1-LS^2)(1-L^4)}\right) \\ = {} & L^{-1}\phi_{[2^2]}(L,S)\,. \end{aligned} \tag{2.11}$$

The generating functions (2.1)–(2.5), their antisymmetric counterparts, and (2.8) are given in Ref. 4. They are reproduced here for the sake of completeness, because of changes in notation, and to correct misprints in Eq. (7.11) of Ref. 4.

of the indicated symmetry based on any SU(2) representation R_l. For example, the plethysm of exchange symmetry $[2^2]$ based on the SU(2) representation R_4 is given by the coefficient of L^4 in the expansion of $\phi_{[2^2]}(L,S)$, Eq. (2.10), namely,

$$2+2S^4+S^6+2S^8+S^{12}\,. \tag{2.12}$$

This implies that the component of the direct product of four $l=4$ representations (angular momentum 2) with exchange symmetry $[2^2]$ contains the SU(2) representations 12, 8 twice, 6, 4 twice, 0 twice.

It is interesting that the spinor representation R_1 never appears in the symmetric or antisymmetric plethysm for three copies of any R (but it does appear just once in the plethysm of mixed symmetry [2,1]).

3. RELATIONS BETWEEN PLETHYSMS

Some seemingly unrelated plethysms apparently have identical SU(2) decompositions.

Consider the plethysm of symmetry $[x^y]$ (rectangular Young tableau) based on the SU(2) representation R_l, with $l=y+z+1$. The six plethysms obtained by permuting the three positive integers x,y,z appear to have the same SU(2) content. That is, the multiplicities of (1.1) satisfy

$$n^{[x^y]}_{y+z-1,s} = n^{[y^x]}_{x+z-1,s} = n^{[z^y]}_{x+y-1,s} = \text{etc.} \quad (x,y,z\geqslant 1)\,. \tag{3.1}$$

Here the first equality holds also when $z=0$. In terms of generating functions

$$\begin{aligned} L^{-y+1}\phi_{[x^y]}(L,S) &= L^{-x+1}\phi_{[y^x]}(L,S) \\ &= F'_{x+y-1}(L,S) = \text{etc.} \end{aligned} \tag{3.2}$$

The generating function F' in (3.2) is a fixed-l ($l=x+y-1$) generating function for rectangular plethysms of y rows; its dummy L carries the number of columns. It is a special case (all $\Lambda_i=0$ except $\Lambda_y=L$) of the fixed-l generating functions for general plethysms described in Ref. 4.

With the help of branching rule tables[7] we have checked numerous examples of (3.1). Although no complete proof exists, the following argument makes (3.1) extremely plausi-

ble. First consider the dimension of the plethysm $[x^y]$ based on $l = y + z - 1$. It is the dimension of the representation (0,...,0,x,0,...,0) (yth label = x; for definition of the notation see Ref. 4) of SU($y + z$). This is found to be

$$\frac{\prod_{i=0}^{x+y+z-1} i! \prod_{j=0}^{x-1} j! \prod_{k=0}^{y-1} k! \prod_{h=0}^{z-1} h!}{\prod_{l=0}^{x+z-1} l! \prod_{m=0}^{y+z-1} m! \prod_{p=0}^{x+y-1} p!}, \tag{3.3}$$

which is symmetrical in x, y, z. Also the highest representation R_s in the plethysm has $s = xyz$, again symmetrical in x, y, z.

As the second case, consider the plethysm of symmetry $[x, 1^{y-1}]$ (hook-shaped Young tableau) based on R_l with $l = y + z - 1$. The two plethysms obtained by permuting x and y, keeping z fixed, seem to have the same SU(2) content. That is the mutiplicities satisfy

$$n^{[x,1^{y-1}]}_{y+z-1,s} = n^{[y,1^{x-1}]}_{x+z-1,s}, \quad x, y > 0. \tag{3.4a}$$

and the generating functions are related by

$$L^x \phi_{[x,1^{y-1}]}(L,S) = L^y \phi_{[y,1^{x-1}]}(L,S). \tag{3.4b}$$

We have applied the same checks as for the relations (3.1). This time the dimension of the plethysm is

$$\frac{(x+y+z-1)!}{(x+y-1)(x-1)!(y-1)!z!}, \tag{3.5}$$

and the highest s is

$$z(x+y-1)+1, \tag{3.6}$$

both symmetric in x and y.

We have a complete proof of the relations (3.1) for $x = 1$ [they then include the relation (3.4) with $x = 1$]. Begin by showing

$$n^{[1^y]}_{y+z-1,s} = n^{[y]}_{z,s} \quad (y > 0), \tag{3.7a}$$

which is equivalent to the conjectured Eq. (8.1)

$$\phi_{[1^p]}(L,S) = L^{p-1} \phi_{[p]}(L,S), \quad 0 < p, \tag{3.7b}$$

of Ref. 4. Equation (3.7) states that the symmetric plethysm of y boxes based on the SU(2) representation R_z has the same SU(2) content as the antisymmetric plethysm of y boxes based on R_{z+y-1}. The proof proceeds by showing that the weights in the antisymmetric plethysm are identical to those in the symmetric one. Start by giving all y weights their maximum values; in the symmetric plethysm they are all equal to z; in the antisymmetric case they are respectively $z+y-1, z+y-3, \ldots, z-y+1$; in both cases the maximum total weight is zy. Now any other assignment of the y weights is obtained from the initial one by specifying $z+1$ integers i_j satisfying

$$0 \leqslant i_0 \leqslant i_1 \leqslant \cdots \leqslant i_z; \quad \sum_{j=0}^{z} i_j = y; \tag{3.8}$$

i_j is the number of weights which have been reduced by j from their original values; in the antisymmetric case, to make the i_j unique it is agreed that no weight will jump past any other when their values are decreased. Since the same sets of i_j describe the symmetric and antisymmetric weights, the two plethysms contain the same weights, and hence the same SU(2) representations. Hence (2.7) and (3.7) are proved.

Because of conjugation symmetry [the yth fundamental representation of SU($y + z$) has the same SU(2) content as the zth] we have

$$n^{[1^y]}_{y+z-1,s} = n^{[1^z]}_{y+z-1,s}, \quad y, z > 0; \tag{3.9}$$

substituting (3.7) in (3.9) shows that

$$n^{[y]}_{z,s} = n^{[z]}_{y,s}. \tag{3.10}$$

Equation (3.10) represents the surprising (to us) result that the symmetric plethysm of y boxes based on R_z has the same SU(2) content as the symmetric plethysm of z boxes based on R_y. In terms of generating functions, (3.10) implies that the Cayley–Sylvester fixed-l generating function $F_l(P,S)$ of Eq. (1.3) (see also Ref. 5) is equal to the fixed-symmetry $\phi_{[p]}(L,S)$ of Eqs. (2.1)–(2.6) when l and P are identified with p and L, respectively.

4. DERIVATION OF GENERATING FUNCTIONS

To derive (2.11), start with Eq. (3.7) of Ref. 4, in which $\Lambda_1, \Lambda_3, \Lambda_4, \Lambda_5$ have been set equal to 0. The result,

$$\mathcal{F}(\Lambda_2, M_1, M_2) = [1 - \Lambda_2^2)(1 - \Lambda_2 M_1^2)(1 - \Lambda_2^2 M_2^2)]^{-1}, \tag{4.1}$$

is the generating function for O(5) representations (labels carried by M_1, M_2) contained in rectangular two-rowed plethysms (number of columns carried by Λ_2) based on the (01) representation of O(5). To convert (4.1) to the generating function for SU(2) representations contained in rectangular two-rowed plethysms based on R_4 of SU(2), substitute it into Eq. (23) of Ref. 2, the generating function for O(5) $\supset$ SU(2) branching rules. The form of (4.1) indicates that only the part of Eq. (23), Ref. 2, even in U_1, U_2 is required; call this part $\mathcal{G}'(U_1^2, U_2^2, A)$. Then, according to (3.2), we find

$$L^{-1} \phi_{[3^2]}(L,S) = L^{-2} \phi_{[2^3]}(L,S) = F'_4(L,S) = (1 - L^2)^{-1} \mathcal{G}'(L, L^2, S), \tag{4.2}$$

where $\phi_{[3^2]}(L,S)$ and $\phi_{[2^3]}(L,S)$ are given in (2.11).

We now sketch the derivation of the mixed-symmetry generating functions (2.9) and (2.10). For each Young tableau it is straightforward to establish counting rules for weights. Thus for $[2,1^2]$ each set of four distinct m's is counted three times and a pair of equal m's with two distinct m's is counted once (the number of times a set of m's is counted depends only on the partition of p, not on the particular m values.[8] One can then write a recurrence formula for multiplicites of (total) weights

$$\begin{aligned} N^{[2,1^2]}_{l,m} = {} & N^{[2,1^2]}_{l-2,m-l} + N^{[1^3]}_{l-2,m-l} + N^{[2,1]}_{l-2,m-l} \\ & + N^{[1^3]}_{l-2,m+l} + N^{[2,1]}_{l-2,m+l} \\ & + N^{[1^2]}_{l-2,m-2l} + N^{[1^2]}_{l-2,m+2l} \\ & + N^{[2]}_{l-2,m} + 2N^{[1^2]}_{l-2,m} + N^{[1]}_{l-2,m-l} \\ & + N^{[1]}_{l-2,m+l}. \end{aligned} \tag{4.3}$$

$N^{[\lambda]}_{l,m}$ is the multiplicity of the total weight m in the plethysm of symmetry $[\lambda]$ based on the representation R_l; the terms on the right-hand side of (4.3) correspond to choices of weights from the representation R_{l-2} together with zero, one, two or three weights $\pm l$. Using $n^{[\lambda]}_{l,s} = N^{[\lambda]}_{l,m} - N^{[\lambda]}_{l,m+2}$, one

finds the recurrence formula

$$
\begin{aligned}
&n^{[2,1^2]}_{l,s} \\
&= n^{[2,1^2]}_{l-2,s} + n^{[1^3]}_{l-2,s-l} \\
&\quad - n^{[1^3]}_{l-2,l-s-2} + n^{[2,1]}_{l-2,s-l} - n^{[2,1]}_{l-2,l-s-2} \\
&\quad + n^{[1^3]}_{l-2,s+l} + n^{[2,1]}_{l-2,s+l} + n^{[1^2]}_{l-2,s-2l} - n^{[1^2]}_{l-2,2l-s-2} \\
&\quad + n^{[1^2]}_{l-2,s+2l} + n^{[2]}_{l-2,s} + 2n^{[1^2]}_{l-2,s} \\
&\quad + n^{[1]}_{l-2,s-l} - n^{[1]}_{l-2,l-s-2} + n^{[1]}_{l-2,s+l} \,. \qquad (4.4)
\end{aligned}
$$

It is understood that $n^{[\lambda]}_{l,s}$ vanishes if either subscript is negative. Now multiply (4.4) by $L^{\,l}S^{\,s}$ and sum over l and s. The result is a recurrence formula

$$
\begin{aligned}
&\phi_{[2,1^2]}(L,S) \\
&= L^{\,2}\phi_{[2,1^2]}(L,S) + L^{\,2}S^{\,2}\phi_{[1^3]}(LS,S) - L^{\,2}\phi_{[1^3]}(LS,S^{-1}) \\
&\quad + L^{\,2}S^{\,2}\,\phi_{[2,1]}(LS,S) - L^{\,2}\,\phi_{[2,1]}(LS,S^{-1}) \\
&\quad + L^{\,2}S^{-2}\,\phi_{[1^3]}(LS^{-1},S) \\
&\quad + L^{\,2}S^{-2}\,\phi_{[2,1]}(LS^{-1},S) + L^{\,2}S^{\,4}\,\phi_{[1^2]}(LS^{\,2},S) \\
&\quad - L^{\,2}S^{\,2}\,\phi_{[1^2]}(LS^{\,2},S^{-1}) \\
&\quad + L^{\,2}S^{-4}\,\phi_{[1^2]}(LS^{-2},S) + L^{\,2}\,\phi_{[2]}(L,S) \\
&\quad + 2L^{\,2}\,\phi_{[1^2]}(L,S) + L^{\,2}S^{\,2}\,\phi_{[1]}(LS,S) \\
&\quad - L^{\,2}\,\phi_{[1]}(LS,S^{-1}) + L^{\,2}S^{-2}\,\phi_{[1]}(LS^{-1},S) \,. \qquad (4.5)
\end{aligned}
$$

It is understood that negative power of S are to be discarded on the right-hand side of (4.5); this may be done with the help of Eq. (7.6) of Ref. 4.

Relation (4.5) expresses the generating function $\phi_{[2,1^2]}$ of (2.9) in terms of generating functions of lower plethysms which are all known.

The derivation of $\phi_{[2,1^2]}(L,S)$ described above can be repeated for any plethysm. It turns out, however, that this is not always the most practicable way of deriving a fixed plethysm generating function. The generating function for a plethysm with a rectangular Young tableau can be found from an appropriate fixed-l plethysm; thus the method used to derive Eq. (4.2) gives very simply the result

$$
\phi_{[2^2]}(L,S) = \frac{L\mathcal{G}(0,L,S)}{1-L} \,, \qquad (4.6)
$$

in agreement with (2.10); the function $\mathcal{G}$ in (4.6) is given by Eq. (23) of Ref. 2.

It is also possible to find a generating function by inspecting plethysms of the symmetry in question for a number of low values of l (method of elementary multiplets[9]). The plethysms are found from $SU(l+1) \supset SU(2)$ branching rules.

A simple label-counting argument shows that the number of denominator factors in a fixed-plethysm generating function equals

$$
r + (p-1)\,b \,, \qquad (4.7)
$$

where r is the number of representation labels (rank) and b is the number of internal labels of the group in question. For SU(2), one has $r = b = 1$; therefore the number of denominator factors is p, the number of boxes in the Young tableau of the plethysm. For SU(3) one has $r = 2$, $b = 3$, and the number of denominator factors is $3p - 1$.

The symmetry (3.7), which relates symmetric to antisymmetric plethysms, can be used to simplify a result of Ref. 4. Equation (7.5) of Ref. 4, satisfied by generating functions $\phi_{[1^p]}(L,S)$ for antisymmetric plethysms implies an analogous equation for generating functions $\phi_{[\,p\,]}(L,S)$ for symmetric plethysms. It turns out to be

$$
\begin{aligned}
&(1-L^{\,2})\,\phi_{[\,p\,]}(L,S) \\
&\quad = LS^{\,p}\,\phi_{[\,p-1\,]}(LS,S) + LS^{-p}\,\phi_{[\,p-1\,]}(L/S,S) \\
&\quad\quad - LS^{\,p-2}\,\phi_{p-1}(LS,1/S) \\
&\quad\quad + \phi_{[\,p-2\,]}(L,S) + \delta_{p0}(1+L) + \delta_{p1} \,, \qquad (4.8)
\end{aligned}
$$

which replaces the much more complicated (7.10) of Ref. 4. Negative powers of S are to be discarded from the expansion of the right-hand side of (4.8); $\phi_{[\,p\,]}(L,S) = 0$ if p is negative. Similarly, Eq. (7.9) of Ref. 4 is equivalent to the much simpler

$$
\begin{aligned}
n^{[\,p\,]}_{ls} &= n^{[\,p\,]}_{l-2,s} + n^{[\,p-1\,]}_{l-1,s-l-p+1} \\
&\quad + n^{[\,p-1\,]}_{p-1,l-l,s+l+p-1} - n^{[\,p-1\,]}_{l-l,l+p-s-3} \\
&\quad + n^{[\,p-1\,]}_{1,s} + \delta_{p0}\,\delta_{l0}\delta_{s0} + \delta_{p0}\delta_{l1}\delta_{s0} + \delta_{p1}\delta_{l0}\delta_{s0} \,. \qquad (4.9)
\end{aligned}
$$

$n^{[\,p\,]}_{ls} = 0$ if any subscript or superscript is negative.

5. CONCLUDING REMARKS

(1) Fixed-l generating functions may be reinterpreted as generating functions for fixed plethysms of type $[\,p\,]$ or $[1^p]$ [Eq. (3.10)]. The form in which they are given by Franklin and Sylvester is inconvenient because their series expansions contain negative terms and consequent cancellations. They are being rederived in completely positive form.[10]

(2) Fixed-plethysm and fixed-l generating functions are not the only kinds one might define. For example consider

$$
\mathcal{S}(P,L,S) = \sum_{p,l,s=0}^{\infty} n^{[\,p\,]}_{ls}\, P^{\,p} L^{\,l} S^{\,s} \,, \qquad (5.1)
$$

the generating function for all symmetric plethysms. It contains the fixed-l generating function (1.3) as the coefficient of $L^{\,l}$ and the fixed-plethysm generating function $\phi_{[\,p\,]}(L,S)$ as the coefficient of $P^{\,p}$.

Multiplying Eq. (4.8) by $P^{\,p}$ and summing over p leads to a functional equation satisfied by $\mathcal{S}(P,L,S)$,

$$
\begin{aligned}
&(1-L^{\,2}-P^{\,2})\,\mathcal{S}(P,L,S) + \frac{PL}{S}\,\mathcal{S}\left(PS,LS,\frac{1}{S}\right) \\
&\quad - PLS\mathcal{S}(PS,LS,S) - \frac{PL}{S}\,\mathcal{S}\left(\frac{P}{S},\frac{L}{S},S\right) \\
&\quad = 1 + L + P \,. \qquad (5.2)
\end{aligned}
$$

Negative powers of S are to be discarded from the expansion of the left-hand side of (5.2). The symmetry [Eq. (3.10)] of $\mathcal{S}(P,L,S)$ in P and L is obvious in (5.2). The solution of (5.2) would contain all the results hitherto obtained for symmetric and antisymmetric plethysms, and much more.

The generating function

$$
\mathcal{A}(P,L,S) = \sum_{p,l,s=0}^{\infty} n^{[1^p]}_{ls}\, P^{\,p} L^{\,l} S^{\,s} \,, \qquad (5.3)
$$

for all antisymmetric plethysms is related to $\mathcal{S}(P,L,S)$ because of (3.7). We find

$$
\mathcal{A}(P,L,S) = L^{-1}[\mathcal{S}(PL,L,S) - 1] \,. \qquad (5.4)
$$

In deriving (5.4) one has to take into account also the trivial

relation $\phi_{[1^0]}(L,S)=\phi_{[0]}(L,S)=(1-L)^{-1}$ excluded in (3.7).

Besides fixed-l and fixed-plethysm generating functions one may define symmetric and antisymmetric fixed-s generating functions defined by the summations

$$\sum_{p,l=0}^{\infty} n_{ls}^{[p]} P^p L^l \quad \text{and} \quad \sum_{p,l=0}^{\infty} n_{ls}^{[1^p]} P^p L^l . \tag{5.5}$$

They are the coefficients of S^s in $\mathscr{S}(P,L,S)$ and $\mathscr{A}(P,L,S)$, respectively. No closed expression is known for such generating functions.

(3) Fixed-plethysm generating functions are not restricted to the group SU(2). For example

$$F_{[2]}(A,B,a,b) = \frac{1+ABab+A^2Bab^2+AB^2a^2b}{(1-Aa^2)(1-Bb^2)(1-A^2b^2)(1-B^2a^2)(1-AB)},$$

and (5.6)

$$F_{[1^2]}(A,B,a,b) = \frac{Ab+Ba+ABab+A^2B^2a^2b^2}{(1-Aa^2)(1-Bb^2)(1-A^2b^2)(1-B^2a^2)(1-AB)},$$

are generating functions for the symmetric and antisymmetric parts of the Clebsch–Gordan product of two equal representations of SU(3). The dummies A,B play the role of L in the case of SU(2); they carry the labels of the representations being multiplied; a,b play the role of S and carry the labels of the representations into which the product decomposes.

(4) Using the well known relation[4] between Young tableaus and irreducible representations of the groups SU(n), one can translate the symmetries (3.1) and (3.4) into the equality of dimensions of certain representations of SU(n) groups of different ranks. Denoting a representation of SU(n) by non-negative integers λ_i, $i=1,\ldots,n$, equal to the number of *columns* of length i in the corresponding tableau, one concludes from three of the six relations (3.1) that the representations

$$\begin{aligned} &\lambda_i = y\delta_{ix}, \quad i=1,2,\ldots,x+z-1,\\ &\lambda_i = z\delta_{iy}, \quad i=1,2,\ldots,y+x-1,\\ &\lambda_i = x\delta_{iz}, \quad i=1,2,\ldots,z+y-1, \end{aligned} \tag{5.7}$$

of the groups SU($x+z$), SU($y+x$), and SU($z+y$) have the same dimension for any positive integers x,y,z. For example, putting $x=1$, $y=2$, $z=3$, one gets the representations (0010), (200), (03) of SU(5), SU(4), SU(3), all being of dimension ten.

Similarly from (3.4), one finds that the representations

$$\begin{aligned} &\lambda_i = (x-1)\delta_{i1}+\delta_{iy}, \quad i=1,2,\ldots,y+z-1,\\ &\lambda_i = (y-1)\delta_{i1}+\delta_{ix}, \quad i=1,2,\ldots,x+z-1, \end{aligned} \tag{5.8}$$

of SU($y+z$) and SU($x+z$) have the same dimension for any non-negative x,y,z. An example is the coincidence of dimensions for (101) of SU(4) and (21) of SU(3).

(5) It is also possible to define fixed-plethysm generating functions based on reducible representations. Rather than discuss the general problem we mention here some simple examples for the group SU(2). Define

$$\phi_{[\lambda]}(L_1,L_2,S) = \sum_{l_1,l_2,s=0}^{\infty} n_{l_1l_2s}^{[\lambda]} L_1^{l_1} L_2^{l_2} S^s , \tag{5.9}$$

where $n_{l_1l_2s}^{[\lambda]}$ is the multiplicity of R_s in the plethysm $[\lambda]$ based on the reducible representation $R_{l_1}\oplus R_{l_2}$. Each $\phi_{[\lambda]}$ is a sum of generating functions corresponding to all possible splittings of the Young tableau $[\lambda]$ into two disconnected tableaus. For that purpose the trivial tableau $\cdot$ with zero boxes has to be taken into account. The lowest cases then are

$\phi_{[1]}(L_1,L_2,S)$

$$= \cdot\,\square + \square\,\cdot = \frac{1}{(1-L_1)(1-L_2S)} + \frac{1}{(1-L_1S)(1-L_2)}, \tag{5.10}$$

$\phi_{[2]}(L_1,L_2,S)$

$$= \cdot\,\square\square + \square\square + \square\square\,\cdot = \frac{1}{(1-L_1)(1-L_2^2)(1-L_2S^2)} + \frac{1}{(1-L_1L_2)(1-L_1S)(1-L_2S)} + \frac{1}{(1-L_1^2)(1-L_1S^2)(1-L_2)}, \tag{5.11}$$

$\phi_{[1^2]}(L_1,L_2,S)$

$$= \cdot\,\begin{array}{|c|}\hline\ \\\hline\ \\\hline\end{array} + \square\square + \begin{array}{|c|}\hline\ \\\hline\ \\\hline\end{array}\,\cdot = \frac{L_2}{(1-L_1)(1-L_2^2)(1-L_2S^2)} + \frac{1}{(1-L_1L_2)(1-L_1S)(1-L_2S)} = \frac{L_1}{(1-L_1^2)(1-L_1S^2)(1-L_2)}, \tag{5.12}$$

$\phi_{[3]}(L_1,L_2,S)$

$$= \cdot\,\square\square\square + \square\square\square + \square\square\square + \square\square\square\,\cdot = \frac{1+L_2^3S^3}{(1-L_1)(1-L_2^2)(1-L_2^2S^2)(1-L_2S^3)} + \frac{1+I}{(1-L_2^2)(1-L_1^2L_2)(1-L_1S)(1-L_2S^2)} + \{L_1\longleftrightarrow L_2\}, \tag{5.13}$$

$\phi_{[1^3]}(L_1,L_2,S)$

$$= \cdot\,\begin{array}{|c|}\hline\ \\\hline\ \\\hline\ \\\hline\end{array} + \square\,\begin{array}{|c|}\hline\ \\\hline\ \\\hline\end{array} + \begin{array}{|c|}\hline\ \\\hline\ \\\hline\end{array}\,\square + \begin{array}{|c|}\hline\ \\\hline\ \\\hline\ \\\hline\end{array}\,\cdot = \frac{L_2^2+L_2^5S^3}{(1-L_1)(1-L_2^2)(1-L_2^2S^2)(1-L_2S^3)} + \frac{L_2+L_1L_2^2S}{(1-L_2^2)(1-L_1^2L_2)(1-L_1S)(1-L_2S^2)} + \{L_1\longleftrightarrow L_2\}, \tag{5.14}$$

$\phi_{[2,1]}(L_1,L_2,S)$

$$= \cdot\cdot\begin{array}{|c|c|}\hline\ &\ \\\hline\ \\\hline\end{array} + (\ \square\,\begin{array}{|c|}\hline\ \\\hline\ \\\hline\end{array} + \square\square\square\ ) + (\ \begin{array}{|c|}\hline\ \\\hline\ \\\hline\end{array}\,\square + \square\square\square\ ) + \begin{array}{|c|c|}\hline\ &\ \\\hline\ \\\hline\end{array}\,\cdot = \frac{L_2S}{(1-L_1)(1-L_2^2)(1-L_2S^2)(1-L_2S)}$$

$$+\frac{1+L_2+L_1L_2S+L_1L_2^2S}{(1-L_1S)(1-L_2^2)(1-L_1^2L_2)(1-L_2S^2)}$$

$$+\{L_1\longleftrightarrow L_2\}\,. \tag{5.15}$$

In (5.10)–(5.15) we indicate symbolically the correspondence between a split Young tableau and its contribution to $\phi_{\{\lambda\}}(L_1,L_2,S)$. In each split tableau the left factor refers to the variable L_1, the right one refers to L_2. The variables S from each factor have been combined by means of the generating function for SU(2) Clebsch–Gordan series (Eq. (10) of Ref. 11].

(6) Finally, one may wonder why the symmetries discussed in Sec. 3 are limited to plethysms with Young tableaux of two types only (rectangular and hook shaped). Is this not a particular case of a more general symmetry? We have only a partial answer to that question: The symmetries [cf. (3.1), (3.4a), (3.7a)] imply relations between two generating functions [cf. (3.2), (3.4b), (3.7b)]. If the two generating functions in each of these relations are to differ by a power of the variable L, then the symmetries we describe are the only ones possible.

Note added in proof: B. G. Wybourne has pointed out to us that Eqs. (3.7a) and (3.10) were known to Murnaghan. See F. D. Murnaghan, Proc. Nat. Acad. Sci. **37**, 439 (1951) and **40**, 832 (1954); also B. G. Wybourne, J. Math phys. **10**, 467 (1969).

ACKNOWLEDGMENTS

We are grateful to J. McKay, R. Stanley, and B. Wybourne for helpful suggestions.

[1] J. J. Sylvester and F. Franklin, Am. J. Math. **2**, 223, 293 (1879); F. Franklin, Am. J. Math. **3**, 178 (1880); J. J. Sylvester, Am. J. Math. **4**, 41 (1881). Sylvester and Franklin refer to Cayley's first and second methods in deriving their generating method, but do not given an explicit reference to Cayley.

[2] R. Gaskell, A. Peccia, and R. T. Sharp, J. Math. Phys. **19**, 727 (1978).

[3] J. Patera, R. T. Sharp, and P. Winternitz, J. Math. Phys. **19**, 2362 (1978); P. E. Desmier and R. T. Sharp, J. Math. Phys. **20**, 74 (1979).

[4] J. Patera and R. T. Sharp, J. Phys. A Math. Nucl. Gen. **13**, 397 (1980).

[5] P. Gard and N. B. Backhouse, J. Phys. A Math. Nucl. Gen. **7**, 1793 (1974); M. A. Rashid, J. Phys. A Math. Nucl. Gen. **10**, L135 (1977); V. V. Mikhailov, J. Phys. A Math. Nucl. Gen. **12**, 2329 (1979).

[6] J. Patera and R. T. Sharp, in *Recent Adlvances in Group Theory and their Application to Spectroscopy*, edited by J. Donini (Plenum, New York, 1979).

[7] W. McKay and J. Patera, *Tables of Dimensions, Indices and Branching Rules of Representations of Simple Lie Algebras* (Dekker, New York, 1980) (to be published).

[8] D. E. Knuth, Pac. J. Math. **34**, 709 (1970).

[9] V. Bargmann and M. Moshinsky, Nucl. Phys. **23**, 177 (1961); V. S. Devi and M. Moshinsky, J. Math. Phys. **10**, 455 (1969); R. T. Sharp and C. S. Lam, J. Math. Phys. **10**, 2033 (1969).

[10] J. Bystrický, J. Patera, and R. T. Sharp (in preparation).

[11] J. Patera and R. T. Sharp, in *Lecture Notes in Physics* (Springer, New York, 1979), Vol. 94, pp. 175–83.

General U(N) raising and lowering operators

M. D. Gould [a)]
Department of Mathematical Physics, University of Adelaide, G.P.O. Box 498, Adelaide, South Australia, 5001

(Received 21 September 1979; accepted for publication 10 December 1979)

It is the aim of this paper to obtain the general form of U(n) raising and lowering operators. The raising and lowering operators constructed previously by several authors are then compared. The Hermiticity properties of these operators are also investigated. The methods presented extend, with trivial modifications, to the orthogonal groups.

PACS numbers: 02.20.Qs, 02.20.Rt

1. INTRODUCTION

In the 1960's a great deal of interest was generated in extending the angular momentum techniques of Wigner and Racah to the general unitary and orthogonal groups. This has led to the introduction of the group-theoretical concept of operators that lower or raise the highest weights of representations of a subgroup contained in an irreducible representation of the group. Such operators may be regarded as a generalization of the raising and lowering operators $L_{\pm}$ appearing in the theory of angular momenta.

Such operators were first constructed for the unitary groups by Nagel and Moshinsky[1] who applied them to the analysis of many body problems.[2] Subsequently raising and lowering operators were constructed for the orthogonal groups by Pang and Hecht[3] and Wong.[4] Following the definition of Nagel and Moshinsky[1] the lowering (raising) operators shall be polynomials of the group generators that, when acting on a basis vector of an irreducible representation of the group which is of given weight with respect to the subgroup, lower (raise) the weight. Furthermore, they shall, when acting on a basis vector of highest weight of the subgroup transform it into a basis vector of highest weight of a lowered (raised) irreducible representation of the subgroup.

It is important to note that the raising and lowering operators for a subgroup are essentially only defined by their action on a state of highest weight for the subgroup. We see therefore that such operators are not unique. Hence the raising and lowering operators constructed previously for O(n) and U(n) are only one particular solution to the problem.

Recently Bincer[5] obtained raising and lowering operators for the orthogonal and unitary groups using methods based on the characteristic identities satisfied by the infinitesimal generators of the group.[6] These operators of Bincer appear in a compact product form which is useful for manipulations. In subsequent independent work of the author[7] an alternative set of raising and lowering operators for O(n) and U(n) were constructed using techniques similar in content to Bincer's. Our operators, like those of Bincer, may also be written in a compact product form.

Recent work of the author[8] shows how these techniques may be extended to obtain the matrix elements of the group generators. Central in this approach is the concept of "simultaneous shift operator" which shifts the representation labels of U(n) and each of its canonical subgroups in a certain prescribed way. These operators may therefore be regarded as generalizations of the raising and lowering operators discussed in this paper. A general procedure for constructing raising and lowering operators for a general semi-simple Lie group is discussed in Ref. 7.

It is the aim of the present paper to investigate the connection between the various raising and lowering operators. We shall obtain the general form for a raising (resp. lowering) operator for U(n). It shall be shown that the raising and lowering operators constructed in Refs. 1 and 5 are identical. By contrast the operators constructed in Ref. 7 are shown to be different. The behavior of these raising and lowering operators under Hermitian conjugation is also investigated. It shall be shown that the raising and lowering operators constructed in Ref. 7 are unique with respect to the property of being Hermitian conjugates of one another.

The techniques employed in this paper are similar in content to Bincer's except for our use of the U(n) contragredient identity. This enables raising operators for U(n) (which are absent in the work of Bincer) to be constructed in analogy with the lowering operators.

Although we shall only discuss the unitary group it is clear that the arguments extend to O(n) with little modification. Also it is of interest to extend these results to the noncompact groups O(n,1) and U(n,1). Patera[9] has shown that the Nagel–Moshinsky operators are also a suitable choice for U(n,1) while Wong[10] has shown that his operators for O(n) extend to O(n,1). Wong and Yeh[11] have also recently investigated the extension of Bincer's operators to O(n,1). One may follow through their derivation to conclude that the raising and lowering operators constructed in Ref. 7 also extend to O(n,1) and U(n,1) as does any general raising (resp. lowering) operator for O(n) and U(n) (defined in the sense of Nagel and Moshinsky).

2. THE CHARACTERISTIC IDENTITIES

The generators a^i_j of the Lie group U(n) may be assembled into a square matrix a which, on an irreducible representation of the group (finite or infinite dimensional) with highest weight $\lambda = (\lambda_1, \lambda_2, ..., \lambda_n)$, satisfies the polynomial identity[6]

$$\prod_{r=1}^{n} (a - \lambda_r - n + r) = 0 .$$

This polynomial identity may be written in a representation

[a)]Present address: School of Physical Sciences, The Flinders University of South Australia, South Australia, 5042.

 0022-2488/81/020267-04$1.00

independent way as

$$\prod_{r=1}^{n}(a-\alpha_r)=0\,, \tag{1}$$

where the operators α_r lie in an algebraic extension of the center of the enveloping algebra.[12] (Note that any symmetric combination of the α_r necessarily lies in the center Z. In particular the coefficients of the identity (1) are central elements.) The eigenvalues of these operators on any representation admitting an infinitesimal character χ_λ (or equivalently on any irreducible representation with highest weight λ) are given by

$$\chi_\lambda(\alpha_r)=\lambda_r+n-r\,.$$

Associated with the matrix a is its "contragredient" $\bar{a}$ with entries given by

$$\bar{a}_j{}^i=-a^i_j\,.$$

The matrix $\bar{a}$ satisfies the polynomial identity

$$\prod_{r=1}^{n}(\bar{a}-\bar{\alpha}_r)=0\,, \tag{2}$$

where the roots $\bar{\alpha}_r$ are related to the α_r by $\bar{\alpha}_r=n-1-\alpha_r$.

By virtue of the identities (1) and (2) one may construct projection operators

$$P[r]=\prod_{l\neq r}\left(\frac{a-\alpha_l}{\alpha_r-\alpha_l}\right),$$

$$\bar{P}[r]=\prod_{l\neq r}\left(\frac{\bar{a}-\bar{\alpha}_l}{\bar{\alpha}_r-\bar{\alpha}_l}\right),$$

which enables arbitrary functions of the matrices a and $\bar{a}$ to be defined by setting

$$P(a)=\sum_{r=1}^{n}P(\alpha_r)P[r]\,,$$

$$P(\bar{a})=\sum_{r=1}^{n}P(\bar{\alpha}_r)\bar{P}[r]\,.$$

The projection operators $P[r]$ and $\bar{P}[r]$ are well defined elements of an extension of the enveloping algebra although they need not be defined on representations where the eigenvalues of some α_r and α_k $(r\neq k)$ coincide. This however cannot occur on finite dimensional representations [nor on unitary representations of the noncompact groups U(p,q)] and hence, for the applications we have in mind, the projectors $P[r]$ and $\bar{P}[r]$ are always well defined.

If ψ (resp. $\psi^\dagger$) is a vector (resp. contragredient vector) operator of U(n) then we may resolve ψ and $\psi^\dagger$ into shift vectors[6]

$$\psi=\sum_{r=1}^{n}\psi[r]\,,\quad \psi^\dagger=\sum_{r=1}^{n}\psi^\dagger[r]\,,$$

which alter the U(n) representation labels according to

$$\alpha_k\psi[r]=\psi[r](\alpha_k+\delta_{kr})\,,$$

$$\alpha_k\psi^\dagger[r]=\psi^\dagger[r](\alpha_k-\delta_{kr})\,.$$

(Note that this shift property also extends to infinite dimensional representations.) Such shift operators may be constructed by applying the projectors $P[r]$ and $\bar{P}[r]$:

$$\psi[r]=P[r]\psi=\psi\bar{P}[r]\,,$$

$$\psi^\dagger[r]=\bar{P}[r]\psi^\dagger=\psi^\dagger P[r]=(\psi[r])^\dagger\,.$$

It was shown in Ref. 7 (see also Ref. 13) that if v_0 is an arbitrary maximal weight state of U(n) with highest weight $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_n)$ then

$$\begin{aligned}&P[r]^i{}_j v_0=0\,,\quad \text{for } j>r \text{ and } i \text{ arbitrary}\,,\\ &\bar{P}[r]_j{}^i v_0=0\,,\quad \text{for } j<r \text{ and } i \text{ arbitrary}\,.\end{aligned} \tag{3}$$

As a consequence of (3) we see that on the maximal weight vector v_0 the shift vectors $\psi[r]$ and $\psi^\dagger[r]$ must satisfy

$$\begin{aligned}&\psi[r]^i v_0=0\,,\quad \text{for } i<r\,,\\ &\psi^\dagger[r]_i v_0=0\,,\quad \text{for } i>r\,.\end{aligned} \tag{4}$$

Using the definition of vector operator equation (4) then implies that $\psi[r]^r v_0$ and $\psi^\dagger[r]_r v_0$ are maximal weight states of weight $\lambda+\Delta_r$ and $\lambda-\Delta_r$, respectively.

3. GENERAL RAISING AND LOWERING OPERATORS

It is our aim here to determine the general form for U(n) raising and lowering operators and to compare the operators constructed in Refs. 1, 5, and 7.

Throughout we shall let ψ denote the U(n) vector operator with components $\psi^i=a^i{}_{n+1}$ $(i=1,\ldots,n)$ whose contragredient has components $\psi^\dagger_i=a^{n+1}{}_i$. The operators $\psi[r]^r$ and $\psi^\dagger[r]_r$ will shift highest weight vectors of U(n) in a finite dimensional irreducible representation of U$(n+1)$. These are the U(n) raising and lowering operators constructed in Ref. 7. For convenience we denote them by $\psi^r{}_n$ and $\psi^{\dagger r}{}_n$, respectively. Finally we denote a maximal weight state of U(n) [i.e., a semi-maximal state of U$(n+1)$] by the pattern $\left|{\lambda_{i,n+1}\atop\lambda_{i,n}}\right\rangle$. Here $\lambda_{i,n+1}$ and $\lambda_{i,n}$, as usual, refer to highest weights of finite dimensional irreducible representations of U$(n+1)$ and U(n), respectively.

Suppose now that $R^r{}_n$ and $L^r{}_n$ are arbitrary raising and lowering operators of U(n) effecting the shifts

$$\left|{\lambda_{i,n+1}\atop\lambda_{i,n}}\right\rangle\longrightarrow\left|{\lambda_{i,n+1}\atop\lambda_{i,n}+\delta_{ir}}\right\rangle,$$

$$\left|{\lambda_{i,n+1}\atop\lambda_{i,n}}\right\rangle\longrightarrow\left|{\lambda_{i,n+1}\atop\lambda_{i,n}-\delta_{ir}}\right\rangle,$$

respectively. According to Nagel and Moshinsky such operators are of the form

$$R^r{}_n=h(a)^r{}_j a^j{}_{n+1}\,, \tag{5}$$

$$L^r{}_n=a^{n+1}{}_j h(a)^j{}_r\,, \tag{6}$$

for a suitable polynomial $h(x)$.

Resolving $\psi^\dagger_i=a^{n+1}{}_i$ into its distinct shift components allows us to write Eq. (6) in the form

$$L^r{}_n=\sum_{l=1}^{n}\psi^\dagger[l\,]_r h(\alpha_l)\,. \tag{7}$$

Now acting on the state $\left|{\lambda_{i,n+1}\atop\lambda_{i,n}}\right\rangle$ the operators $\psi^\dagger[l\,]_r$, $l<r$, vanish by virtue of Eq. (4). Hence, acting on the state $\left|{\lambda_{i,n+1}\atop\lambda_{i,n}}\right\rangle$, Eq. (7) reduces to

$$L^r{}_n\left|{\lambda_{i,n+1}\atop\lambda_{i,n}}\right\rangle=\sum_{l\geq r}\psi^\dagger[l\,]_r h(\alpha_l)\left|{\lambda_{i,n+1}\atop\lambda_{i,n}}\right\rangle. \tag{8}$$

However each operator $\psi^\dagger[l\,]_r$ (for $l>r$) effects the shifts $\lambda_{i,n}\to\lambda_{i,n}-\delta_{il}$. Hence in order to obtain the required shifts

we must have $h(\alpha_l) = 0$ for $l > r$. Hence $h(x)$ must be divisible by the polynominal $\Pi_{l>r}(x - \alpha_l)$. In the limiting case where $h(x) = \Pi_{l>r}(x - \alpha_l)$ we obtain the operators constructed by Bincer.[5] We have shown then that a general lowering operator is of the form

$$\psi^r{}_n\beta + \sum_{l<r}\psi^\dagger[l]_r\beta_l ,$$

where β and β_l are invariant multiplies of the identity [or equivalently of the form $a^{n+1}{}_j h(a)^j{}_r$ where $h(x)$ is divisible by $\Pi_{l>r}(x - \alpha_l)$].

Note also that the Nagel–Moshinsky operators $L^r{}_n$ are of the form $a^{n+1}{}_j h(a)^j{}_r$ for a suitable polynomial $h(x)$ which, according to the remarks above, is divisible by the polynomial $\Pi_{l>r}(x - \alpha_l)$. However it is well known that the Nagel–Moshinsky operators $L^r{}_n$ are homogeneous of degree $n - r + 1$ in the group generators from which it follows that $h(x)$ is of degree $n - r$ which is precisely the degree of $\Pi_{l>r}(x - \alpha_l)$. Accordingly we must have

$$h(x) = c\prod_{l>r}(x - \alpha_l) ,$$

where c is a constant dependent on the roots α_l.

From Eq. (8) acting on the state $|^{\lambda_{i,n+1}}_{\lambda_{i,n}}\rangle$ the Nagel–Moshinsky operators $L^r{}_n$ reduce to

$$L^r{}_n \left|\begin{matrix}\lambda_{i,n+1}\\ \lambda_{i,n}\end{matrix}\right\rangle = \psi^{\dagger r}{}_n c\prod_{l>r}(\alpha_r - \alpha_l)\left|\begin{matrix}\lambda_{i,n+1}\\ \lambda_{i,n}\end{matrix}\right\rangle .$$

By comparing the normalization of our lowering operators $\psi^{\dagger r}{}_n$ with the normalization of the $L^r{}_n$ the constant c may be determined. By this means we obtain

$$c = \prod_{l>r}\left(\frac{\alpha_r - \alpha_l - 1}{\alpha_r - \alpha_l}\right).$$

Hence the Nagel–Moshinsky lowering operators may be written

$$L^r{}_n = a^{n+1}{}_j g(a)^j{}_r , \tag{9}$$

where $g(x)$ is the polynomial

$$g(x) = \prod_{l>r}(x - \alpha_l)\left(\frac{\alpha_r - \alpha_l - 1}{\alpha_r - \alpha_l}\right).$$

These are essentially the operators constructed by Bincer[5] (up to multiplication by an invariant multiple of the identity).

Equation (9) is just one representation of the Nagel–Moshinsky operators. Expanding the polynomial $g(x)$ into powers of x we may write

$$g(x) = \sum_{k=0}^{n-r} X^{n-r-k} S_k \prod_{l>r}\left(\frac{\alpha_r - \alpha_l - 1}{\alpha_r - \alpha_l}\right),$$

where S_k is a polynomial in the $\alpha_l (l > r)$. By replacing the α_r with Lie algebra elements $\epsilon_r = a^r{}_r + n - r$ we obtain the Nagel–Moshinsky operators in their original form.[1]

Using Eq. (4) one may show, by the same techniques, that a general raising operator [see Eq. (5)] is of the form

$$R^r{}_n = h(a)^r{}_j a^j{}_{n+1} ,$$

where $h(x)$ is divisible by the polynomial $\Pi_{l<r}(x - \alpha_l)$. In terms of the matrix $\bar{a}$ a general raising operator may more usefully be written

$$R^r{}_n = a^j{}_{n+1}\, g(\bar{a})_j{}^r ,$$

where $g(x)$ is necessarily divisible by $\Pi_{l<r}(x - \bar{\alpha}_l)$. The limiting case where $g(x) = \Pi_{l<r}(x - \bar{\alpha}_l)$ gives the raising analogue of Bincer's lowering operators (although these operators do not appear in the work of Bincer). The raising operators of Nagel and Moshinsky may be writen

$$\begin{aligned} R^r{}_n &= \Pi_{l<r}\left(\frac{\alpha_r - \alpha_l - 1}{\alpha_r - \alpha_l}\right)(a - \alpha_l)^r{}_i a^i{}_{n+1} \\ &= a^i{}_{n+1}\prod_{l<r}(\bar{a} - \bar{\alpha}_l)_i{}^r\left(\frac{\bar{\alpha}_r - \bar{\alpha}_l - 1}{\bar{\alpha}_r - \bar{\alpha}_l}\right). \end{aligned} \tag{10}$$

4. HERMITICITY PROPERTIES

It was shown by Nagel and Moshinsky that their raising and lowering operators are not Hermitian conjugates [as one may show by comparing Eqs. (9) and (10)]. It is natural then to determine under what conditions the Hermitian conjugate of a raising operator is a lowering operator (and vice versa). We answer this by showing that our raising and lowering operators are unique with respect to the property of being Hermitian conjugates.

In the last section it was shown that a general lowering operator may be written in the form

$$L^r{}_n = \psi^{\dagger r}{}_n\beta + \sum_{l<r}\psi^\dagger[l]_r\beta_l , \tag{11}$$

where β and β_l are constants dependent on the roots α_l. Similarly a general raising operator may be written in the form

$$R^r{}_n = \gamma\psi^r{}_n + \sum_{l>r}\gamma_l\psi[l]^r . \tag{12}$$

Comparing Eqs. (11) and (12) we see that $(L^r{}_n)^\dagger$ cannot possibly be a raising operator unless $\beta_l = 0$ for $l < r$; i.e., unless $L^r{}_n$ is an invariant multiple of $\psi^{\dagger r}{}_n$. (An analogous statement holds for $R^r{}_n$.) Accordingly we see that the raising and lowering operators $\psi^r{}_n$ and $\psi^{\dagger r}{}_n$ (constructed in Ref. 7) are unique with respect to the property of being Hermitian conjugates.

ACKNOWLEDGMENTS

It is my pleasure to thank Professor H.S. Green for his interest and encouragement. The author also acknowledges the financial support of a Commonwealth Post-Graduate Research Award.

[1] J.G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965); Rev. Mex. Fis. **14**, 29 (1965).
[2] M. Moshinsky, *Many-body problems and Other Selected Topics in Theoretical Physics* (Gordon and Breach, New York, 1966).
[3] J.C. Pang and K.T. Hecht, J. Math. Phys. **8**, 1233 (1967).
[4] M.K.F. Wong, J. Math. Phys. **8**, 1899 (1967).
[5] A.M. Bincer, J. Math. Phys. **18**, 1870 (1977); **19**, 1173 (1978); **19**, 1179 (1978).
[6] H.S. Green, J. Math. Phys. **12**, 2106 (1971); A.J. Bracken and H.S. Green, J. Math. Phys. **12**, 2099 (1971).
[7] M.D. Gould, "An infinitesimal approach to semi-simple Lie Groups and raising and lowering operators of O(N) and U(N)," to appear in J. Math. Phys.

[8]M.D. Gould, "On the matrix elements of the U(N) generators," submitted for publication in J. Math. Phys.
[9]J. Patera, J. Math. Phys. **14**, 279 (1973).
[10]M.K.F. Wong, J. Math. Phys. **15**, 25 (1974); M.K.F. Wong and H.Y. Yeh, J. Math. Phys. **16**, 800 (1975).
[11]M.K.F. Wong and Hsin-Yang Yeh, J. Math. Phys. **4**, 629 (1979).
[12]D.M. O'Brien, A. Cant, and A.L. Carey, Ann. Inst. Henri Poincaré A **26**, 405 (1977); M.D. Gould, Ann. Inst. Henri Poincaré A **24**, 51 (1978).
[13]M.D. Gould, J. Aust. Math. Soc. B **20**, 401 (1978).

Expansion of a function about a displaced center

M. A. Rashid
Department of Mathematics, Ahmadu Bello University, Zaria, Nigeria

(Received 11 October 1979; accepted for publication 18 January 1980)

An extremely simple closed expression is obtained for the coefficients which appear in the expansion of a function of a special type about a displaced center. A conjecture about the vanishing of a certain coefficient which appear in the expansion of a Slater-type orbital about a displaced center is also proved.

PACS numbers: 02.30.Bi, 02.30.Mv

1. INTRODUCTION

Sharma[1] has obtained a general and closed expression for the coefficients which appear in the expansion of functions of a special type about a displaced center. In our review of Sharma's work,[2] we pointed out that the four summations in his expression for the coefficients b_v $(slLM)$ could very easily be reduced to two. In the present paper, we show that we can eliminate three of his summations to obtain a simple expression for these expansion coefficients. Also we utilize our expression to prove a conjecture about the vanishing of the expansion coefficients $F_{k'k}(NlLM)$ appearing in the expansion of a Slater type orbital whenever $k' > l + N$.

To facilitate comparison with Sharma's work, we have used the notations of his paper throughout.

2. DISPLACED CENTER EXPANSION

Let (r,θ,ϕ) and (R,Θ,Φ) be the spherical polar coordinates of the same point P with respect to parallel set of coordinate axes at A and B respectively where B is at a distance a from A along the positive z axis. In this situation $\phi = \Phi$. We try to express a function

$$(1/R)f_{NL}(R)Y_L^M(\Theta,\Phi),$$

which is centered at B as

$$(1/R)f_{NL}(R)Y_L^M(\Theta,\Phi) = \sum_l (1/r)\alpha_l(NLM\,|a,r)Y_l^m(\theta,\phi), \tag{1}$$

i.e., as a linear combination of functions centered at A.

The problem is to find a simple closed expression for the coefficients $\alpha_l(NLM\,|a,r)$ which appear in the Eq. (1) above. Sharma[1] obtained the following closed expression for these coefficients[3]:

$$\alpha_l(NLM\,|a,r) = (1/a)\sum_v (r/a)^{2v-l}\sum_s b_v(slLM)\int_{|a-r|}^{a+r} f_{NL}(R)(R/a)^{2s-L}\,dR, \tag{2}$$

where

$$\begin{aligned} b_v(slLM) = {} & \frac{(-1)^L}{2}\left(\frac{(2L+1)(2l+1)(L+M)!(l-M)!}{(L-M)!(l+M)!}\right)^{1/2} \\ & \times \sum_{pqp'q'} (-1)^{p+q+p'+q'} 2^{2(p+p'-L-l)}(2L-2p)!(2l-2p')!/[\,p!(L-p)!q! \\ & \times (s-p-q')!(L+M-s-p-q+q')!p'!(l-p')!q'!(v-q-p')!(l-M-v+q-p'-q')!]. \end{aligned} \tag{3}$$

We shall adopt the following convenient conventions throughout the paper:

(i) The factorial notation is used even when the argument is not necessarily a nonnegative integer. Quite generally $x! = \Gamma(x+1)$.

(ii) The ranges of the summations appearing in the various expressions are omitted since these are fixed by the nonnegativity of the arguments of the factorials with *integral* arguments.

We remark that in Eq. (3) above, the presence of $(L-p)!((l-p)!)$ does not contribute any extra restriction on the ranges of the summations since from $L-M \geqslant 0, q \geqslant 0$, $s-p-q' \geqslant 0$, $L+M-s-p-q+q' \geqslant 0 (l+M \geqslant 0, q \geqslant 0, s-q-p' \geqslant 0, l-M-v+q-p'-q' \geqslant 0)$, we conclude that $L-p \geqslant 0 (1-p \geqslant 0)$. This remark is useful since, the duplication formula

$$(2z)! = \Gamma(2z+1) = (1/\sqrt{\pi})2^{2z}\Gamma(z+\tfrac{1}{2})\Gamma(z+1) = (1/\sqrt{\pi})2^{2z}(z-\tfrac{1}{2})!(z)!, \tag{4}$$

for the gamma functions, enables us to replace $2^{2p-2L}(2L-2p)!/(L-p)!$ and $2^{2p'-2l}(2l-2p')!/(l-p')!$, by $(1/\sqrt{\pi})(L-p-\tfrac{1}{2})!$ and $(1/\sqrt{\pi})(l-p'-\tfrac{1}{2})!$, respectively. This results in an expression for the coefficient b_v $(slLM)$ wherein the *finite*

 0022-2488/81/020271-04$1.00

p,p' summations can be performed using[4]

$$\sum_{\delta}(-1)^{\delta}\frac{(a-\delta)!}{\delta!(b-\delta)!(c-\delta)!}=\frac{(a-b)!(a-c)!}{b!c!(a-b-c)!}, \tag{5}$$

to arrive at

$$b_v(slLM)=\frac{(1)^L}{2\pi}\left(\frac{(2L+1)(2l+1)(L+M)!(l-M)!}{(L-M)!(l+M)!}\right)^{1/2}$$
$$\times\sum_{qq'}(-1)^{q+q'}\frac{(l-\frac{1}{2}-s+q')!(l-\frac{1}{2}-v+q)!(-M-\frac{1}{2}+s+q-q')!(M-\frac{1}{2}+v-q+q')!}{q!q'!(s-q')!(v-q)!(L-M-s-q+q')!(l-M-v+q-q')!(-M-\frac{1}{2}+q)!(M-\frac{1}{2}+q')!} \tag{6}$$

Our aim is to replace the double summation in the above equation by a single summation. First we use[5]

$$\frac{(M-\frac{1}{2}+v-q+q')!}{(M-\frac{1}{2}+q')!(L+M-s-q+q')!(v-q)!}=\sum_{pp'}\frac{(-L-\frac{1}{2}+s+v)!}{(v-q-p)!(-L-\frac{1}{2}+v+s-p)!(L+M-s-v+q'+p)!},$$

and

$$\frac{(-M-\frac{1}{2}+s+q-q')!}{(-M-\frac{1}{2}+q)!(l-M-v+q-q')!(s-q')!}=\sum_{\sigma}\frac{(-l-\frac{1}{2}+s+v)}{\sigma!(s-q'-\sigma)!(-l-\frac{1}{2}+s+v-\sigma)!(l-M-s-v+q+\sigma)!}$$

and perform the q,q' summations utilizing.

$$\sum_{\delta}(-1)^{\delta}\frac{(a+\delta)!}{\delta!(b-\delta)!(c+\delta)!}=(-1)^{b}\frac{a!(a-c)!}{b!(b+c)!(a-b-c)!}, \tag{5'}$$

which is essentially the same as Eq. (5) and can be obtained from it by taking $b-\delta$ as the new b. This results in

$$b_v(slLM)=\frac{(-1)^{L+M}}{2\pi^2}\left(\frac{(2L+1)(2l+1)(L+M)!(l-M)!}{(L-M)!(l+M)!}\right)^{1/2}$$
$$\times(L-\tfrac{1}{2}-s)!(l-\tfrac{1}{2}-v)!(-L-\tfrac{1}{2}+s+v)!(-l-\tfrac{1}{2}+s+v)!$$
$$\times\sum_{\sigma p}\frac{(M-\frac{1}{2}+s-\sigma)!(-M-\frac{1}{2}+v-p)!}{\sigma!p!(s-\sigma)!(v-p)!(L+M-v-\sigma+p)!(l-M-s+\sigma-p)!(-L-\frac{1}{2}+s+v-p)!}$$
$$(-l-\tfrac{1}{2}+s+v-\sigma)!, \tag{7}$$

on making use of

$$(M-\tfrac{1}{2}+s-v+p-\sigma)!(-M-\tfrac{1}{2}-s+v+\sigma-p)!=\Gamma(M+\tfrac{1}{2}+s-v+p-\sigma)!\Gamma(-M+\tfrac{1}{2}-s+v+\sigma+p)$$
$$=\frac{\pi}{\sin\pi(M+\frac{1}{2}+s-v-\sigma+p)}=\pi(-1)^{M+s+v+\sigma+p}. \tag{8}$$

On comparing Eqs. (6) and (7) we note that though Eq. (7) still contains two summations, the number of factorials within the new summations is two less.

Next we replace[6]

$$\frac{(-M-\frac{1}{2}+v-p)!}{(L+M-v-\sigma+p)!(l-M-s+\sigma-p)!(-l-\frac{1}{2}+v+s-\sigma)!},$$

by

$$\frac{1}{(L+l-s-v)!}\sum_{t}(-1)^{t}\frac{(L-\frac{1}{2}-\sigma-t)!}{t!(L+M-v-\sigma+p-t)!(l-\frac{1}{2}+s+v-\sigma-t)!}$$
$$=\frac{1}{(L+l-s-v)!}\sum_{t}(-1)^{L+M+\sigma+t}\frac{(-M-\frac{1}{2}+t)!}{(L+M-\sigma-t)!(-v+p+t)!(-L-l-M-\frac{1}{2}+s+v+t)!},$$

in Eq. (7) and perform the *two* σ and p summations utilizing Eq. (5) and[7]

$$\sum_{\delta}\frac{1}{\delta!(a-\delta)!(b-\delta)!(c-b+\delta)!}=\frac{(a+c)!}{a!b!c!(a+c-b)!}, \tag{9}$$

to obtain

$$b_v(slLM)=\frac{1}{2\pi^2}\left(\frac{(2L+1)(2l+1)(L+M)!(l-M)!}{(L-M)!(l+M)!}\right)^{1/2}\frac{(M-\frac{1}{2})!(L-s-\frac{1}{2})!(l-v-\frac{1}{2})!(-l-\frac{1}{2}+s+v)!}{s!v!(L+l-s-v)!}$$
$$\times\sum_{t}(-1)^{t}\frac{(-M-\frac{1}{2}+t)!(-L-\frac{1}{2}+s+v+t)!}{t!(L+M-t)!(-L-\frac{1}{2}+t)!(-L-l-M-\frac{1}{2}+s+v+t)!}. \tag{10}$$

In the above equation, the coefficient $b_v(slLM)$ has been expressed as a *single* summation though the expression lacks symmetry between L and l. But we can indeed restore it by making use of the type of arguments given above to obtain the final

expression

$$b_v(slLM) = \frac{(-1)^L}{2\pi}\,[(2L+1)(2l+1)(L+M)!(L-M)!(l+M)!(l-M)!]^{1/2} \times \frac{(L-\frac{1}{2}-s)!(l-\frac{1}{2}-v)!}{s!v!(L+l-s-v)!}\sum_t \frac{(-M-\frac{1}{2}+s+v-t)!}{t!(L-M-t)!(l-M-t)!(-M-\frac{1}{2}-t)!(2M+t)!}, \tag{11}$$

which has the obvious symmetries

$$b_v(slLM) = b_v(slL(-M)) = b_s(vLlM)(-1)^{L+l}. \tag{12}$$

The above two equations are the main results of the present paper. In a future publication, we hope to present interesting applications of our results. We conclude this paper by proving that the coefficient $F_{k'k}(NlLM)$ in Sharma's paper indeed vanishes for $k' > N + l$. This result was conjectured by Sharma by noting that it leads to physically correct large r behavior of the α's for Slater orbitals.

3. Vanishing of $F_{k'k}(NlLM)$ for $k' > l + N$

The coefficients $F_{k'k}(NlLM)$ are expressed in terms of the coefficients $b_v(slLM)$ by[8]

$$F_{k'k}(NlLM) = \sum_{sv} \frac{b_v(slLM)(N-L+2s)!}{(k'-2v)!(N-L-k-k'+2s+2v)!}, \tag{13}$$

which takes the form [on using Eq. (11) and writing $s = u - v$]

$$F_{k'k}(NlLM) = \frac{(1)^L}{2\pi}\,[(2L+1)(2l+1)(L+M)!(L-M)!(l+M)!(l-M)!]^{1/2} \sum_{tu} \frac{(-M-\frac{1}{2}+u-t)!}{t!(L-M-t)!(l-M-t)!(-M-\frac{1}{2}-t)!(2M+t)!(L+l-u)!(N-L-k-k'+2u)!} E_{k'}, \tag{14}$$

where

$$E_{k'} = \sum_v \frac{(L-\frac{1}{2}-u+v)!(l-\frac{1}{2}-v)!(N-L+2u-2v)!}{v!(u-v)!(k'-2v)!}. \tag{15}$$

We show below that the quantity $E_{k'}$ vanishes whenever $k' > l + N$. This will establish that $F_{k'k}(NlLM) = 0$ for $k' > l + N$.

Now from $N - L + 2u - 2v \geqslant 0$, $\frac{1}{2}(N-L) + u - v \geqslant 0$. Suppose initially that $\frac{1}{2}(N-L-k') + u$ is an integer (necessarily nonnegative since from Eq. (14) $N - L - k' + 2u \geqslant k \geqslant 0$). Then

$$\frac{[\frac{1}{2}(N-L)+u-v]!}{(u-v)![(k'/2)-v]!} = \sum_{pp'} \frac{(\frac{1}{2}(N-L))!(\frac{1}{2}(N-L-k')+u)!}{(\frac{1}{2}(N-L-p)!(\frac{1}{2}(N-L-k')+u-p)!(\frac{1}{2}(-N+L+k')-v+p)!}, \tag{16}$$

and

$$\frac{(l-\frac{1}{2}-v)!(\frac{1}{2}(N-L-1)+u-v)!}{(\frac{1}{2}(k'-1)-v)!} = \sum_\sigma (-1)^\sigma \frac{(l-k'/2)!(\frac{1}{2}(N-L-k')+u)!(\frac{1}{2}(N-L-k'-1)+l+u-v-\sigma)!}{\sigma!(l-(k'/2)-\sigma)!(\frac{1}{2}(N-L-k')+u-\sigma!)}, \tag{17}$$

which are valid since on the right, we have terminating series. In Eq. (17), we will have to replace $(l-k'/2)!/(l-k'/2-\sigma)!$ by $(-1)^\sigma(-l+(k'/2)-1+\sigma)!/(-l+(k'/2)-1)!$ in case $l-(k'/2)$ is a negative integer. Note also that for a Slater-type orbital $N - L \geqslant 0$.

Substituting from Eqs. (16) and (17) in the expression for $E_{k'}$ in Eq. (15), we note that the v-summation can be performed which results in

$$E_{k'} = 2^{N-L-k'+2u}\sum_{\sigma p} \frac{[\frac{1}{2}(N-L)!(l-(k'/2)]![(\frac{1}{2}(N-L-k')+u]^2}{p!\sigma!(l-(k'/2)-\sigma)!(\frac{1}{2}(N-L)-p)!} \times \frac{(L-\frac{1}{2}-u)!(\frac{1}{2}(N+L-k')+l-\sigma)!(N-L+l-\frac{1}{2}-k'+u-\sigma-p)!}{(\frac{1}{2}(N-L-k')+u-\sigma)!(\frac{1}{2}(N-L-k')+u-p)!(\frac{1}{2}(-N-L+k')+p)!(N+l-k'-\sigma-p)!}. \tag{18}$$

Now $\frac{1}{2}(N+L-k')+l-\sigma = (L-l-u)+(\frac{1}{2}(N-L-k')+u-\sigma) \geqslant 0$, since each of $L+l-u$, $\frac{1}{2}(N-L-k')+u-\sigma$ is nonnegative whereas for $k' > N+l$, $N+l-k'-\sigma-p < 0$. Thus $E_{k'} = 0$ for $k' > N + l$.

However, if $\frac{1}{2}(N-L-k')+u$ is not an integer, $\frac{1}{2}(N-L-k'-1)+u$ will be a (nonegative integer) and we can modify the above argument slightly to arrive at the same conclusion.

[1]R.R. Sharma, Phys. Rev. A **13**, 517 (1976).
[2]Review No. 15183 by M.A. Rashid published in the Math. Rev. **53** (June 1977).
[3]This is obtained by combining Eqs. (16a) and (17) in Ref. 1.
[4]See Eq. (A.1.2) in A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P. Princeton, New Jersey, 1960).
[5]This is an application of Eq. (9) in reverse.
[6]This is an application of Eq. (5) in reverse.
[7]See Eq. (A.1.1) in A.R. Edmonds *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, New Jersey, 1960).
[8]See Eq. (23c) in Ref. 1.

On some properties of solutions of Helmholtz equation [a)]

A. G. Ramm
University of Michigan, Department of Mathematics, Ann Arbor, Michigan 48109

(Received 12 May 1980; accepted for publication 12 September 1980)

We give a new method to prove results of the following type. Let: $(\nabla^2 + k^2)u = 0$ in $D_R = \{x|x| \geqslant R\}$, $k^2 > 0$. (1) If $u \in L^2(D_R)$, then $u \equiv 0$ in D_R. (2) If $|x|^m u(x) \to 0$ as $|x| \to \infty$, $x_1^2 + \cdots + x_{N-1}^2 \leqslant c x_N^{-2p}, p > 0$, $m = 1, 2, 3, \ldots$, $|x|(\partial u/\partial |x| / - iku) \underset{|x| \to \infty}{\to} 0$ then $u \equiv 0$ in D_R.

PACS numbers: 02.30.Jr

1. INTRODUCTION

Some of the above results were proved by a different method in Refs. 1 and 2, but the new method of the proof is of interest in our opinion. We start with the following theorem.

Theorem 1: Let

$$(\nabla^2 + k^2)u = 0 \quad \text{in } D_R, \quad k^2 > 0. \tag{1}$$

and

$$u \in L^2(D_R). \tag{2}$$

Then $u \equiv 0$ in D_R.

Proof. From (1) and (2) it follows that (see Appendix)

$$\nabla u \in L^2(D_R) \tag{3}$$

and

$$u(x) = \int_{S_R} \left\{ g^+(x,t,k) \frac{\partial u}{\partial N} - u \frac{\partial g^+(x,t,k)}{\partial N} \right\} dt = \int_{S_R} \left(g^- \frac{\partial u}{\partial N} - u \frac{\partial g^-}{\partial N} \right) dt, \tag{4}$$

where N is the unit normal to the sphere $S_R = \{x : |x| = R\}$ directed outside of D_R,

$$g^{\pm}(x,y,k) = \frac{\exp(\pm ik|x-y|)}{4\pi|x-y|},$$
$$|x-y| = (r^2 - 2r|y|\cos\widehat{xy} + |y|^2)^{1/2}, \quad r = |x|. \tag{5}$$

Now the main idea can be explained. We analytically continue functions (4) on the complex plane $z = r\exp(i\psi)$ (see also Refs. 3–5). From (4) it follows that ($\omega = x/|x|^{-1}$)

$$u(x) = u(r,\omega) = u(z,\omega) = \frac{\exp(ikz)}{z} f_1(z,\omega) = \frac{\exp(-ikz)}{z} f_2(z,\omega), \tag{6}$$

where $f_1(z,\omega)$ and $f_2(z,\omega)$ are analytic in z for $|z| > R$ and bounded near infinity. Thus

$$f_j(z,\omega) = \sum_{s=0}^{\infty} f_{js}(\omega) z^{-s}, \quad j = 1,2. \tag{7}$$

But in this case (6) implies that $f_1 = f_2 \equiv 0$. Indeed, if $z = iy$, $y \to +\infty$ then $([\exp(ikz)]/z) f_1(z,\omega)$ in (6) goes to zero exponentially, while $([\exp(-ikz)]/z) f_2(z,\omega)$ goes to infinity exponentially unless $f_2 \equiv 0$. Thus $u(z,\omega) \equiv 0$, $u(r,\omega) \equiv 0$ in D_R.

Let us show how the idea works in a different problem.

2. RATE OF DECREASE OF SOLUTIONS TO HELMHOLTZ EQUATION

Theorem 2: Let (1) hold, u satisfies the radiation condition and

$$|x|^m u(x) \to 0 \quad \text{as} \quad x \to \infty, \quad \rho \leqslant c|x_3|^{-p}$$
$$c = \text{const} > 0, \quad 0 \leqslant p, \quad m = 1,2,3,\ldots, \tag{8}$$

where $\rho = (x_1^2 + x_2^2)^{1/2}$, then $u \equiv 0$ in D_R.

Proof: Since u satisfies (1) and the radiation condition, we can use the first equality in (4) and the third equality in (6). If $p = 0$ the condition (8) says that $u(x)$ decreases faster than any negative power of $|x|$ at infinity in the cylinder $\rho \leqslant c$. From this and (6) it follows that $f_1 \equiv 0$ for ω directed along the axis x_3. By shifting the origin a little, we conclude that $f_1 = 0$ along any ray in the cylinder $\rho \leqslant c$. Thus $u \equiv 0$ in this cylinder and by unique continuation theorem for solutions of homogeneous elliptic equations $u \equiv 0$ in D_R. If $p > 0$ our argument is a little more complicated. In this case let us write the equation $\rho = c|x_3|^{-p}$ in the spherical coordinates: $r^{p+1}\cos^p\theta \cdot \sin\theta = c$. For large r the angle θ is near 0, and $\theta = \theta(r)$ is an analytic bounded function of cr^{-1-p} for large r. Let us prove that $u = 0$ in the body $r^{p+1}\cos^p\theta \sin\theta \leqslant c$. From this and the unique continuation theorem we conclude that $u \equiv 0$ in D_R. Let us take in (4) $x = (r, \theta_b(r))$, where $\theta_b(r)$ is constructed as $\theta(r)$ but instead of c we use $0 < b < c$. For simplicity we shall write $\theta(r)$ instead of $\theta_b(r)$ in what follows. Then $\omega = \omega(r)$ and $\omega(r)$ is an analytic and bounded function of the argument br^{-1-p} for large r. From this it follows that $f_1(r,\omega(r))$ will be analytic and bounded near infinity on an appropriate Riemann surface (which by the way will be finite-sheeted for rational p. Since we can always find a rational number $p_1 > p$ such that the body $r^{1+p}\cos^p\theta \sin\theta = b$ contains the body $r^{1+p_1}\cos^{p_1}\theta \sin\theta = b$, we can consider only finite-sheeted Riemannian surfaces). If $f_1(z)$ decreases faster than any negative power of z on such a surface, $f_1 \equiv 0$. Thus $u(x) = 0$ on any curve $r^{p+1}\cos^p\theta \sin\theta = b \leqslant c$ and we conclude that $u \equiv 0$ in D_R.

3. GENERALIZATIONS

(1) We can consider general elliptic equations with constant coefficients in $\mathbb{R}^N$.

(2) It is possible to consider the case when $u(x)$ is a solution of (1) in a domain with infinite boundary.

[a)]Supported by AFOSR 800204.

 0022-2488/81/020275-02$1.00

1. APPENDIX

Lemma 1: From (1) and (2) inclusion (3) follows. Let

$$g(r) = \int_R^r dr\, r^2 \int_{S_1} |u(r,\omega)|^2\, d\omega,$$

$g(r)$ increases monotonically and $g(\infty) < \infty$,

$$g'(r) = r^2 \int_{S_1} |u(r,\omega)|^2\, d\omega \geqslant 0, g'' = 2r \int_{S_1} |u|^2\, d\omega + 2r^2 \int_S \frac{\partial u(r,)}{\partial r} u(r,\omega)\, d\omega.$$

Here we assume without any loss of generality that u is a real valued function [since the coefficients of Eq. (1) are real]. If $g''(r_n) \to 0$ as $r_n \to \infty$, then

$$\int_{S_{r_n}} \frac{\partial u}{\partial r} u\, dS_{r_n} \to 0 \quad \text{as} \quad z_n \to \infty. \tag{A1}$$

But from (1) it follows that

$$\int_{R \leq |x| \leq r_n} |\nabla u|^2\, dx = k^2 \int_{R \leq |x| \leq r_n} |u|^2\, dx + \int_{S_{r_n}} u \frac{du}{dr}\, dS_{r_n} + \int_{S_R} u \frac{\partial u}{\partial N}\, dS_R. \tag{A2}$$

From (A1), (A2), and (2) we get (3). If $g''(r)$ does not go to zero whatever sequence $r_n \to \infty$ we choose, then $|g''| \geqslant \epsilon > 0$ for all $r \geqslant R_1 \geqslant R$. If $g'' \geqslant \epsilon$, then $g'(r) \to +\infty$. This is impossible because of (2). If $g'' \leqslant -\epsilon$, then $g'(r) \to -\infty$. Again this is impossible because of (2). This completes the proof.

[1] O. Arena and W. Littman, "Farfield behavior of solution to P.D.E.," Ann. Sc. Norm. Super. Pisa 2B, 807–27 (1972).

[2] T. Kato, "Growth properties of solutions of the reduced wave equation with variable coefficient," Commmun. Pure Appl. Math. **12**, 402–25 (1959).

[3] A. G. Ramm, "About the absence of the discrete positive spectrum of the Laplace operator of the Dirichlet problem in some domains with infinite boundaries," Vestn. Leningr. Univ. Mat. Mekh. Astron. **13**, 153–6 (1964); **1**, 176 (1966).

[4] A. G. Ramm, "Nonselfadjoint operators in diffraction and scattering, "Math. Meth. Appl. Sci. **2**, 327–46 (1980).

[5] A. G. Ramm, *Theory and Applications of Some New Classes of Integral Equations* (Springer, New York, 1980).

Gauge equivalence of exactly integrable field theoretic models

J. Honerkamp
Fakultät für Physik der Universität Freiburg, D-7800 Freiburg, West Germany

(Received 20 November 1979; accepted for publication 18 January 1980)

Exactly integrable field theoretic models are constructed which are gauge equivalent to the n-component or $m\cdot n$ component nonlinear Schrödinger equations and to the O(n) nonlinear σ-model. We obtain the CP^n-Heisenberg model or the Grassmann–Heisenberg model and the generalized sine-Gordon model respectively. Consequences for the conserved quantities are discussed.

PACS numbers: 02.30.Jr, 03.65.Fd, 11.10.Np, 11.10.Lm

1. INTRODUCTION

Recently Zakharov and Takhtadzhyan[1] and, independently, D. Chudnovsky and G. Chudnovsky,[2] pointed out, that the nonlinear Schrödinger equation (NLSE) and the equation of a Heisenberg ferromagnet (HF) are gauge equivalent. Gauge equivalence means that the two pairs of linear differential equations

$$\phi_{i,x} = U_i(x,t;\lambda)\phi_i, \quad \begin{cases} i=1, & \text{for NLSE} \\ i=2, & \text{for HF} \end{cases} \tag{1.1a}$$

$$\phi_{i,t} = V_i(x,t;\lambda)\phi_i \tag{1.1b}$$

where U_i, V_i in these cases are polynomials in λ of order one resp. two, can be obtained from each other by a gauge transformation independent of λ:

$$\phi_1 = g\phi_2; \quad U_1 = gU_2g^{-1} + g_xg^{-1}; \quad V_1 = gV_2g^{-1} + g_tg^{-1}. \tag{1.2}$$

Because the consistency conditions of (1),

$$U_{i,t} - V_{i,x} + [U_i, V_i] = 0, \tag{1.3}$$

lead to the nonlinear differential equations defining the models, the gauge transformation constitutes a field coordinate transformation by which the nonlinear field equations and also the conserved quantities transform into each other.

In this note we show in section 2 that a gauge equivalent model to the n-component nonlinear Schrödinger equation presented by Nogami and Warke[3] and by D. Chudnovsky and G. Chudnovsky[2] is a generalized Heisenberg model where instead of a set of spin components $\{S_i\}$ a matrix

$$S \in CP^n = \mathrm{U}(n+1)/\mathrm{U}(n)\otimes\mathrm{U}(1)$$

represents the field coordinates.

A slight modification allows us to discuss a model which is gauge equivalent to a Heisenberg model for a matrix S on a Grassmannian manifold $\mathrm{U}(n+m)/\mathrm{U}(n)\otimes\mathrm{U}(m)$ or even for $S \in \mathrm{GL}(n+m)/\mathrm{GL}(n)\otimes\mathrm{GL}(m)$.

In Sec. 3 we discuss a model which is gauge equivalent to the nonlinear O(n) invariant σ-model.[4] The gauge equivalent models are, for $n=3$, the sine-Gordon model and, for $n>3$, their higher generalizations, which have also been obtained by another method.[5] In Sec. 4 we discuss some consequences for the conserved quantities of those models which are related by a coordinate transformation. As an example we do this in the O(3) σ-model (sine-Gordon model).

2. THE HEISENBERG MODELS

We start with the n-component nonlinear Schrödinger equation[3] for the $2n$ fields. $(\varphi_1,\ldots,\varphi_n)$, $(\chi_1,\ldots,\chi_n)$:

$$-i\varphi_{j,t} + \varphi_{j,xx} + 2\varphi_j \sum_{k=1}^{n} \varphi_k\chi_k = 0 \tag{2.1a}$$

and

$$i\chi_{j,t} + \chi_{j,xx} + 2\chi_j \sum_{k=1}^{n} \varphi_k\chi_k = 0. \tag{2.1b}$$

These equations are the compatibility equations for the linear system

$$\phi_x = U_1\phi, \quad \phi_t = V_1\phi, \tag{2.2}$$

where U_1, V_1 are $(n+1)\otimes(n+1)$ matrices and polynomials in λ:

$$U_1 = i\lambda\Gamma + A_0, \quad V_1 = B_0 + 2\lambda A_0 + 2i\lambda^2\Gamma, \tag{2.3}$$

with

$$\Gamma = \mathrm{diag}(1,-1,-1,\ldots,-1), \tag{2.4}$$

$$A_0 = \begin{pmatrix} 0 & \varphi_1\varphi_2,\ldots,\varphi_n \\ -\chi_1 & \\ -\chi_2 & 0 \\ \vdots & \\ -\chi_n & \end{pmatrix} \tag{2.5}$$

$$B_0 = -i\Gamma A_{0,x} + i\Gamma A_0^2. \tag{2.6}$$

We define $g(x,t)$ as a solution of (2.2) for $\lambda = 0$. Then

$$g_x = A_0g, \tag{2.7}$$

$$g_t = B_0g, \tag{2.8}$$

and

$$U_2 = g^{-1}U_1g - g^{-1}g_x = g^{-1}i\lambda\Gamma g + g^{-1}A_0g - g^{-1}g_x = i\lambda S,$$

where

$$S = g^{-1}\Gamma g. \tag{2.9}$$

Furthermore

$$V_2 = g^{-1}V_1g - g^{-1}g_t = 2\lambda g^{-1}A_0g + 2i\lambda^2 S. \tag{2.10}$$

By use of $\Gamma A_0\Gamma = -A_0$ one can show that

$$SS_x = 2g^{-1}g_x = 2g^{-1}A_0g. \tag{2.11}$$

Hence

$$V_2 = \lambda SS_x + 2i\lambda^2 S. \tag{2.12}$$

The linear system now reads:

$$\phi_{2,x} = i\lambda S\phi_2, \tag{2.13a}$$

$$\phi_{2,t} = (\lambda SS_x + 2i\lambda^2 S)\phi_2. \tag{2.13b}$$

 0022-2488/81/020277-05$1.00

Because of $S = g^{-1}\Gamma g$, we have as condition on S

$$S^2 = 1, \tag{2.14a}$$

$$\mathrm{tr}S = -(n-1), \tag{2.14b}$$

and the compatibility equation for (2.13) is

$$iS_t = \tfrac{1}{2}[S, S_{xx}]. \tag{2.15}$$

For $\varphi_i = \chi_i^+$ we have $g \in SU(n+1)$ and because of $S = g^{-1}\Gamma g$ with Γ by (2.4):

$$S \in CP^n = U(n+1)/U(n) \otimes U(1).$$

Equations (2.1) read

$$i\chi_{j,t} + \chi_{j,xx} + 2\chi_j \sum_{k=1}^{n} |\chi_k|^2 = 0. \tag{2.16}$$

In the case where χ_i, φ_i are independent complex coordinates we have $S \in GL(n+1)/GL(n) \otimes U(1)$. This approach can be immediately generalized to the Heisenberg equations of motions (2.15), (2.14a), and $\mathrm{tr}S = -(n-m)$ for $S \in U(n+m)/U(n) \otimes U(m)$ or even for $S \in GL(n+m)/GL(n) \otimes GL(m)$ if we start with the linear system (2.2) and (2.3), where now, (say $m < n$)

$$\Gamma = \mathrm{diag}\Big(\overbrace{1\,1\,1\,1}^{m}\ \overbrace{-1,-1,-1,-1}^{n}\Big), \tag{2.17}$$

$$A_0 = \begin{pmatrix} & & \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1n} \\ & 0 & \vdots & & & \vdots \\ & & \varphi_{m1} & \varphi_{m2} & \cdots & \varphi_{mn} \\ -\chi_{11} & \cdots & -\chi_{m1} & & & \\ & & & & 0 & \\ -\chi_{1n} & \cdots & -\chi_{mn} & & & \end{pmatrix}, \tag{2.18}$$

$$B_0 = -i\Gamma A_{0,x} + i\Gamma A_0^2.$$

Note, that again $\Gamma A_0 \Gamma = -A_0$, hence with $S = g^{-1}\Gamma g$, again

$$2g^{-1}A_0 g = SS_x. \tag{2.19}$$

The field equations for the $2m \cdot n$ complex fields are

$$A_{0,t} + i\Gamma A_{0,xx} - 2iA_0^3 = 0. \tag{2.20}$$

We call these equations the Grassmann Schrödinger equations or the $m \cdot n$-component Schrödinger equations.

Examples:

(i) $m = 1$, $n = 1$; (2.16) is the usual nonlinear Schrödinger equation $S \in CP^1 = SU(2)$. Equations (2.14) and (2.15) are the equations for the O(3) invariant Heisenberg model.

(ii) $n = 2$, $m = 2$; $\varphi_{ij} = \chi_{ij}^+$. Now

$$\Gamma = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{2.21}$$

and

$$A_0 = \begin{pmatrix} 0 & 0 & \chi_{11}^+ & \chi_{12}^+ \\ 0 & 0 & \chi_{21}^+ & \chi_{22}^+ \\ -\chi_{11} & -\chi_{21} & 0 & 0 \\ -\chi_{12} & -\chi_{22} & 0 & 0 \end{pmatrix}.$$

Hence (2.20) reads ($a = 1,\ldots,m = 2$; $b = 1,\ldots,n = 2$)

$$-i\chi_{a,b,t} + \chi_{a,b,xx} + 2\sum_{k,j=1}^{2} \chi_{kb}(\chi_{kj}^+ \chi_{aj}). \tag{2.22}$$

Of course, from (2.22) one infers immediately the general structure of the mn-component Schrödinger equation in terms of the field components. These theories are especially interesting because a quantization procedure analogous to the usual nonlinear Schrödinger equation[6] is certainly available. The consequences of the gauge equivalence for the quantized version of the models could especially be studied at first in the context of these nonrelativistic theories.

3. THE σ MODELS

We start with the system of linear differential equations for the $O(n+1)$ nonlinear σ-model as given by Pohlmeyer[4]:

$$\phi_{1,\xi} = (1-\zeta)q^a q_\xi^b \mathcal{F}^{ab}\phi_1 \equiv U_1\phi_1, \tag{3.1a}$$

$$\phi_{1,\eta} = (1-\zeta^{-1})q^a q_\eta^b \mathcal{F}^{ab}\phi_1 \equiv V_1\phi_1. \tag{3.1b}$$

Here ξ and η are the light cone variables: $\xi = \frac{1}{2}(t+x)$; $\eta = \frac{1}{2}(t-x)$. The indices a, b run from 0,1,...,n and the $\mathcal{F}^{ab}$ are generators of the $O(n+1)$. In terms of the basis elements of the Clifford algebra[7] [$\mathbb{C}_n = \{\Gamma^1,\ldots,\Gamma^n\}$] we have

$$\mathcal{F}^{ab} = -\tfrac{1}{4}[\Gamma^a, \Gamma^b],$$

$$\mathcal{F}^{0a} = \tfrac{1}{2}\Gamma^a.$$

The q^a ($a = 0,1,\ldots,n$) are components of a $(n+1)$-dimensional unit vector, $q^a q^a = 1$; we choose also $q_\xi^a q_\xi^a = 1$ and $q_\eta^a q_\eta^a = 1$. Hence also e.g. $q^a q_\eta^a = 0$, etc. The equations (3.1) are compatible if

$$q_{\eta\xi}^a + (q_\eta \cdot q_\xi)q^a = 0. \tag{3.2}$$

The gauge transformation

$$U_2 = g^{-1}U_1 g - g^{-1}g_x,$$

$$V_2 = g^{-1}V_1 g - g^{-1}g_t,$$

leads to the expressions

$$C_1 = g^{-1}q^a q_\xi^b \mathcal{F}^{ab} g, \tag{3.3a}$$

$$C_2 = g^{-1}q^a q_\eta^b \mathcal{F}^{ab} g, \tag{3.3b}$$

with

$$g \in O(n+1).$$

Proposition: One can find a $g \in O(n+1)$, so that

$$C_1 = -(i/2)\Gamma^1, \tag{3.4a}$$

$$C_2 = -(i/2)e_i\Gamma^i, \tag{3.4b}$$

where e_i ($i = 1,\ldots,n$) are the components of a n-dimensional unit vector: $\Sigma_{i=1}^n e_i^2 = 1$. g is not yet completely specified by (3.4a) and (3.4b) because any transformation g_0 generated by $\mathcal{F}^{ij}$ ($i,j \neq 0,1$) leads to

$$g_0^{-1}\Gamma^1 g_0 = \Gamma^1,$$

$$g_0^{-1}e_i\Gamma^i g_0 = e_i\Gamma'^i = e_i'\Gamma^i.$$

Proof: (a) We have

$$C_1^2 = g^{-1}q^a q_\xi^b q^c q_\xi^d \mathcal{F}^{ab}\mathcal{F}^{cd} g = -\tfrac{1}{4},$$

because of

$$q^a q^a = 1, \quad q_\xi^a q_\xi^a = 1, \quad q^a q_\xi^a = 0.$$

Hence we may choose a g which transforms $q^a q^b_\xi \mathcal{F}^{ab}$ to a diagonal matrix, which has to be a constant matrix. If we choose Γ^1 to be the diagonal matrix of $\mathbb{C}_n$ we have

$$C_1 = -(i/2)\Gamma^1.$$

(b) In the same manner we show

$$C_2^2 = -\tfrac{1}{4}.$$

Then C_2 may be represented by

$$C_2 = -(i/2)(e'_1\Gamma'^i + \cdots + e'_n\Gamma'^n),$$

with

$$\{\Gamma'^i, \Gamma'^j\}_+ = 2\delta^{ij}, \quad \sum_{i=1}^n e_i'^2 = 1.$$

Because for $q^a_\eta \to q^a_\xi$, $C_2 \to C_1$, the matrix Γ^1 is representable by a linear combination of the $\{\Gamma'^1,\ldots,\Gamma'^n\}$. By a further transformation which leaves Γ^1 invariant, one may arrange C_2 in its standard form

$$C_2 = -(i/2)e_i\Gamma^i.$$

Then we obtain

$$U_2 = (i/2)\zeta\Gamma^1 + A_0, \tag{3.5a}$$

$$V_2 = (i/2)(1/\zeta)e_i\Gamma^i + B_0, \tag{3.5b}$$

where

$$A_0 = -g^{-1}g_\xi - (i/2)\Gamma^1; \quad B_0 = -g^{-1}g_\eta - (i/2)e_i\Gamma^i. \tag{3.5c}$$

Now, the compatibility equation for U_2, V_2,

$$U_{2,\eta} - V_{2,\xi} + [U_2, V_2] = 0, \tag{3.6}$$

leads to

$$e_{i,\xi}\Gamma^i = [A_0, e_i\Gamma^i], \tag{3.7}$$

$$[\Gamma^1, B_0] = 0, \tag{3.8}$$

$$A_{0,\eta} - B_{0,\xi} + e_i\mathcal{F}^{1i} + [A_0, B_0] = 0. \tag{3.9}$$

Equation (3.7) can be solved by the ansatz

$$A_0 = \sum_{e=2}^n a_e\mathcal{F}^{1e} + \sum_{e,k\neq 1} c_{ek}\mathcal{F}^{en}.$$

Because g until now is only specified up to a transformation generated by the $\mathcal{F}^{ek}$ $(e,k \neq 1,0)$ we may choose such a g so that $c_{ek} = 0$. Then we obtain from (3.7)

$$e_{k,\xi}\Gamma^k = a_k e_i[\mathcal{F}^{1k}, \Gamma^i] = -\sum_{k=2}^n (a_k e_k)\Gamma^1 + \sum_{k=2}^n a_k\Gamma^k e_1.$$

Hence

$$e_{1,\xi} = -\sum_{k=2}^n a_k e_k \tag{3.10}$$

$$e_{k,\xi} = a_k e_1, \quad k = 2,\ldots,n. \tag{3.11}$$

From (3.11) we infer

$$a_k = \frac{e_{k,\xi}}{e_1}, \tag{3.12}$$

whereby (3.10) then reads

$$e_{1,\xi}e_1 = -\sum_{k=2}^n e_{k,\xi}e_k$$

which, of course, is true.

From (3.8) we get

$$B_0 = c_1\Gamma^1 + \sum_{i,j\neq 1,0} c_{ij}\mathcal{F}^{ij}, \quad c_{ij} = -c_{ji}, \tag{3.13}$$

so that we obtain finally from (3.9)

$$a_{k,\eta}\mathcal{F}^{1k} - c_{1,\xi}\Gamma^1 - c_{ij,\xi}\mathcal{F}^{ij} + e_k\mathcal{F}^{1k} + c_1 a_k\Gamma^k + 2a_e c_{ke}\mathcal{F}^{1k} = 0. \tag{3.14}$$

Hence $c_{1,\xi} = 0$, $c_{ij,\xi} = 0$, $c_1 = 0$. Then $c_{ij} = c_{ij}(\eta)$ and

$$a_{k,\eta} + e_k + 2a_e c_{ke}(\eta) = 0. \tag{3.15}$$

But the term $c_{ke}\mathcal{F}^{ke}$ in V_2 can again be transformed away so that we end up with the linear system.

$$U_2 = i\frac{\zeta}{2}\Gamma^1 + \sum_{k=2}^n \frac{e_{k,\xi}}{e_1}\mathcal{F}^{1k}, \tag{3.16}$$

$$V_2 = \frac{i}{2\zeta}e_i\Gamma^i, \tag{3.17}$$

and the field equations are

$$(e_{k,\xi}/e_1)_\eta + e_k = 0, \quad k = 2,\ldots,n. \tag{3.18}$$

Equation (3.18) was derived in Ref. 5 by another method, the associated linear problem indicated by the matrices in (3.16), and (3.17) has also been found by Eichenherr and Pohlmeyer.[8]

Note that by (3.5c) we have

$$-g^{-1}g_\xi = (g^{-1})_\xi g = \frac{i}{2}\Gamma^1 + \sum_{k=2}^n \frac{e_{k,\xi}}{e_1}\mathcal{F}^{1k},$$

$$-g^{-1}g_\eta = (g^{-1})_\eta g = \frac{i}{2}e_k\Gamma^k.$$

Hence

$$g^{-1}(\xi,\eta) = \phi_2(\zeta,\eta;\zeta = 1). \tag{3.19}$$

Examples: (i) $n = 2$ [O(3) model]. We choose

$$e_1 = \cos\alpha, \quad e_{1,\xi} = -\alpha_\xi \sin\alpha,$$

$$e_2 = \sin\alpha, \quad e_{2,\xi} = \alpha_\xi \cos\alpha, \tag{3.20}$$

$$e_{2,\xi}/e_1 = \alpha_\xi;$$

hence

$$\alpha_{\xi\eta} + \sin\alpha = 0,$$

with

$$\Gamma^1 = \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \Gamma^2 = \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{3.21}$$

$$\mathcal{F}^{12} = \frac{1}{2}\begin{pmatrix} 0 & -1 \\ +1 & 0 \end{pmatrix};$$

then

$$U_2 = \frac{i}{2}\zeta\sigma^3 - \frac{i}{2}\alpha_\xi\sigma^2 = \frac{i}{2}\zeta\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & -\alpha_\xi \\ +\alpha_\xi & 0 \end{pmatrix}, \tag{3.22}$$

$$V_2 = \frac{1}{\zeta}\frac{i}{2}[\cos\alpha\sigma^3 + \sin\alpha\sigma^1] = \frac{i}{2\zeta}\begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}. \tag{3.23}$$

(ii) $n = 3$ [O(4) model]. We choose

$$e_1 = \cos\alpha, \quad e_{1,\xi} = -\alpha_\xi \sin\alpha,$$
$$e_2 = \sin\alpha\cos\omega, \quad e_{2,\xi} = \alpha_\xi \cos\alpha\cos\omega - \omega_\xi \sin\alpha\sin\omega, \tag{3.24}$$
$$e_3 = \sin\alpha\sin\omega, \quad e_{3,\xi} = \alpha_\xi \cos\alpha\sin\omega + \omega_\xi \sin\alpha\cos\omega,$$

and

$$a_2 = \alpha_\xi \cos\omega - \omega_\xi \tan\alpha \sin\omega, \tag{3.25a}$$
$$a_3 = \alpha_\xi \sin\omega + \omega_\xi \tan\alpha \cos\omega. \tag{3.25b}$$

The field equation can be written as

$$\alpha_{\xi\eta} + \tan\alpha\omega_\xi\omega_\eta + \sin\alpha = 0, \tag{3.26a}$$
$$\omega_{\xi\eta} + \alpha_\xi\omega_\eta + \omega_\xi\alpha_\eta \frac{1}{\cos^2\alpha} = 0. \tag{3.26b}$$

With

$$\Gamma^1 = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \Gamma^2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \Gamma^3 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix},$$
$$\mathscr{F}^{12} = \frac{1}{2}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathscr{F}^{13} = \frac{1}{2}\begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix},$$

we obtain (compare with Ref. 4)

$$U_2 = \frac{1}{2}\begin{pmatrix} 0 & (-\alpha_\xi - i\omega_\xi \tan\alpha)e^{i\omega} \\ (\alpha_\xi - i\omega_\xi \tan\alpha)e^{-i\omega} & 0 \end{pmatrix} + i\frac{\zeta}{2}\begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{3.27a}$$

$$V_2 = \frac{i}{2\zeta}\begin{pmatrix} +\cos\alpha & \sin\omega e^{i\omega} \\ \sin\alpha e^{-i\omega} & -\cos\alpha \end{pmatrix}. \tag{3.27b}$$

4. IMPLICATIONS FOR THE CONSERVED QUANTITIES

In order to study the consequences of these equivalences let us restrict ourselves to the nonlinear O(3)σ-model.

In terms of the angle coordinate α the linear equations read [see (3.22) and (3.23)]

$$\partial_\xi\phi_2 = \left[i\zeta\frac{1}{2}\sigma_3 + \frac{1}{2}\begin{pmatrix} 0 & -\alpha_\xi \\ \alpha_\xi & 0 \end{pmatrix}\right]\phi_2 \equiv U_2\phi_2, \tag{4.1}$$

$$\partial_\eta\phi_2 = \frac{i}{2\zeta}\begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}\phi_2 \equiv V_2\phi_2. \tag{4.2}$$

We now change from the light-cone variables to coordinates x,t:

$$\xi = \tfrac{1}{2}(t+x), \quad \eta = \tfrac{1}{2}(t-x).$$

If we require $\alpha(x,t) \to 0$ for $|x| \to \infty$ we obtain

$$\partial_x\phi_2 = \tfrac{1}{2}(\partial_\xi - \partial_\eta)\phi_2 \underset{|x|\to\infty}{\to} (i/2)(\sigma_3/2)(\zeta - 1/\zeta)\phi_2. \tag{4.3}$$

We define two sets of fundamental Jost solutions $\chi_2^\pm$ to $\partial_x\phi_2 = \tfrac{1}{2}(U_2 - V_2)\phi_2$ by the asymptotic conditions:

$$\chi_2^\pm \underset{x\to\pm\infty}{\to} e^{i(\sigma_3/4)(\zeta - 1/\zeta)x}. \tag{4.4}$$

The transition matrix between both fundamental sets of solutions then is defined by[9]

$$\chi_2^+ = \tau_2\chi_2^-, \quad \tau_2(\zeta) = \begin{pmatrix} A_2(\zeta) & B_2(\zeta) \\ -B_+^2(\zeta) & A_2^+(\zeta) \end{pmatrix}. \tag{4.5}$$

Doing the same for σ-model coordinates we obtain

$$\partial_x\phi_1 = \tfrac{1}{2}[\tfrac{1}{2}i(\zeta - 1)(\mathbf{q}\times\mathbf{q}_\xi) - \tfrac{1}{2}i(1/\zeta - 1)(\mathbf{q}\times\mathbf{q}_\eta)]\cdot\boldsymbol{\sigma}\phi_1. \tag{4.6}$$

Now by (3.3), (3.4) we have, in this case,

$$(\mathbf{q}\times\mathbf{q}_\xi)\cdot\boldsymbol{\sigma} = g\sigma_3 g^{-1}, \tag{4.7}$$
$$(\mathbf{q}\times\mathbf{q}_\eta)\cdot\boldsymbol{\sigma} = g(\cos\alpha\sigma^3 + \sin\alpha\sigma^1)g^{-1}, \tag{4.8}$$

and by (3.19), e.g.,

$$g^{-1}(x) = \chi_2^+(x;\zeta = 1).$$

Hence for $x \to +\infty$ we obtain $g^{-1} \to 1$; for $x \to -\infty$ we obtain $g^{-1} \to \tau_2(\zeta = 1)$, and therefore

$$(\mathbf{q}\times\mathbf{q}_\xi)\cdot\boldsymbol{\sigma}, (\mathbf{q}\times\mathbf{q}_\eta)\cdot\boldsymbol{\sigma} \underset{x\to+\infty}{\to} \sigma_3, \tag{4.9}$$

$$(\mathbf{q}\times\mathbf{q}_\xi)\cdot\boldsymbol{\sigma}, (\mathbf{q}\times\mathbf{q}_\eta)\cdot\boldsymbol{\sigma} \underset{x\to-\infty}{\to} \tau_2^{-1}(1)\sigma_3\tau_2(1) \equiv :t. \tag{4.10}$$

Therefore

$$\partial_x\phi_1 \underset{x\to+\infty}{\to} \tfrac{1}{4}i(\zeta - 1/\zeta)\sigma_3\phi_1, \tag{4.11}$$

$$\partial_x\phi_1 \underset{x\to-\infty}{\to} \tfrac{1}{4}i(\zeta - 1/\zeta)t\phi_1. \tag{4.12}$$

We may define again a set of fundamental solutions $\chi_1^\pm$ by

$$\chi_1^+(x,\zeta) \underset{x\to+\infty}{\to} e^{i\sigma_3(\zeta - 1/\zeta)x/4}, \tag{4.13}$$

$$\chi_1^-(x,\zeta) \underset{x\to-\infty}{\to} e^{it(\zeta - 1/\zeta)x/4}, \tag{4.14}$$

and also a transition matrix by

$$\chi_1^+ = \tau_1\chi_1^-, \quad \tau_1 = \begin{pmatrix} A_1(\zeta) & B_1(\zeta) \\ -B_1^+(\zeta) & A_1^+(\zeta) \end{pmatrix}.$$

On the other hand we have, especially because of $g\phi_2 = \phi_1$,

$$\chi_2^+ = g^{-1}\chi_1^+, \tag{4.15}$$

because in the limit $x \to +\infty$ both sides agree. For $x \to -\infty$ we obtain

$$\tau_2(\zeta)e^{i(\sigma_3/2)(\zeta - 1/\zeta)x} \simeq \tau_2(\zeta = 1)\tau_1(\zeta)e^{i\sigma_3(\zeta - 1/\zeta)x/4},$$

or

$$\tau_1(\zeta) = \tau_2^{-1}(1)\tau_2(\zeta), \tag{4.16}$$

which agrees with

$$\tau_1(\zeta = 1) = \mathbb{1}$$

If we parametrize

$$\tau_2(1) = \begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix}, \quad \alpha\alpha^* + \beta\beta^* = 1,$$

then

$$A_1(\zeta) = \alpha^* A_2(\zeta) + \beta B_2^+(\zeta) \tag{4.17}$$
$$B_1(\zeta) = \alpha^* B_2(\zeta) - \beta A_2^+(\zeta). \tag{4.18}$$

It is interesting to discuss the time behavior. We know that[9]

$$A_2(\zeta,t) = A_2(\zeta,0),$$
$$B_2(\zeta,t) = B_2(\zeta,0)e^{(i/4)(\zeta + 1/\zeta)t},$$

and therefore $\beta(t) = \beta(0)e^{it/2}$, $\alpha(t) = \alpha(0)$. Then $A_1(\zeta,t)$ and $B_1(\zeta,t)$ are not time-independent in general, but the combination

$$\alpha(0)A_1(\zeta) - \beta e^{i/2t} B_1^+(\zeta) = A_2(\zeta). \tag{4.19}$$

If we expand the right-hand side in a Laurent series about $\zeta = 0$ we obtain

$$A_2(\zeta) = \sum_{n=-\infty}^{+\infty} I_n^{(2)} \zeta^n,$$

and the $I_\mu^{(2)}$ are the local conserved quantities of the sine–Gordon model. Doing the same expansion for the left-hand side is not of great help *unless* $\beta = 0$.

If $\beta = B_2(\zeta = 1) = 0$ (i.e. the reflection coefficient for zero momentum is zero) we obtain also in the nonlinear σ-model conserved quantities as coefficients in Laurent expansion of the Jost function $A_1(\zeta)$ about $\zeta = 0$, and these are related to the local conserved quantities of the sine–Gordon equation by a normalization factor $\alpha = A_2(0)$ and by a coordinate transformation. These conserved quantities of the σ-model should be the local ones. The nonlocal charges[10] are obtained by expanding $A_1(\zeta)$ about $w = 0$, $w = (1 - \zeta)/(1 + \zeta)$ and equation (4.19) would mean, for $\beta = 0$, that these are related in the same way to nonlocal charges of the sine–Gordon equation.

But one should stress that our asymptotic conditions (4.9) and (4.10) are very different from the condition in Refs. 4 or 10, and that all these conclusions are a consequence of $\beta = 0$. This point deserves further study.

Gauge transformations on the linear system provide a powerful tool for the construction of field-coordinate transformations. They show that some models are classically equivalent and differ only by the choice of coordinates. The quantized version of these models may become really different because of different quantization rules.

The theories which are gauge equivalent to the $\mathbb{C}P^n\sigma$-models are given elsewhere.[11]

ACKNOWLEDGMENTS

The author thanks H. Eichenherr and K. Pohlmeyer for valuable discussions.

[1]V.E. Zahkharov and L.A. Takhtadzhyan, Teor. Mat. Fiz. **38**, 26 (1979) [Theor. Math. Phys. **17** (1979)].
[2]D.V. Chudnovsky and G.V. Chudnovsky, Preprint Saclay DPh-P/79.
[3]J. Nogami and C.S. Warke, Phys. Lett. **59**, A 251 (1976); I.M. Gelfand and L.A. Diku, Funct. Anal. Appl. **11**, No. 2. (1976); D.F. Chudnovsky and G.V. Chudnovsky, Lett. Nuovo Cimento **22**, 47 (1978).
[4]K. Pohlmeyer, Commun. Math. Phys. **46**, 207 (1976).
[5]K. Pohlmeyer and K.-H. Rehren, J. Math. Phys. **20**, 2628 (1979).
[6]See, e.g., J. Honerkamp, P. Weber, and A. Wiesler, Nucl. Phys. B **152**, 266 (1979); L.D. Faddeev and E.K. Sklyanin, Dokl. Akad. Nauk SSSR **243**, 1430 (1978); E.K. Sklyanin, Dokl. Akad. Nauk SSSR **244**, 1337 (1979); H.B. Thacker and D.W. Wilkinson Phys. Rev. D **19**, 3660 (1979).
[7]H. Boerner, *Darstellungen von Gruppen* (Springer-Verlag, Berlin, Göttingen, Heidelberg, 1955).
[8]H. Eichenherr and K. Pohlmeyer, Phys. Lett. **89B**, 76 (1979).
[9]See, e.g., D.W. McLaughlin, J. Math. Phys. **16**, 96 (1975); M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur; Stud. Appl. Math. **53**, 249 (1974).
[10]M. Lüscher and K. Pohlmeyer, Nucl. Phys. B **137**, 46 (1978).
[11]H. Eichenherr and J. Honerkamp, J. Math. Phys. **22**, 374 (1981).

Integrals over the Fermi function

R. Rockmore
Department of Physics, Serin Physics Laboratory, Piscataway, New Jersey 08854

(Received 27 August 1980; accepted for publication 17 October 1980)

A method for evaluating integrals over the Fermi distribution function using results from Mellin–transform theory is presented. The connection of this approach with the operational result of Blankenbecler is explicated. The method is used to calculate the profile function for a Fermi distribution.

PACS numbers: 02.30.Qy

INTRODUCTION

Some time ago, in a brief note with the same title, Blankenbecler[1] presented a formal device for evaluating integrals over the Fermi distribution function. In several special cases, i.e., the scattering form factors for the Fermi distribution and the "modified Gaussian" distribution,[2] the result was shown to be expressible in closed form. This concise method was later used by Kittel[3] in a textbook calculation of the linear term in the heat capacity of a free electron gas.

Here a powerful alternative approach to such integrals is presented in terms of the transform calculus. The connection of this more prosaic approach with Blankenbecler's[1] operational result is explicated. The present technique is shown to be useful in dealing with the calculation in the impact-parameter representation of the profile function in the case of a Fermi distribution.

Method: Following Blankenbecler,[1] the required integral is of the form

$$\int_0^\infty h(x)[\exp(x-y)+1]^{-1}\,dx$$

$$=\int_0^\infty h(x)\exp(y-x)[\exp(y-x)+1]^{-1}\,dx, \tag{1}$$

where, for the moment, x and y are teken to be the energy and the Fermi energy, respectively, in units of kT; it is assumed[1] that h may be integrated once to yield the function $H(x)$. The point of departure for the present discussion is a result from the theory of the Mellin transfrom,[4]

$$\frac{f}{1+f}=\frac{1}{2\pi i}\int_L \frac{\pi f^z dz}{\sin\pi z}, \tag{2}$$

where L is a path extending from $\zeta - i\infty$ to $\zeta + i\infty$, for $0<\zeta<1$. With the identification,

$$f=\exp(y-x), \tag{3}$$

the integral (1) may be written as

$$\frac{1}{2\pi i}\int_L \frac{\pi dz}{\sin\pi z}\int_0^\infty h(x)e^{z(y-x)}\,dx. \tag{4}$$

After integration by parts, this becomes

$$\frac{1}{2\pi i}\int_L \frac{\pi e^{zy}dz}{\sin\pi z}\left[H(x)e^{-zx}\Big|_0+\int_0^\infty H(x)ze^{-zx}\,dx\right]. \tag{5}$$

The further assumption, consistent with the applications in Ref. 1, that $H(0)=0$ is now made, *although this is not at all essential to the present method*. Thus the integral (5) may now be written as

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{\pi z e^{zy}}{\sin\pi z}dz\int_0^\infty H(x)e^{-zx}\,dx, \tag{6}$$

where the contour L has been shifted to the imaginary axis since $z=0$ is no longer a singular point. Introducing ∂, the derivative with respect to y, expression (6) can be put into the form

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{\pi\partial dz}{\sin\pi\partial}e^{zy}\int_0^\infty H(x)e^{-zx}\,dx, \tag{7}$$

and, after an additional transformation of variable, $z=is$, one finds

$$\frac{\pi\partial}{\sin\pi\partial}\int_0^\infty dx\,H(x)\frac{1}{2\pi}\int_{-\infty}^{\infty}dx\,e^{is(y-x)}$$

$$=\frac{\pi\partial}{\sin\pi\partial}\int_0^\infty dx\,H(x)\delta(x-y)$$

$$=[\pi\partial\csc\pi\partial\,]H(y); \tag{8}$$

the last result is Blankenbecler's operational form.

Variants of this approach based on other results from Mellin-transform theory are possible. Thus, the integration of expression (1) by parts yields

$$\int_0^\infty h(x)[\exp(x-y)+1]^{-1}\,dx$$

$$=-\int_0^\infty h(x)\,d\ln[\exp(y-x)+1]$$

$$=h(0)\ln[\exp(y)+1]+\int_0^\infty \ln[\exp(y-x)+1]\,dh(x); \tag{9}$$

this suggests using the result

$$\ln(1+f)=\frac{1}{2\pi i}\int_{\zeta-i\infty}^{\zeta+i\infty}\frac{\pi f^z}{z\sin\pi z}dz,\quad 0<\zeta<1 \tag{10}$$

in the evaluation

$$\int_0^\infty dx\,h'(x)\ln[\exp(y-x)+1]$$

$$=\frac{1}{2\pi i}\int_{\zeta-i\infty}^{\zeta+i\infty}\frac{\pi e^{yz}dz}{z\sin\pi z}\int_0^\infty dx\,h'(x)e^{-xz}. \tag{11}$$

Thus, for example, the recent discussion of the form factor for a Fermi distribution,

$$F(z)=(4\pi/q)\rho_o\,\mathrm{Im}I, \tag{12}$$

where

$$I=-i\frac{d}{dq}\int_0^\infty e^{iqr}\frac{dr}{1+e^{(r-c)/\beta}}, \tag{13}$$

 0022-2488/81/020282-02$1.00

[ρ_o is the normalization ($F(0) = A$), c is the nuclear shape radius parameter, and β the diffuseness parameter] given by Amado, Dedonder, and Lenz[5] is now simplified considerably through the introduction of the Mellin transform (10). One has

$$I = \beta \frac{d}{dq}\left[q\int_0^\infty dr\, e^{iqr}\ln(e^{-(r-c)/\beta}+1)\right]$$
$$= \beta \frac{d}{dq}\left(\frac{q}{2\pi i}\int_{\zeta-i\infty}^{\zeta+i\infty}\frac{dz\pi}{z\sin\pi z}\int_0^\infty dr e^{iqr-z(r-c)/\beta}\right)$$
$$= \frac{\beta^2}{2i}\frac{d}{dq}\left(q\int_{\zeta-i\infty}^{\zeta+i\infty}dz\frac{e^{zc/\beta}}{z\sin\pi z(z-iq\beta)}\right),$$
$$(0<\zeta<1). \tag{14}$$

Closing the contour in the left half-plane, one obtains I as the sum of contributions from the poles at $z = iq\beta$ and $z = 0$, -1, ···. Since $F(q)$ receives *no* contribution from the pole at $z = 0$, it is easy to see that $F(q)$ can be represented by the contribution from the pole $z = iq\beta$ *alone*,

$$F(q) = (8\pi^2\rho_o/q)\beta e^{-\pi\beta q}(\pi\beta\sinh qc - c\cosh qc), \tag{15}$$

as long as q is in the region $1/\pi\beta \ll q \ll c/\pi\beta^2$.

CALCULATION OF THE PROFILE FUNCTION FOR A FERMI DISTRIBUTION

It was recently noted[5] that a description of the nuclear shape in terms of the Fermi distribution function is a valuable tool in elucidating various characteristic features of hadron-nucleus elastic scattering at high energy. The profile function $t(b)$ in the eikonal amplitude,[6]

$$f(q) = ik\int_0^\infty J_0(qb)(1-e^{-\gamma t(b)})b\,db, \tag{16}$$

given by

$$t(b) = \int_{-\infty}^{\infty}\rho((z^2+b^2)^{1/2})dz, \tag{17}$$

plays a central role in that discussion. As a final illustration of the approach to integrals over the Fermi distribution function detailed above a useful expression for $t(b)$ is presented in this case.

$$t(b) = 2\rho_0\int_b^\infty dr\frac{r}{(r^2-b^2)^{1/2}}\int_{\sigma-i\infty}^{\sigma+i\infty}\frac{ds}{2\pi i}\pi\frac{e^{-sr/\beta}}{\sin\pi s}e^{sc/\beta}$$
$$= -i\rho_0 b\int_{\sigma-i\infty}^{\sigma+i\infty}ds\frac{e^{sc/\beta}}{\sin\pi s}K_1\left(\frac{sb}{\beta}\right), \quad (0<\sigma<1) \tag{18}$$

since

$$K_1(z) = \int_1^\infty e^{-zt}\frac{tdt}{(t^2-1)^{1/2}}. \tag{19}$$

Evaluating $t(b)$ at the nuclear radius ($b \to c_+$) for example, one finds[5] the asymptotic expression in powers of the small parameter $\tau = \beta/c$,

$$t(c) = -i\rho_0 c\int_{\sigma-i\infty}^{\sigma+i\infty}ds\frac{e^{sc/\beta}}{\sin\pi s}\left(\frac{\pi\beta}{2sc}\right)^{1/2}$$
$$\times[1+3\beta/(8sc) - 15\beta^2/(128s^2c^2)+\cdots]e^{-sc/\beta}$$
$$= \rho_0(2\pi\beta c)^{\frac{1}{2}}[(1-\sqrt{2})\zeta(\tfrac{1}{2}) + \tfrac{3}{8}(1-1/\sqrt{2})\zeta(\tfrac{3}{2})\tau$$
$$-\tfrac{15}{128}(1-1/2\sqrt{2})\zeta(\tfrac{5}{2})\tau^2+\cdots], \tag{20}$$

after closing the contour in the right half-plane. On the other hand, for small impact parameters at, say $b = 0$, one finds, using the result

$$K_1(z)\sim 1/z \quad (z\to 0) \tag{21}$$

and closing the contour in the left half-plane,

$$t(0)\simeq -i\rho_0\beta\int_{\sigma-i\infty}^{\sigma+i\infty}ds\frac{e^{sc/\beta}}{s\sin\pi s}\simeq 2\rho_0(c+\beta e^{-c/\beta}), \tag{22}$$

where only the contributions from the poles at $s = 0, -1$ have been retained; in this region one finds that the profile function becomes proportional to the nuclear radius.[5]

[1] R. Blankenbecler, Am. J. Phys. **25**, 279 (1957).
[2] E. Hahn, D. Ravenhall and R. Hofstadter, Phys. Rev. **101**, 1136 (1956).
[3] C. Kittel, *Elementary Statistical Physics* (Wiley, New York, 1958), Sec. 20.
[4] H. Hochstadt, *The Functions of Mathematical Physics* (Wiley-Interscience, New York, 1971), pg. 80.
[5] R. D. Amado, J. -P. Dedonder and F. Lenz, Phys. Rev. C **21**, 647 (1980).

[6] As in Ref. 5, $\gamma = (\sigma_{\rm Tot}/2)(1-ir)$, where r = ratio of real to imaginary part of the forward amplitude; b is the impact parameter and q the momentum transfer.

The set of all projective limits of a projective system of state operators

Wilhelm Ochs
Fakultät für Physik, Universität München, Theresienstr. 37, München, Germany

(Received 17 August 1979; accepted for publication 28 December 1979)

Let $\{W_K\}$ be a projective system of state operators defined on the finite tensor products of some family $\{\mathscr{H}_t\}$ of Hilbert spaces. We prove that all projective limits of $\{W_K\}$ on the *complete* tensor product $\otimes_{t\in T}\mathscr{H}_t$ can be generated from every single projective limit by certain operations. In addition, we provide a necessary and sufficient condition for $\{W_K\}$ to have projective limits on *incomplete* tensor products of the $\mathscr{H}_t$'s.

PACS numbers: 02.30.Tb, 03.65.Ca, 02.50.Cw

1. INTRODUCTION

In a preceding paper,[1] we established necessary and sufficient conditions for a projective system $\{W_K : K\in\mathscr{F}(T)\}$ of state operators to have a projective limit on the *complete* tensor product $\otimes_{t\in T}\mathscr{H}_t$. In particular, we showed that, in contrast to projective systems of probability measures,[2] a projective system of state operators may have none or many projective limits. Therefore, the purpose of the present paper is to clarify the structure of the set of all projective limits belonging to a given projective system of state operators.

In Sec. 2, we recapitulate the basic notions of Ref. 1 and compile the required facts about infinite tensor products of Hilbert spaces. In Sec. 3 we investigate the structure of the set $\mathfrak{L}$ of all projective limits of a projective system. In its first part, a Choquet-type theorem is established for $\mathfrak{L}$ and it is proven that $\mathfrak{L}$ is the norm closed convex hull of its strongly exposed points. In the second part of Sec. 3 it is shown that $\mathfrak{L}$ can be obtained from any element $W\in\mathfrak{L}$ by the relation

$$\mathfrak{L} = \mathfrak{u}^{-1}(\mathfrak{u}W),$$

where $\mathfrak{u}$ belongs to a class of explicitly given mappings from the set of all state operators on $\otimes_{t\in T}\mathscr{H}_t$ into itself. In Sec. 4 we consider special projective limits which live on *incomplete* tensor products, and we give a necessary and sufficient condition for such projective limits to exist. It turns out that this kind of projective limit is unique if it exists.

2. MATHEMATICAL PRELIMINARIES

Throughout this paper, "Hilbert space" means "complex Hilbert space of dimension $\geqslant 1$." If $\mathscr{K}$ is a Hilbert space, $\mathscr{B}(\mathscr{K})$ denotes the Banach space of all bounded linear operators on $\mathscr{K}$ equipped with the operator norm $\|\cdot\|$, and $\mathscr{T}(\mathscr{K})$ denotes the Banach space of all trace class operators equipped with the trace norm $\|\cdot\|_1$. A *state operator* (STO) on $\mathscr{K}$ is an element of $\mathscr{S}(\mathscr{K}) := \{X\in\mathscr{T}(K): X\geqslant 0, \operatorname{tr}(X)=1\}$. "Projection" will always mean "orthogonal projection." For $0\neq\varphi\in\mathscr{K}$, $P(\varphi)$ denotes the projection onto $\operatorname{span}\{\varphi\}$. The range of a mapping f is denoted by $\operatorname{rng} f$.[3]

In the sequel we consider an arbitrary nonempty collection $\{\mathscr{H}_t : t\in T\}$ of Hilbert spaces which we keep fixed for the rest of this paper. By $\mathscr{F}(T)$ we denote the directed set of all finite nonempty subsets of T directed by inclusion. To every $\emptyset\neq M\subseteq T$ we associate the complete tensor product $\mathscr{H}^M := \otimes_{t\in M}\mathscr{H}_t$;[4] for $\mathscr{H}^T$ we write $\hat{\mathscr{H}}$. The unit operator on $\mathscr{H}^M$ is denoted by $\mathbf{1}_M$. An element $\alpha = \otimes_{t\in T}\alpha_t$ in $\hat{\mathscr{H}}$ with $\|\alpha_t\| = 1$ $(\forall t\in T)$ is called a *product unit vector* (PUV). Two PUV's α,β are called *equivalent* (written $\alpha\sim\beta$), if $\Sigma_{t\in T}|1-\langle\alpha_t,\beta_t\rangle| < \infty$. The set of all $(\sim)$-equivalence classes $\mathbf{a}, \mathbf{b}, \cdots$ is denoted by Γ. The PUV's in $\mathbf{a}\in\Gamma$ span a closed subspace $\hat{\mathscr{H}}_{\mathbf{a}}$ of $\hat{\mathscr{H}}$ which is called the *incomplete tensor product* (ICT) of the $\mathscr{H}_t$'s *with respect to* $\mathbf{a}$. If $\alpha = \otimes_t\alpha_t$ is a PUV in $\mathbf{a}$, then the set of all PUV's $\otimes_t\beta_t$ such that $\alpha_t = \beta_t$ for all but a finite number of $t\in T$ is total in $\hat{\mathscr{H}}_{\mathbf{a}}$; in this sense, $\hat{\mathscr{H}}_{\mathbf{a}}$ is generated by any PUV in $\mathbf{a}$. If we want to refer to a particular generating PUV in $\mathbf{a}$, say α, we also write $\hat{\mathscr{H}}[\alpha]$ for $\hat{\mathscr{H}}_{\mathbf{a}}$. The complete tensor product is the direct sum of all ICT's, $\hat{\mathscr{H}} = \oplus_{\mathbf{a}\in\Gamma}\hat{\mathscr{H}}_{\mathbf{a}}$. $Q_{\mathbf{a}}$ denotes the projection from $\hat{\mathscr{H}}$ onto $\hat{\mathscr{H}}_{\mathbf{a}}$. If $\{\epsilon_t : t\in D\}$, $D\overset{\subseteq}{\neq} T$, is a collection of unit vectors $\epsilon_t\in\mathscr{H}_t$, then we abbreviate $\otimes_{t\in T}\epsilon_t$ by $\epsilon[D]$.

Let $\mathscr{Z}$ be the set of all families $\{z_t : t\in T\}$ of complex numbers of modulus one. If $\alpha = \otimes_t\alpha_t$ is a PUV in $\hat{\mathscr{H}}$ and $\mathfrak{z} = \{z_t\}\in\mathscr{Z}$, then $\mathfrak{z}\alpha := \otimes_{t\in D}(z_t\alpha_t)$ is again a PUV in $\hat{\mathscr{H}}$. $\mathfrak{z}\alpha$ is equivalent to α if and only if $\Pi_{t\in T} z_t$ converges (i.e., $\Sigma_t|\arg z_t|$ converges). Two PUV's α,β are called *weakly equivalent* (written $\div$), if there is a $\mathfrak{z}\in\mathscr{Z}$ with $\alpha\sim\mathfrak{z}\beta$. The equivalence relation "$\div$" is compatible with and weaker than the relation "$\sim$." So it induces an equivalence relation in Γ by

$$\mathbf{a}\div\mathbf{b} :\Leftrightarrow (\exists\ \alpha\in\mathbf{a},\quad \beta\in\mathbf{b})\quad \alpha\sim\beta$$

for which we use the same symbol. The closed subspace spanned by the set of PUV's which are weakly equivalent to a PUV $\beta\in\mathbf{b}$ is called the *weak incomplete tensor product* (WICT) of the $\mathscr{H}_t$'s *with respect to* $\mathbf{b}$ and is denoted by $\hat{\mathscr{H}}_{w\mathbf{b}}$ or $\hat{\mathscr{H}}_w[\beta]$,

$$\hat{\mathscr{H}}_{w\mathbf{b}} = \bigoplus_{\mathbf{a}\div\mathbf{b}}\hat{\mathscr{H}}_{\mathbf{a}} \tag{2.1}$$

To every $\mathfrak{z}\in\mathscr{Z}$ there is a unique unitary operator $U(\mathfrak{z})$ on $\hat{\mathscr{H}}$ such that $U(\mathfrak{z})\alpha = \mathfrak{z}\alpha$ and

$$U(\mathfrak{z})(\hat{\mathscr{H}}[\alpha]) = \hat{\mathscr{H}}[\mathfrak{z}\alpha]$$

for all PUV's $\alpha\in\hat{\mathscr{H}}$. Hence every $U(\mathfrak{z})$ leaves all WICT's invariant. To every $\mathbf{b}\div\mathbf{a}$ there is a $\mathfrak{z}\in\mathscr{Z}$ such that $U(\mathfrak{z})(\hat{\mathscr{H}}_{\mathbf{a}}) = \hat{\mathscr{H}}_{\mathbf{b}}$. For $\varphi\in\hat{\mathscr{H}}$, $E[\varphi]$ denotes the projection onto $\overline{\operatorname{span}}\{U(\mathfrak{z})\varphi : \mathfrak{z}\in\mathscr{Z}\}$. A *cylinder operator* is an element in $\mathscr{B}(\hat{\mathscr{H}})$ of the form $Y\otimes\mathbf{1}_{T\setminus K}$ with $Y\in\mathscr{B}(\mathscr{H}^K)$ and $K\in\mathscr{F}(T)$.

Lemma 2.1[5]: (i) If $\xi\in\hat{\mathscr{H}}_{\mathbf{b}}$ for some $\mathbf{b}\in\Gamma$, then $E[\xi]$ commutes with all $U(\mathfrak{z})$ and $Q_{\mathbf{a}}$, $E[\xi]Q_{\mathbf{b}} = P(\xi)$, and $E[\xi]Q_{\mathbf{c}} =$ for all $\mathbf{c}\in\Gamma$ which are not weakly equivalent to $\mathbf{b}$.

(ii) Every cylinder operator commutes with all $U(\mathfrak{z})$ and $Q_{\mathbf{a}}$.

 0022-2488/81/020284-06$1.00

For $\emptyset \neq A \subseteq B \subseteq T$, $\Theta(A,B)$ denotes the *partial trace* from $\mathcal{T}(\mathcal{H}^B)$ to $\mathcal{T}(\mathcal{H}^A)$; for $\Theta(A,T)$ we write Θ_A.[1] A family $\{W_K : K \in \mathcal{F}(T)\}$ of STO's $W_K \in \mathcal{T}(\mathcal{H}^K)$ is called a *projective system, if* $W_K = \Theta(K,H)\, W_H$ for all $K,H \in \mathcal{F}(T)$ with $K \subseteq H$. A STO $V \in \mathcal{T}(\hat{\mathcal{H}})$ is called a *projective limit* of the projective system $\{W_K\}$, if $W_K = \Theta_K V$ for all $K \in \mathcal{F}(T)$. The set of all projective limits of the projective system $\{W_K : K \in \mathcal{F}(T)\}$ is denoted by $\mathfrak{L}(\{W_K\})$.

3. THE STRUCTURE OF $\mathfrak{L}(\{W_K\})$

If T is finite, then our problem is trivial: Every projective system $\{W_K : K \in \mathcal{F}(T)\}$ of STO's with finite T has exactly one projective limit, namely W_T. We therefore assume in the rest of this paper that T is infinite. Since the projective system $\{W_K : K \in \mathcal{F}(T)\}$ is kept fixed throughout the paper, we will write $\mathfrak{L}$ for $\mathfrak{L}(\{W_K\})$ and $\mathcal{F}$ for $\mathcal{F}(T)$.

Lemma 3.1: As a subset of the Banach space $\mathcal{T}(\hat{\mathcal{H}})$, $\mathfrak{L}$ is bounded, closed and convex.

Proof: Boundedness of $\mathfrak{L}$ is obvious and convexity follows from the linearity of the trace and the partial trace. Let $(X_i)_{i \in \mathbb{N}}$ be a sequence in $\mathfrak{L}$ and $Y \in \mathcal{T}(\hat{\mathcal{H}})$ such that $\|X_i - Y\|_1 \to 0$. By Theorem A.3 of Ref. 1 we have

$$\|W_K - \Theta_K Y\|_1 = \|\Theta_K(X_i - Y)\|_1 \leq \|X_i - Y\|_1,$$

which implies that $W_K = \Theta_K Y$ for all $K \in \mathcal{F}$. Hence $Y \in \mathfrak{L}$ and $\mathfrak{L}$ is closed. ■

The following propositions show that $\mathfrak{L}$ is generated by its extremal points.

Definition: Let X be a (real or complex) Banach space and B a bounded subset of X. $x \in B$ is called a *strongly exposed point of* B if there exists a functional $f \in X^*$ with $\|f\| = 1$ such that $f(x) > f(y)$ for all $y \in B$, $y \neq x$ and such that $\|y_n - x\| \to 0$ for every sequence (y_n) in B with $f(y_n) \to f(x)$.

Proposition 3.2: $\mathfrak{L}$ is the $\|\cdot\|_1$-closed convex hull of its strongly exposed points.

Proof: We show in the Appendix that $\mathcal{T}(\hat{\mathcal{H}})$ has the "Radon–Nikodym property." So the proposition follows from Lemma 3.1 and from a theorem of Phelps.[6,7] ■

Since every strongly exposed point is extremal, Proposition 3.2 implies the Krein–Milman type assertion that $\mathfrak{L}$ is the closed convex hull of its extremal points. With regard to the extremal points of $\mathfrak{L}$, there even holds a Choquet type theorem. For C a convex set, ex$C$ denotes the set of all extremal points of C.

Proposition 3.3: For every $W \in \mathfrak{L}$ there is a complete Borel probability measure μ on $\mathfrak{L}$ such that

$$W = \text{Bochner} - \int_{\mathfrak{L}} x d\mu(x)$$

and $\mu(\text{ex}\mathfrak{L}) = 1$.

Proof: Let $C(\Gamma)$ denote the set of all countable subsets of Γ. For $A \in C(\Gamma)$ we define

$$\hat{\mathcal{H}}_A := \bigoplus_{\mathfrak{a} \in A} \hat{\mathcal{H}}_{\mathfrak{a}},$$

$$\mathcal{S}_A := \{W \in \mathcal{S}(\hat{\mathcal{H}}) : \text{rng} W \subseteq \hat{\mathcal{H}}_A\},$$

and

$$\mathfrak{L}_A := \mathcal{L} \cap \mathcal{S}_A.$$

$\hat{\mathcal{H}}_A$ is a separable closed subspace of $\hat{\mathcal{H}}$ and $\mathfrak{L}_A$ is a separable subset of the Banach space $\mathcal{T}(\hat{\mathcal{H}})$. Since $\mathfrak{L}$ and $\mathcal{S}_A$ are both bounded, closed and convex, $\mathfrak{L}_A$ is also bounded, closed and convex. And since $\mathcal{T}(\hat{\mathcal{H}})$ has the "Radon–Nikodym property," it follows from a theorem of Edgar that, to every $V \in \mathfrak{L}_A$, there is a probability measure μ on the universally measurable subsets of $\mathfrak{L}_A$ such that $V = \int x d\mu(x)$ as a Bochner integral and $\mu(\text{ex}\mathfrak{L}_A) = 1$.[7,8] Now $\text{ex}\mathfrak{L}_A = \mathfrak{L}_A \cap \text{ex}\mathfrak{L}$ since $\mathfrak{L}_A$ is a *face* of $\mathfrak{L}$ (i.e., from $W = pV_1 + (1-p)V_2$, $0 < p < 1$ and $W \in \mathfrak{L}_A$, $V_i \in \mathfrak{L}$ it follows that $V_i \in \mathfrak{L}_A$ for $i = 1,2$). Hence the trivial continuation of μ from $\mathfrak{L}_A$ to $\mathfrak{L}$ yields the assertion for every $W \in \mathfrak{L}_A$. Thus the theorem is proven since $\mathfrak{L} = \cup\{\mathfrak{L}_A : A \in C(\Gamma)\}$. ■

Unfortunately, these propositions give little information about the concrete structure of $\mathfrak{L}$ since we failed to specify all extremal or strongly exposed points of $\mathfrak{L}$. Therefore, we will try a new approach to gain insight into the structure of $\mathfrak{L}$. To this end we consider operations which generate new elements of $\mathfrak{L}$ from given ones.

Lemma 3.4: Let $\{R_i : i \in I\}$ be a family of mutually orthogonal projections on $\hat{\mathcal{H}}$ such that $\Sigma_{i \in I} R_i = \mathbf{1}_T$ and that every $Q_{\mathfrak{a}}$ is smaller than or equal to some R_i. Then, for every $V \in \mathcal{S}(\hat{\mathcal{H}})$,[9]

$$V \in \mathfrak{L} \Longleftrightarrow \left(\sum_{i \in I} R_i V R_i\right) \in \mathfrak{L}.$$

Proof: Let $V \in \mathcal{S}(\hat{\mathcal{H}})$. By Lemma 2.1(ii), every R_i commutes with all cylinder operators, and so we get

$$\begin{aligned}
\text{tr}[V(Y \otimes \mathbf{1}_{T \setminus K})] &= \text{tr}\left[\left(\sum_{i \in I} R_i\right) V (Y \otimes \mathbf{1}_{T \setminus K})\right] \\
&= \sum_{i \in I} \text{tr}[R_i V (Y \otimes \mathbf{1}_{T \setminus K}) R_i] \\
&= \sum_{i \in I} \text{tr}[R_i V R_i (Y \otimes \mathbf{1}_{T \setminus K})] \\
&= \text{tr}\left[\left(\sum_{i \in I} R_i V R_i\right)(Y \otimes \mathbf{1}_{T \setminus K})\right]
\end{aligned}$$

for all $K \in \mathcal{F}$ and $Y \in \mathcal{B}(\mathcal{H}^K)$. Hence

$$\Theta_K V = \Theta_K \left(\sum_{i \in I} R_i V R_i\right)$$

for all $K \in \mathcal{F}$ which yields the assertion. ■

The prescription

$$X \mapsto \mathfrak{q}X := \sum_{\mathfrak{a} \in \Gamma} Q_{\mathfrak{a}} X Q_{\mathfrak{a}} \tag{3.1}$$

uniquely defines a mapping $\mathfrak{q}$ from $\mathcal{S}(\hat{\mathcal{H}})$ into itself, and Lemma 3.4 implies that

$$W \in \mathfrak{L} \Longleftrightarrow \mathfrak{q}W \in \mathfrak{L} \tag{3.2}$$

for all $W \in \mathcal{S}(\hat{\mathcal{H}})$.

Lemma 3.5: Let $W = \Sigma_{i \in I} c_i V_i$ be an arbitrary countable convex decomposition of a STO $W \in \mathcal{S}(\hat{\mathcal{H}})$ in $V_i \in \mathcal{S}(\hat{\mathcal{H}})$ and let $\{U_i : i \in I\}$ be a family of unitary operators on $\hat{\mathcal{H}}$ of type $U(\mathfrak{z})$. Then

$$W \in \mathfrak{L} \Longleftrightarrow \left(\sum_{i \in I} c_i U_i V_i U_i^*\right) \in \mathfrak{L}.$$

Proof: As a convex sum of STO's, $W_0 := \Sigma_i c_i U_i V_i U_i^*$ is

also a STO. From Lemma 2.1(ii) we conclude that

$$\begin{aligned}&\mathrm{tr}[W_0(Y\otimes \mathbb{1}_{T\setminus K})]\\&\quad=\sum_{i\in I}c_i\,\mathrm{tr}[U_iV_iU_i^*(Y\otimes \mathbb{1}_{T\setminus K})]\\&\quad=\sum_{i\in I}c_i\,\mathrm{tr}[V_i(Y\otimes \mathbb{1}_{T\setminus K})]=\mathrm{tr}[W(Y\otimes \mathbb{1}_{T\setminus K})]\end{aligned}$$

for all $K\in\mathcal{F}$, $Y\in\mathcal{B}(\mathcal{H}^K)$ which proves the assertion. ■

Let σ be a mapping from Γ into itself such that $\sigma(\mathbf{a})\div\mathbf{a}$ and $\mathbf{a}\div\mathbf{b}\Rightarrow\sigma(\mathbf{a})=\sigma(\mathbf{b})$, and let $\{U_\mathbf{a}:\mathbf{a}\in\Gamma\}$ be a family of unitary operators of type $U(\mathfrak{H})$ such that $U_\mathbf{a}(\hat{\mathcal{H}}_\mathbf{a})=\hat{\mathcal{H}}_{\sigma(\mathbf{a})}$ for all $\mathbf{a}\in\Gamma$. Then

$$\mathfrak{u}_\sigma W:=\sum_{\mathbf{a}\in\Gamma}U_\mathbf{a}Q_\mathbf{a}WQ_\mathbf{a}U_\mathbf{a}^* \tag{3.3}$$

uniquely determines a mapping from $\mathcal{S}(\hat{\mathcal{H}})$ into itself, and we conclude from Lemmas 3.4 and 3.5 that

$$W\exists\mathfrak{L}\Longleftrightarrow\mathfrak{u}_\sigma W\in\mathfrak{L} \tag{3.4}$$

for all $W\in\mathcal{S}(\hat{\mathcal{H}})$ and all mappings $\mathfrak{u}_\sigma$ of type (3.3).

Lemmas 3.4 and 3.5 show how to construct new elements of $\mathfrak{L}$ from known ones. The following theorem even shows that all of $\mathfrak{L}$ can be obtained in this way from every single element.

Theorem 3.6: For all mappings $\mathfrak{u}_\sigma$ of type (3.3) and all $W\in\mathfrak{L}$,

$$\mathfrak{L}=\mathfrak{u}_\sigma^{-1}(\mathfrak{u}_\sigma W)\,. \tag{3.5}$$

The proof of this theorem is based on

Lemma 3.7: Let M_1, M_2 be two STO's in $\mathfrak{L}$, each with a countable convex decomposition

$$M_i=\sum_{\mathbf{a}\in A_i}c_{i\mathbf{a}}X_{i\mathbf{a}} \tag{3.6}$$

in STO's $X_{i\mathbf{a}}\in\mathcal{S}(\hat{\mathcal{H}})$ such that $A_i\subseteq\Gamma$ and $\mathrm{rng}X_{i\mathbf{a}}\subseteq\hat{\mathcal{H}}_a$ for all $\mathbf{a}\in A_i$ and $i=1,2$. If $\mathbf{a}\div\mathbf{b}$ implies that $\mathbf{a}=\mathbf{b}$ for all $\mathbf{a},\mathbf{b}\in A_1\cup A_2$, then $M_1=M_2$.

Proof: Assume that $\mathbf{a}\div\mathbf{b}\Rightarrow\mathbf{a}=\mathbf{b}$ for all $\mathbf{a},\mathbf{b}\in A_1\cup A_2$ and consider an arbitrary element $\mathbf{b}\in A_1\cup A_2$. Let β be any PUV in $\hat{\mathcal{H}}_\mathbf{b}$ and φ_K any unit vector in $\mathcal{H}^K$, $K\in\mathcal{F}$. Then we infer from $M_1,M_2\in\mathfrak{L}$ that

$$\begin{aligned}&\mathrm{tr}[M_1P(\varphi_K\otimes\beta[H\setminus K])\otimes \mathbb{1}_{T\setminus H}]\\&\quad=\mathrm{tr}[M_2P(\varphi_K\otimes\beta[H\setminus K])\otimes \mathbb{1}_{T\setminus H}]\end{aligned} \tag{3.7}$$

for all $H\in\mathcal{F}$ with $K\subset H$. The net

$$\{P(\varphi_K\otimes\beta[H\setminus K])\otimes \mathbb{1}_{T\setminus H}:H\in\mathcal{F}\}$$

of projections converges strongly to $E[\Phi]$ with $\Phi:=\varphi_K\otimes\beta[T\setminus K]$.[10] Hence, since the trace is normal, (3.7) yields

$$\mathrm{tr}(M_1E[\Phi])=\mathrm{tr}(M_2E[\Phi])\,. \tag{3.8}$$

If $\mathbf{b}\in A_k$, $k\in\{0,1\}$, it follows from $\Phi\in\hat{\mathcal{H}}_\mathbf{b}$, from Lemma 2.1(i) and from our assumptions that

$$\begin{aligned}&\mathrm{tr}(M_kE[\Phi])\\&\quad=\sum_{\mathbf{a}\in A_k}c_{k\mathbf{a}}\,\mathrm{tr}(X_{k\mathbf{a}}E[\Phi])\\&\quad=c_{k\mathbf{b}}\,\mathrm{tr}(X_{k\mathbf{b}}E[\Phi])=c_{k\mathbf{b}}\,\mathrm{tr}(Q_\mathbf{b}X_{k\mathbf{b}}E[\Phi])\\&\quad=c_{k\mathbf{b}}\,\mathrm{tr}(X_{k\mathbf{b}}P(\Phi))=c_{k\mathbf{b}}\langle\Phi,X_{k\mathbf{b}}\Phi\rangle\,.\end{aligned} \tag{3.9}$$

If $\mathbf{b}\notin A_k$, it follows from the same reasons that

$$\mathrm{tr}(M_kE[\Phi])=0\,. \tag{3.10}$$

Hence we conclude from (3.8) to (3.10) that $c_{1\mathbf{b}}\langle\Phi,X_{1\mathbf{b}}\Phi\rangle=c_{2\mathbf{b}}\langle\Phi,X_{2\mathbf{b}}\Phi\rangle$ for all $\mathbf{b}\in A_1\cap A_2$ and that $c_{k\mathbf{b}}\langle\Phi,X_{k\mathbf{b}}\Phi\rangle=0$ for $\mathbf{b}\in A_k\setminus(A_1\cap A_2)$. Since the set of all vectors of the form $\varphi_K\otimes\beta[T\setminus K]$ is dense in $\hat{\mathcal{H}}_\mathbf{b}$, it follows that $c_{1\mathbf{b}}X_{1\mathbf{b}}=c_{2\mathbf{b}}X_{2\mathbf{b}}$ for all $\mathbf{b}\in A_1\cap A_2$ and $c_{k\mathbf{b}}X_{k\mathbf{b}}=0$ for all $\mathbf{b}\in A_k\setminus(A_1\cap A_2)$ and $k=1,2$. Hence $M_1=M_2$. ■

Proof of Theorem 3.6: Let X,Y be arbitrary elements of $\mathfrak{L}$ and let $\mathfrak{u}_\sigma$ be an arbitrary mapping of type (3.3), with the associated family $\{U_\mathbf{a}:\mathbf{a}\in\Gamma\}$ of unitary operators. Choose any diagonal representation

$$X=\sum_{\nu\in A}x_\nu P(\varphi_\nu)\,,\quad Y=\sum_{\mu\in B}y_\mu P(\psi_\mu) \tag{3.11}$$

of X,Y [that means: (φ_ν) and (ψ_μ) are orthonormal sequences and $x_\nu,y_\mu>0$ for all ν,μ]. Then the STO's $\mathfrak{u}_\sigma X$, $\mathfrak{u}_\sigma Y$ can be written in the form

$$\begin{aligned}\mathfrak{u}_\sigma X&=\sum_{\nu\in A}\sum_{\mathbf{e}\in E_\nu}x_\nu c_{\nu\mathbf{e}}^2P(U_\mathbf{e}\varphi_{\nu\mathbf{e}})=\sum_{\mathbf{a}\in S}g_\mathbf{a}X_\mathbf{a}\,,\\\mathfrak{u}_\sigma Y&=\sum_{\mu\in B}\sum_{\mathbf{e}\in F_\mu}y_\nu d_{\mu\mathbf{e}}^2P(U_\mathbf{e}\psi_{\mu\mathbf{e}})=\sum_{\mathbf{b}\in R}h_\mathbf{b}Y_\mathbf{b}\,,\end{aligned} \tag{3.12}$$

where $E_\nu:=\{\mathbf{e}\in\Gamma:Q_\mathbf{e}\varphi_\nu\neq0\}$, $F_\mu:=\{\mathbf{e}\in\Gamma:Q_\mathbf{e}\psi_\mu\neq0\}$,

$$\begin{aligned}&c_{\nu\mathbf{e}}:=\|Q_\mathbf{e}\varphi_\nu\|,\quad d_{\mu\mathbf{e}}:=\|Q_\mathbf{e}\psi_\mu\|\,,\\&\varphi_{\nu\mathbf{e}}:=(c_{\nu\mathbf{e}})^{-1}Q_\mathbf{e}\varphi_\nu\quad\text{for }\mathbf{e}\in E_\nu\,,\\&\psi_{\mu\mathbf{e}}:=(d_{\mu\mathbf{e}})^{-1}Q_\mathbf{e}\psi_\mu\quad\text{for }\mathbf{e}\in F_\mu\,,\\&g_\mathbf{a}:=\mathrm{tr}\Big(\sum_{\nu\in A}\sum_{\sigma(\mathbf{e})=\mathbf{a}}x_\nu c_{\nu\mathbf{e}}^2P(U_\mathbf{e}\varphi_{\nu\mathbf{e}})\Big)\,,\\&h_\mathbf{b}:=\mathrm{tr}\Big(\sum_{\mu\in B}\sum_{\sigma(\mathbf{e})=\mathbf{b}}y_\mu d_{\mu\mathbf{e}}^2P(U_\mathbf{e}\psi_{\mu\mathbf{e}})\Big)\,,\\&S:=\{\mathbf{a}\in\Gamma:g_\mathbf{a}\neq0\}\,,\quad R:=\{\mathbf{b}\in\Gamma:h_\mathbf{b}\neq0\}\,,\\&X_\mathbf{a}:=g_\mathbf{a}^{-1}\sum_{\nu\in A}\sum_{\sigma(\mathbf{e})=\mathbf{a}}x_\nu c_{\nu\mathbf{e}}^2P(U_\mathbf{e}\varphi_{\nu\mathbf{e}})\,,\end{aligned}$$

and

$$Y_\mathbf{b}:=h_\mathbf{b}^{-1}\sum_{\mu\in B}\sum_{\sigma(\mathbf{e})=\mathbf{b}}y_\mu d_{\mu\mathbf{e}}^2P(U_\mathbf{e}\psi_{\mu\mathbf{e}})\,.$$

Trivially, $\sum_{\mathbf{a}\in S}g_\mathbf{a}=1$, $\sum_{\mathbf{b}\in R}h_\mathbf{b}=1$ and $X_\mathbf{a}$, $Y_\mathbf{b}\in\mathcal{S}(\hat{\mathcal{H}})$ with $\mathrm{rng}X_\mathbf{b}\subseteq\hat{\mathcal{H}}_\mathbf{a}$, $\mathrm{rng}Y_\mathbf{b}\subseteq\hat{\mathcal{H}}_\mathbf{b}$, for all $\mathbf{a}\in S$, $\mathbf{b}\in R$. Moreover, $(S\cup R)\subset\mathrm{rng}\sigma$ so that $\mathbf{a},\mathbf{b}\in(S\cup R)$ and $\mathbf{a}\div\mathbf{b}$ imply that $\mathbf{a}=\mathbf{b}$. Hence we conclude from Lemma 3.7 that $\mathfrak{u}_\sigma X=\mathfrak{u}_\sigma Y$. ■

Corollary 3.8: For every $\mathbf{a}\in\Gamma$, $\mathfrak{L}$ contains at most one STO X with $\mathrm{rng}X\subseteq\hat{\mathcal{H}}_\mathbf{a}$.

Proof: If $\mathfrak{L}$ contains elements X,Y with $\mathrm{rng}X\subseteq\hat{\mathcal{H}}_\mathbf{a}$, $\mathrm{rng}Y\subseteq\hat{\mathcal{H}}_\mathbf{a}$, for some $\mathbf{a}\in\Gamma$, then we infer from Lemma 3.7 that $X=Y$. ■

Corollary 3.9: Suppose that $\mathfrak{L}$ contains a STO W with $\mathrm{rng}W\subseteq\hat{\mathcal{H}}_{w\mathbf{a}}$, for some $\mathbf{a}\in\Gamma$. Then

(i) $\mathrm{rng}X\subseteq\hat{\mathcal{H}}_{w\mathbf{a}}$ for all $X\in\mathfrak{L}$;

(ii) To every $\mathbf{b}\in\Gamma$ with $\mathbf{b}\div\mathbf{a}$, $\mathfrak{L}$ contains exactly one STO $W_\mathbf{b}$ with $\mathrm{rng}W_\mathbf{b}\subseteq\hat{\mathcal{H}}_\mathbf{b}$, and all these STO's are unitarily equivalent by some unitary operator of type $U(\mathfrak{H})$.

Proof: Assertion (i) follows immediately from (3.5) and the construction of the mappings $\mathfrak{u}_\sigma$. To prove (ii) we assume elements $\mathbf{b},\mathbf{c}\in\Gamma$ which are weakly equivalent to $\mathbf{a}$. Then there

exist mappings $\mathfrak{u}_\sigma$, $\mathfrak{u}_\mu$ of type (3.3) such that $\sigma(\mathbf{x}) = \mathbf{b}$, $\mu(\mathbf{x}) = \mathbf{c}$ for all $\mathbf{x} \div \mathbf{a}$. Hence $W_\mathbf{b} := \mathfrak{u}_\sigma W$ and $W_\mathbf{c} := \mathfrak{u}_\mu W$ are contained in $\mathfrak{L}$ with $\mathrm{rng} W_\mathbf{b} \subseteq \hat{\mathscr{H}}_\mathbf{b}$, $\mathrm{rng} W_\mathbf{c} \subseteq \hat{\mathscr{H}}_\mathbf{c}$. By Corollary 3.8, $W_\mathbf{b}$ and $W_\mathbf{c}$ are the only elements in $\mathfrak{L}$ with ranges in $\hat{\mathscr{H}}_\mathbf{b}$ and $\hat{\mathscr{H}}_\mathbf{c}$, respectively. According to (2.1) *infra*, there is a unitary operator U_1 of type $U(\mathfrak{H})$ with $U_1(\hat{\mathscr{H}}_\mathbf{b}) = \hat{\mathscr{H}}_\mathbf{c}$. Hence $\mathrm{rng}(U_1 W_\mathbf{b} U_1^*) \subseteq \hat{\mathscr{H}}_\mathbf{c}$. By Lemma 3.5, $U_1 W_\mathbf{b} U_1^* \in \mathfrak{L}$ and so we see from the above uniqueness argument that $U_1 W_\mathbf{b} U_1^* = W_\mathbf{c}$. ■

Up to now, no statement concerning the cardinality of $\mathfrak{L}$ has been made. In view of the following proposition, we recall the general assumption of this section, viz. that T is infinite.

Proposition 3.10: If $\mathfrak{L} \neq \emptyset$, then $\mathfrak{L}$ contains uncountably many mutually orthogonal STO's of the form $P(\psi)$ (called *pure states*).

Proof: By assumption, $\mathfrak{L}$ contains an element X. As a STO, X has a diagonal representation of the form (3.11). Hence $\mathfrak{q}X$ [cf. (3.1)] can be written as

$$\mathfrak{q}X = \sum_{\nu \in A} \sum_{\mathbf{e} \in E_\nu} x_\nu c_{\nu\mathbf{e}}^2 \, P(\varphi_{\nu\mathbf{e}}) \tag{13.13}$$

with the symbols introduced in (3.12) *infra*. Going over to a single index in (3.13) we get a countable convex decomposition of $\mathfrak{q}X$ into pure states,

$$\mathfrak{q}X = \sum_{i \in I} y_i \, P(\chi_i), \tag{3.13a}$$

where every unit vector χ_i is contained in some ICT. Since T is infinite, every WICT in $\hat{\mathscr{H}}$ contains uncountably many ICT's.[11] So we can choose, to every $i \in I$, uncountably many unitary operators U_{il} of type $U(\mathfrak{H})$ such that the U_{il}'s all lie in *different* ICT's. Hence the unit vectors

$$\Phi_l := \sum_{i \in I} \sqrt{y_i} \; U_{il} \chi_i$$

are mutually orthogonal and

$$\mathfrak{q} P(\Phi_l) = \sum_{i \in I} y_i \, U_{il} \, P(\chi_i) \, U_{il}^* .$$

From (3.13a) and Lemma 3.5 we conclude that $P(\Phi_l) \in \mathfrak{L}$ for all l. ■

We close this section by exhibiting two classes of strongly exposed points of $\mathfrak{L}$.

Proposition 3.11: (i) Let $\mathscr{K}$ be a Hilbert space and $0 \neq \varphi \in \mathscr{K}$. The pure state $P(\varphi)$ is a strongly exposed point [in the Banach space $\mathscr{T}(\mathscr{K})$] of any subset of $\mathscr{S}(\mathscr{K})$ which contains $P(\varphi)$.

(ii) If $\mathfrak{L}$ contains an element Y with $\mathrm{rng} Y \subseteq \hat{\mathscr{H}}_\mathbf{b}$ for some $\mathbf{b} \in \Gamma$, then Y is a strongly exposed point of $\mathfrak{L}$.

Proof: (i) It is obviously sufficient to prove that $P(\varphi)$ is a strongly exposed point of $\mathscr{S}(\mathscr{K})$. The functional $f \in \mathscr{T}(\mathscr{K})^*$, defined by $f(X) := \langle \varphi, X\varphi \rangle$, exposes $P(\varphi)$ in $\mathscr{S}(\mathscr{K})$ since $\|f\| = 1$ and $\langle \varphi, P(\varphi)\varphi \rangle = 1 > \langle \varphi, X\varphi \rangle$ for all $X \in \mathscr{S}(\mathscr{K})$, $X \neq P(\varphi)$. Let W be a STO in $\mathscr{S}(\mathscr{K})$ and

$$W = \sum_{i \in I} w_i \, P(\alpha_i)$$

a diagonal representation of W. Then

$$\begin{aligned}
&\|W - P(\varphi)\|_1^2 \\
&\quad = \Big\| \sum_i w_i (P(\alpha_i) - P(\varphi)) \Big\|_1^2 \\
&\quad \leqslant \sum_i w_i \, \|P(\alpha_i) - P(\varphi)\|_1^2 \\
&\quad = 4 \Big\{ \sum_i w_i (1 - |\langle \alpha_i, \varphi \rangle|^2) \Big\} = 4(1 - \langle \varphi, W\varphi \rangle) \\
&\quad = 4[f(P(\varphi)) - f(W)],
\end{aligned}$$

where we used the equation $\|P(\alpha) - P(\beta)\|_1^2 = 4(1 - |\langle \alpha, \beta \rangle|^2)$ from the proof of Corollary A.4(f) in Ref. 1. So $P(\varphi)$ is a strongly exposed point of $\mathscr{S}(\mathscr{K})$.

(ii) If α is a unit vector in $\mathscr{K}$ and R any projection in $\mathscr{B}(\mathscr{K})$, then one easily checks that

$$\|RP(\alpha)R^\perp + R^\perp P(\alpha) R\|_1 \leqslant 2\|R^\perp \alpha\| = 2\langle \alpha, R^\perp \alpha \rangle^{1/2} . \tag{3.14}$$

Now, let $Y \in \mathfrak{L}$ and $\mathrm{rng} Y \subseteq \hat{\mathscr{H}}_\mathbf{b}$. Then we choose an injective mapping $\gamma: \Gamma \to \Gamma$ such that $\gamma(\mathbf{x}) = \gamma(\mathbf{y}) \Longleftrightarrow \mathbf{x} \div \mathbf{y}$, $\gamma(\mathbf{x}) \div \mathbf{x}$, and $\gamma(\mathbf{a}) = \mathbf{b}$ for all $\mathbf{a} \div \mathbf{b}$. Since $Q_\mathbf{x} Y Q_\mathbf{x} = 0$ for all $\mathbf{x} \neq \mathbf{b}$, we find $\mathfrak{u}_\gamma Y = Y$ and, by Theorem 3.4,

$$\mathfrak{L} = \{ W \in \mathscr{S}(\hat{\mathscr{H}}) : \mathfrak{u}_\gamma W = Y \} .$$

Hence to every $W \in \mathfrak{L}$ belongs a subset $\{\mathbf{a}_i \in \Gamma : i \in \mathbf{N}\}$ of Γ such that $\mathbf{a}_1 = \mathbf{b}$ and

$$Y = \sum_{i=1}^{\infty} U_{\mathbf{a}_i} Q_{\mathbf{a}_i} W Q_{\mathbf{a}_i} U_{\mathbf{a}_i}^* , \tag{3.15}$$

where we can assume that $U_{\mathbf{a}_1} = \mathbf{1}_T$. Let $W \in \mathfrak{L}$ and let

$$W = \sum_{j \in J} w_j \, P(\varphi_j)$$

be a diagonal representation of W. Then

$$\|W - Q_\mathbf{b} W Q_\mathbf{b}\|_1 \leqslant \|Q_\mathbf{b}^\perp W Q_\mathbf{b}^\perp\|_1 + \|Q_\mathbf{b} W Q_\mathbf{b}^\perp + Q_\mathbf{b}^\perp W Q_\mathbf{b}\|_1 , \tag{3.16}$$

$$\|Q_\mathbf{b}^\perp W Q_\mathbf{b}^\perp\|_1 = \mathrm{tr}(Q_\mathbf{b}^\perp W) = 1 - \mathrm{tr}(Q_\mathbf{b} W) , \tag{3.17}$$

$$\begin{aligned}
&\|Q_\mathbf{b} W Q_\mathbf{b}^\perp + Q_\mathbf{b}^\perp W Q_\mathbf{b}\|_1 \\
&\quad \leqslant \sum_{j \in J} w_j \|Q_\mathbf{b} \, P(\varphi_j) \, Q_\mathbf{b}^\perp + Q_\mathbf{b}^\perp \, P(\varphi_j) \, Q_\mathbf{b}\|_1 .
\end{aligned} \tag{3.18}$$

Equations (3.14) and (3.18) yield

$$\begin{aligned}
&\|Q_\mathbf{b} W Q_\mathbf{b}^\perp + Q_\mathbf{b}^\perp W Q_\mathbf{b}\|_1 \leqslant 2 \sum_{j \in J} w_j \, \langle \varphi_j, Q_\mathbf{b}^\perp \varphi_j \rangle^{1/2} \\
&\leqslant 2 \Big(\sum_{j \in J} w_j \langle \varphi_j, Q_\mathbf{b}^\perp \varphi_j \rangle \Big)^{1/2} = 2[1 - \mathrm{tr}(Q_\mathbf{b} W)]^{1/2} .
\end{aligned} \tag{3.19}$$

Combining (3.16), (3.17), and (3.19) we get

$$\|W - Q_\mathbf{b} W Q_\mathbf{b}\|_1 \leqslant 3[1 - \mathrm{tr}(Q_\mathbf{b} W)]^{1/2} . \tag{3.20}$$

From (3.15), (3.20) and from

$$\begin{aligned}
&\Big\| \sum_{i=2}^{\infty} U_{\mathbf{a}_i} Q_{\mathbf{a}_i} W Q_{\mathbf{a}_i} U_{\mathbf{a}_i}^* \Big\|_1 \\
&\quad \leqslant \sum_{i=2}^{\infty} \|Q_{\mathbf{a}_i} W Q_{\mathbf{a}_i}\|_1 \\
&\quad = \sum_{i=2}^{\infty} \mathrm{tr}(Q_{\mathbf{a}_i} W) = 1 - \mathrm{tr}(Q_\mathbf{b} W)
\end{aligned}$$

we finally arrive at

$$\|Y - W\|_1^2 \leqslant 16[1 - \mathrm{tr}(Q_\mathbf{b} W)] \tag{3.21}$$

for all $W\in\mathfrak{L}$. So, if $(V_n)_{n\in\mathbb{N}}$ is any sequence in $\mathfrak{L}$ with $\mathrm{tr}(Q_\mathfrak{b} V_n)\to\mathrm{tr}(Q_\mathfrak{b} Y)=1$, then we see from (3.21) that $\|V_n - Y\|_1\to 0$. Moreover, the functional $X\mapsto\mathrm{tr}(Q_\mathfrak{b} X)$ in $\mathcal{T}(\hat{\mathcal{H}})^*$ exposes Y in $\mathfrak{L}$ because Corollary 3.8 and $Y\leqslant Q_\mathfrak{b}$ imply that $\mathrm{tr}(Q_\mathfrak{b} W) < \mathrm{tr}(Q_\mathfrak{b} Y)=1$ for all $W\in\mathfrak{L}$, $W\neq Y$. So Y is strongly exposed in $\mathfrak{L}$. ■

4. PROJECTIVE LIMITS ON INCOMPLETE TENSOR PRODUCTS

The ambiguity of the projective limit of a projective system $\{W_K:K\in\mathcal{F}(T)\}$ of STO's is due to the special construction of the (so called) *complete* tensor product $\hat{\mathcal{H}}=\otimes_{t\in T}\mathcal{H}_t$ which, in some respect, is "too big." To avoid this ambiguity, one can think of allowing only those projective limits whose ranges are contained in ICT's which are the smallest subspaces of $\hat{\mathcal{H}}$ that can reasonably be regarded as tensor products on their own. As guaranteed by Corollary 3.8, these particular projective limits are unique if they exist at all.

Theorem 4.1: The projective system $\{W_K:K\in\mathcal{F}(T)\}$ has a projective limit with its range contained in $\hat{\mathcal{H}}[\alpha]$ if and only if the operator net

$$(W_K\otimes P(\alpha[T\setminus K]))_K$$

converges weakly to a STO, say V, on $\hat{\mathcal{H}}$. In this case, $\mathrm{rng}V\subseteq\hat{\mathcal{H}}[\alpha]$ and V is the *unique* projective limit with this property.

Proof: (I) Assume that $(W_K\otimes P(\alpha[T\setminus K]))_K$ converges weakly to $V\in\mathcal{S}(\hat{\mathcal{H}})$. Then[12]

$$\lim_K\|W_K\otimes P(\alpha[T\setminus K])-V\|_1=0. \tag{4.1}$$

Since $Q_\mathfrak{b}(W_K\otimes P(\alpha[T\setminus K]))=0$ for all $\mathfrak{b}\neq[\alpha]$,

$$\mathrm{rng}V\subseteq\hat{\mathcal{H}}[\alpha]. \tag{4.2}$$

If $V=\Sigma_{i\in I}v_iP(\psi_i)$ is a diagonal representation of V, then to every $i\in I$ there is a net $(\varphi_{iK})_K$ of unit vectors $\varphi_{iK}\in\mathcal{H}^K$ such that

$$\lim_K\|\psi_i-\varphi_{iK}\otimes\alpha[T\setminus K]\|=0.$$

Hence

$$\lim_H\left\{\sup_{\substack{K\in\mathcal{F}\\K\supset H}}\|\varphi_{iK}-\varphi_{iH}\otimes\alpha[K\setminus H]\|\right\}=0 \tag{4.3}$$

and

$$\lim_K\left|\left|V-\sum_{i\in I}v_iP(\varphi_{iK}\otimes\alpha[T\setminus K])\right|\right|_1=0. \tag{4.4}$$

From (4.1), (4.4), and from

$$\begin{aligned}&\left|\left|W_K-\sum_{i\in I}v_iP(\varphi_{iK})\right|\right|_1\\&=\left|\left|\left\{W_K-\sum_{i\in I}v_iP(\varphi_{iK})\right\}\otimes P(\alpha[T\setminus K])\right|\right|_1\\&\leqslant\|W_K\otimes P(\alpha[T\setminus K])-V\|_1\\&\quad+\left|\left|V-\sum_{i\in I}v_iP(\varphi_{iK}\otimes\alpha[T\setminus K])\right|\right|_1,\end{aligned}$$

we get

$$\lim_K\left|\left|W_K-\sum_{i\in I}v_iP(\varphi_{iK})\right|\right|_1=0. \tag{4.5}$$

Finally, we conclude from (4.3), (4.5) and Theorem 2.1 of Ref. 1 that $V\in\mathfrak{L}$, and we conclude from (4.2) and Corollary 3.6 that V is the only element in $\mathfrak{L}$ with $\mathrm{rng}V\subseteq\hat{\mathcal{H}}[\alpha]$.

(II) Assume that $V\in\mathfrak{L}$, $\mathrm{rng}V\subseteq\hat{\mathcal{H}}[\alpha]$ and let $V=\Sigma_{i\in I}v_iP(\psi_i)$ be some diagonal representation of V. By specializing part (II) of the proof of Theorem 2.1 in Ref. 1 to this case, we see that, to every $i\in I$, there is a family $\{\varphi_{iK}:K\in\mathcal{F}\}$ of unit vectors $\varphi_{iK}\in\mathcal{H}^K$ such that

$$\lim_K\|\varphi_{iK}\otimes\alpha[T\setminus K]-\psi_i\|=0\quad(\forall\ i\in I) \tag{4.6}$$

and

$$\lim_K\left|\left|W_K-\sum_{i\in I}v_iP(\varphi_{iK})\right|\right|_1=0. \tag{4.7}$$

With $\Phi_{iK}:=\varphi_{iK}\otimes\alpha[T\setminus K]$ we infer from (4.6), (4.7), and from a dominated convergence argument that

$$\lim_K\left|\left|W_K\otimes P(\alpha[T\setminus K])-\sum_{i\in I}v_iP(\Phi_{iK})\right|\right|_1=0 \tag{4.8}$$

and

$$\lim_K\left|\left|V-\sum_{i\in I}v_iP(\Phi_{iK})\right|\right|_1=0. \tag{4.9}$$

From (4.8), (4.9), and from

$$\begin{aligned}&\|W_K\otimes P(\alpha[T\setminus K])-V\|_1\\&\leqslant\left|\left|W_K\otimes P(\alpha[T\setminus K])-\sum_{i\in I}v_iP(\Phi_{iK})\right|\right|_1\\&\quad+\left|\left|\sum_{i\in I}v_iP(\Phi_{iK})-V\right|\right|_1,\end{aligned}$$

it follows that $(W_K\otimes P(\alpha[T\setminus K]))_K$ converges weakly to V. ■

Suppose now that only those projective limits are to be considered whose ranges are contained in ICT's. In this case, it seems to be more appropriate to deal from the outset with incomplete tensor products only, defined independently and not as subspaces of $\hat{\mathcal{H}}$,[13] and to adapt the notion of the projective limit to the ICT as the proper tensor product of Hilbert spaces. If we want to refer to the ICT associated with the family $(\alpha_t)_{t\in T}$ of unit vectors $\alpha_t\in\mathcal{H}_t$ and defined independently of $\hat{\mathcal{H}}$, then we use the notation $\otimes_{t\in T}(\mathcal{H}_t,\alpha_t)$ instead of $\hat{\mathcal{H}}[\alpha]$. By $\mathrm{tr}_{[\alpha]}$ and tr_K we denote the trace functionals on $\mathcal{T}(\otimes_{t\in T}(\mathcal{H}_t,\alpha_t))$ and $\mathcal{T}(\mathcal{H}^K)$, respectively, and $Y_{[\alpha]}$ denotes the cylinder operator on $\otimes_{t\in T}(\mathcal{H}_t,\alpha_t)$ associated with $Y\in\mathcal{B}(\mathcal{H}^K)$.

Definition: Let $\{W_K:K\in\mathcal{F}(T)\}$ be a projective system of STO's W_K on $\mathcal{H}^K$. A STO V on $\otimes_{t\in T}(\mathcal{H}_t,\alpha_t)$ is called an $[\alpha]$-*projective limit* of $\{W_K\}$, if

$$\mathrm{tr}_K(W_KY)=\mathrm{tr}_{[\alpha]}(VY_{[\alpha]})$$

for all $Y\in\mathcal{B}(\mathcal{H}^K)$ and $K\in\mathcal{F}(T)$.

By identifying the product vectors in $\hat{\mathcal{H}}$ and $\otimes_{t\in T}(\mathcal{H}_t,\alpha_t)$ which are associated with the same family $(\beta_t)_{t\in T}$ of unit vectors, we get a canonical isomorphism between $\otimes_{t\in T}(\mathcal{H}_t,\alpha_t)$ and $\hat{\mathcal{H}}[\otimes_t\alpha_t]$. This isomorphism im-

plies that the projective system $\{W_K : K \in \mathcal{F}(T)\}$ has an $[\alpha]$-projective limit V if and only if it has a projective limit X on $\hat{\mathcal{H}}$ with $\mathrm{rng} X \subseteq \hat{\mathcal{H}}[\alpha]$ and that, in this case, V can be identified with $X \restriction \hat{\mathcal{H}}[\alpha]$. We therefore obtain the following necessary and sufficient condition for the existence of $[\alpha]$-projective limits, a result which has been independently established by Bartoszewicz without resort to Theorem 2.1 of Ref. 1.[14]

Corollary 4.2: A projective system $\{W_K : K \in \mathcal{F}(T)\}$ has an $[\alpha]$-projective limit if and only if the net $(W_K \otimes P(\alpha[T \setminus K]))_K$ in $\mathcal{L}(\otimes_{t \in T}(\mathcal{H}_t, \alpha_t))$ converges weakly to a STO, say V, on $\otimes_{t \in T}(\mathcal{H}_t, \alpha_t)$. In this case, V is the *unique* $[\alpha]$-projective limit of $\{W_K\}$.

ACKNOWLEDGMENTS

It is a pleasure to thank D. Castrigiano and H. Spohn for their critical reading of the manuscript. Special thanks are due to H.W. von Weizsäcker for calling my attention to References 6 to 8 and for explaining the proof of Proposition 3.3 to me.

APPENDIX

In this appendix we show that $\mathcal{T}(\hat{\mathcal{H}})$ has the Radon–Nikodym property. For the standard notions of vector measure theory used below, the reader is referred to Ref. 7.

Definition: A (real or complex) Banach space X has the *Radon–Nikodym property* (RNP) if, for every finite measure space $(\Omega, \mathcal{A}, \mu)$ and every μ-continuous vector measure $G : \mathcal{A} \to X$ of bounded variation,

$$G(E) = \text{Bochner} - \int_E g \, d\mu$$

for some $g \in L_1(\mu, X)$ and all $E \in \mathcal{A}$.

Theorem A.1[7]: The following statements about a (real or complex) Banach space X are equivalent:

(i) X has the RNP;

(ii) every closed subspace of X has the RNP;

(iii) every separable closed subspace of X is isomorphic to a subspace of a separable dual space.

Corollary A.2: For every Hilbert space $\mathcal{K}$, the Banach space $\mathcal{T}(\mathcal{K})$ has the RNP.

Proof: Let $\mathcal{Y}$ be any separable closed subspace of $\mathcal{T}(\mathcal{K})$. Hence there is a sequence $(A_i)_{i \in \mathbb{N}}$ in $\mathcal{Y}$ such that

$$(\forall C \in \mathcal{Y})(\forall \epsilon > 0)(\exists j \in \mathbb{N}) \quad \|C - A_j\|_1 \leqslant \epsilon. \tag{A1}$$

To every compact operator D on $\mathcal{K}$, we define the closed subspace

$$L(D) := (\overline{\mathrm{rng}\ D}) \vee ((\ker D)^{\perp})$$

of $\mathcal{K}$ which is separable. Hence the closed subspace

$$\mathcal{H} := \bigvee_{i=1}^{\infty} L(A_i)$$

of $\mathcal{K}$ is also separable. If Q denotes the projection onto $\mathcal{H}$, then $A_i = QA_iQ$ for all $i \in \mathbb{N}$. Hence

$$\|C - A_i\|_1 \geqslant \|Q(C - A_i)Q\|_1 = \|QCQ - A_i\|_1 \tag{A2}$$

for all $C \in \mathcal{Y}$ and $i \in \mathbb{N}$. Eqs (A1) and (A2) imply that $C = QCQ$ or, equivalently, $L(C) \subseteq \mathcal{H}$ for all $C \in \mathcal{Y}$. Hence $\mathcal{Y}$ is isomorphic to a subsapce of $\mathcal{T}(\mathcal{H})$. Now, $\mathcal{T}(\mathcal{H})$ is separable since $\mathcal{H}$ is separable, and it is well known that $\mathcal{T}(\mathcal{H})$ is a dual space.[3] So we infer from Theorem A.1 that $\mathcal{T}(\mathcal{K})$ has the RNP. ■

[1] W. Ochs, J. Math. Phys. **20**, 1842 (1979).
[2] H. Bauer, *Wahrscheinlichkeitstheorie und Grundzüge der Masstheorie* (de Gruyter, Berlin, 1968).
[3] For details of these notions we refer to M. Reed and B. Simon, *Methods of Modern Mathematical Physics. Vol. I: Functional Analysis* (Academic, New York, 1972) and to R. Schatten, *Norm Ideals of Completely Continuous Operators* (Springer, Berlin, 1960).
[4] J. von Neumann, Composito Math. **6**, 1 (1938).
[5] This follows from Lemmas 6.3.1, 6.3.2, and 6.3.5 of Ref. 4.
[6] R.P. Phelps, J. Funct. Anal. **16**, 78 (1974).
[7] J. Diestel and J.J. Uhl, Jr., *Vector Measures* (Amer. Math. Soc., Providence, 1977).
[8] G.A. Edgar, Proc. Amer. Math. Soc. **49**, 354 (1975).
[9] From now on we will frequently use infinite sums of positive operators without further explanation. These operator sums are to be understood as the limits of the nets of their finite partial sums in the strong operator topology. In all cases occurring, these limits exist because all partial sums are uniformly bounded by the unit operator.
[10] This follows from the proof of Lemma 6.3.3 in Ref. 4.
[11] See Lemma 6.4.1 of Ref. 4.
[12] A. Wehrl, Reports Math. Phys. **10**, 159 (1976).
[13] See Appendix a in A. Guichardet, *Symmetric Hilbert Spaces and Related Topics*, Lecture Notes in Math. 261 (Springer, Berlin, 1972).
[14] A. Bartoszewicz, Colloq. Math. **39**, 141 (1978).

The domain of definition of bundle representations

R. N. Sen
Department of Mathematics, Ben Gurion University of the Negev, Beersheba 84120, Israel

(Received 7 August 1979; accepted for publication 11 January 1980)

It is proved that the equivalence class of bundle representations of a group G on the product bundle $\mathscr{B}_0$ with total space $B_0 = X \times Y$ includes all representations of G on bundles $\mathscr{B}$ which are homeomorphic (but not necessarily naturally homeomorphic) to the product $X \times Y$, provided that G has the same action on the fibres of $\mathscr{B}_0$ and $\mathscr{B}$. The group A of the bundle $\mathscr{B}$ is immaterial.

PACS numbers: 02.40. + m

1. INTRODUCTION

In the definition of bundle representations,[1,2] the underlying total space B_0 was taken as a *natural* product, $B_0 = X \times \mathscr{H}$. Since there exist bundles which are homeomorphic,[3] but not naturally homeomorphic, to the product, it is reasonable to ask how one can represent groups on such bundles. The answer is the following: The equivalence class of bundle representations contains all representations on bundles which are homeomorphic to the product, irrespective of whether this homeomorphism is natural or otherwise. In this paper we shall prove this assertion.

We shall start with a remark on terminology. Steenrod[4] defines a *product bundle* as one with a distinguished fibre-preserving homeomorphism to the product, and with the identity as the group of the bundle. Dieudonné[5] calls such bundles *trivial*. To differentiate between these, and bundles which are homeomorphic but *not naturally homeomorphic* to the product, Steenrod introduces the notion of *G-equivalence* to the product.[6] Dieudonné calls such bundles *trivializable*,[7] and simultaneously emphasizes the distinction between trivial and trivializable bundles.

If one is interested only in topological invariants, which is generally the case in the mathematical literature, there is no point in distinguishing between trivial and trivializable bundles. Consequently, in much of the mathematical literature one encounters the term "trivial bundles" without further specification. A topologically trivial bundle may be either trivial or trivializable in the sense of Dieudonné. In Steenrod's book the statement that a bundle is "equivalent to the product" is stipulated to include the case that it is G-equivalent to the product.[8]

However, in some physical applications of fibre bundle theory which are currently under investigation, the nontopological distinction between trivial and trivializable bundles appears to be physically significant. In the geometrical formulations of some gauge theories, a gauge transformation corresponds to the passage from the one global trivialization to another; the principal bundles involved are homeomorphic[9] to the product, but there exists no distinguished homeomorphism.[10,11]

In Sec. 2 we shall define group representations on general fibre bundles (not necessarily vector bundles), and in Sec. 3 we shall establish the main result via two elementary propositions, one of which is proved in detail. Finally we shall sketch, in the Appendix, the construction of trivial and trivializable bundles according to the Whitney–Steenrod theorem, in order to provide a complete clarification of the difference between the two.

2. GENERAL DEFINITION OF GROUP REPRESENTATIONS BY BUNDLE MAPS

Let $\mathscr{B} = [B, X, \pi, A, Y]$ be a fibre bundle[12] with total space B, bases space X, projection π, group A, and fibre Y. Let G be a topological group. We define a bundle representation of G on $\mathscr{B}$ to be a family $\{\lambda(g)\}$ of invertible bundle maps[13] $\lambda(g): B \to B$, indexed by $g \in G$, which satisfy the following conditions[2]:

(i) $\lambda(e)$ = the identity map, e is the identity of G:

(ii) $\lambda(g_1) \circ \lambda(g_2) = \lambda(g_1 g_2)\ \forall g_1, g_2 \in G$;

(iii) the map $\bar{\lambda}: G \times B \to B$ defined by $\bar{\lambda}(g, b) = \lambda(g)b$ is continuous (here $b \in B$).

Clearly the family of bundle maps $\{\lambda(g)\}$ induces a continuous action $\{\bar{\lambda}(g)\}$ of G on X.

3. THE MAIN RESULT

We now assume that (a) $\mathscr{B}$ is homeomorphic to $\mathscr{B}_0 = X \times Y$ {the bundle $\mathscr{B}_0$ is the collection $[B_0, X, \pi_0, Y]$}, but (b) there exists no natural or distinguished homeomorphism between B and B_0. We denote ϕ_i a global trivialization of B i.e.,

$$\phi_i: X \times Y \to B \tag{1}$$

is a homeomorphism such that

$$\pi \circ \phi_i(x, y) = x\ \forall x \in X,\ y \in Y. \tag{2}$$

Here i belong to an indexing set J.

Proposition 1: A global trivialization ϕ_i of B induces, from the representation $\{\lambda(g)\}$ of G on $\mathscr{B}$, a bundle representation (in the ordinary sense) $\{h_i(g)\}$ of G on $\mathscr{B}_0$, where

$$h_i(g) = \phi_i^{-1} \circ \lambda(g) \circ \phi_i. \tag{3}$$

If $\{h_i(g)\}$ and $\{h_j(g)\}$ are the bundle representations of G on $\mathscr{B}_0$ induced by the trivializations ϕ_i and ϕ_j, respectively, then

$$h_i(g) = \Omega_{ji}^{-1} h_j \Pi g) \Omega_{ji}, \tag{4}$$

where Ω_{ji} is a coordinate transformation on the fibres[14] of $\mathscr{B}_0$, given by

$$\Omega_{ji} = \phi_j^{-1} \circ \phi_i, \tag{5}$$

 0022-2488/81/020290-03$1.00

which acts on each fibre as an element of the group A of the bundle $\mathscr{B}$.

Proof: (a) $h_i(g)$ is defined by the commutative diagram of Fig. 1:

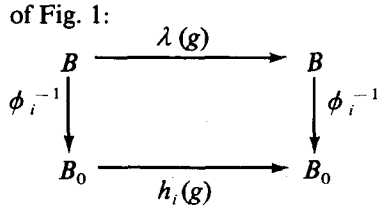


FIG. 1.

Since $\lambda(g)$ and ϕ_i are fibre-preserving homeomorphisms, so is $h_i(g)$. The algebraic representation properties [i.e., (i), (ii) of Sec. 2] of $\{h_i(g)\}$ follow by inspection. The continuity property of $\tilde{h}_i$ [defined by analogy with $\tilde{\lambda}$, (iii) of Sec. 2; cf. also (iii), Eq. (6), Ref. 2] are set in evidence by the diagram in Fig. 2:

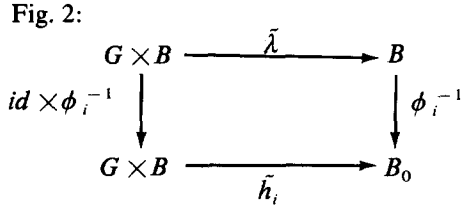


FIG. 2.

In Fig. 2$(id \times \phi_i^{-1})(g, b) = (g, \phi_i^{-1}(b))$, $b \in B$, and one verifies immediately that the diagram commutes.

(b) For every pair ϕ_i, ϕ_j of global trivializations, we define a continuous map[15]

$$a_{ji}: X \to A,$$

by

$$a_{ji}(x) = \phi_{j,x}^{-1} \circ \phi_{i,x},$$

where the map $\phi_{i,x}: Y \to \pi^{-1}(x)$ is defined by

$$\phi_{i,x}(y) = \phi_i(x, y).$$

The $a_{ji}(x)$ satisfy

$$a_{ji}(x) a_{ik}(x) = a_{jk}(x), \tag{6}$$

and

$$[a_{ji}(x)]^{-1} = a_{ij}(x). \tag{7}$$

Now consider the map

$$\Omega_{ji} = \phi_j^{-1} \circ \phi_i: B_0 \to B_0$$

and evaluate $\phi_j^{-1} \circ \phi_i(x,y)$. Let $\phi_i(x, y) = b$, then $\pi(b) = x$, and $\phi_j^{-1} \circ \phi_i(x,y) = \phi_j^{-1}(b) = (x', y')$ (say). Then $b = \phi_j(x', y')$ and $\pi(b) = x'$, whence $x' = x$. Thus we have proved that Ω_{ji} acts as the identity on the base space, and is therefore a coordinate transformation on the fibres. We have, furthermore

$$a_{ji}(x) = \Omega_{ji}(x),$$

and, as is evident upon inspection, the Ω_{ji} satisfy the same relations (6) and (7) as the $a_{ji}(x)$.

Proposition 2: Let $\{h(g)\}$ be a bundle representation of G on $\mathscr{B}_0$ and let ϕ_i be a trivialization of B. Then $\{\lambda_i(g)\}$, defined by

$$\lambda_i(g) = \phi_i \circ h(g) \circ \phi_i^{-1}$$

is a representation (in the sense of Sec. 2) of G on $\mathscr{B}$. If ϕ_i, ϕ_j are two trivializations of B and $\{\lambda_i(g)\}$, $\{\lambda_j(g)\}$ the corresponding representations of G on $\mathscr{B}$, then

$$\lambda_j(g) = T_{ji} \circ \lambda_i(g) \circ T_{ji}^{-1},$$

where $T_{ji}: B \to B$, defined by

$$T_{ji} = \phi_j \circ \phi_i^{-1},$$

is a coordinate transformation on the fibres of $\mathscr{B}$; that is, the representations $\{\lambda_j(g)\}$ and $\{\lambda_i(g)\}$ are equivalent. Moreover, the T_{ji} satisfy the relations

$$T_{ji} T_{ik} = T_{jk},$$

$$T_{ji}^{-1} = T_{ij}.$$

Proof: Similar to the proof of Proposition 1. The results we have obtained above can be summarized in the following theorem.

Theorem: The equivalence class of bundle representations of the group G on $\mathscr{B}_0 = [B_0, X, \pi_0, Y]$ contains all representations of G *on all bundles* $\mathscr{B} = [B, X, \pi, A, Y]$ which are homeomorphic to $\mathscr{B}_0$ by fibre-preserving homeomorphisms, *irrespective of the group* A (provided that the group G has the same action on the fibres of $\mathscr{B}_0$ and $\mathscr{B}$).

4. CONCLUDING REMARKS

The result proved above may appear surprising at first glance, but ceases to remain so after a moment's reflection. The notion of a bundle representation, if properly formulated, ought to be invariant under fibre-preserving topological equivalences. Such a definition is easy to give but harder to use for constructive or computational purposes. The situation is exactly the same as in the theory of linear representations of groups. There a linear representation is defined in a coordinate-free manner as a homomorphism of the group into the group of linear transformations of a vector space. However, to compute one often has to fall back upon *matrix* representations, which require a choice of basis. One then eliminates the basis dependence by passage to equivalence classes. An "abstract" linear representation may be considered as an (appropriately defined) equivalence class of "concrete" matrix representations. So it is with bundle representations. An "abstract" bundle representation is the equivalence class, under coordinate transformations, of a "concrete" bundle representation which is defined, explicitly and constructively, in terms of a specific global trivialization. Since we admit *all fibre-preserving homeomorphims* as coordinate transformations, the nontopological distinction between trivial and trivializable bundles disappears in the passage from the concrete to the abstract.

As a byproduct of these considerations we arrive at the following conclusion. Suppose that a bundle with group G is G-equivalent to the product. If we want to construct a physical theory (using this bundle) in which G has a physical meaning, then the mathematical structure which is significant is a *G-structure*[16] *on the base space* X, and not the topological structure of the bundle. In such contexts principal bundles are only a convenient means for studying the geometry of G-structures.

Many applications of fibre bundle theory which are being considered appear to be of this kind. These include, apart from the usual gauge theories based on Minkowski space, the works of A. Böhm[17] and W. Drechsler.[18] Moreover, Minkowski space itself appears to admit a (3,2) de Sitter group structure, first noticed by Dirac.[19] van Dam and Bie-

denharn[20] have made the remarkable suggestion that this structure be used to bypass O'Raifeartaigh's theorem.[21] In view of these works, it might be useful, for achieving greater clarity, to shift the emphasis in such contexts from fibre bundles to G-structures. The latter should be viewed as the primary objects with physical meaning, and principal bundles as auxiliary mathematical constructs for studying them.

APPENDIX: TRIVIAL AND TRIVIALIZABLE BUNDLES

In all cases, a fibre bundle is an object which is assembled, painstakingly, from its ingredients by the Whitney–Steenrod construction, given in Sec. 3 of Ref. 4. We follow the notations and terminology of this reference.

Recall that a coordinate bundle is called a product bundle if there is just *one* coordinate neighborhood $V = X$, *and* the group G of the bundle consists of the identity alone. Applying the Whitney–Steenrod construction to this case, we obtain the total space $B_0 = X \times Y$, the natural projection $p: B_0 \to X$, and the identity map as the coordinate function. A fibre bundle is a strict equivalence class of coordinate bundles. To determine this class, we take any open cover $\{V_i\}$ of X consisting of more than one chart, and $G = \{e\}$. Working out the Whitney–Steenrod construction for this case, we find again that the total space is $X \times Y$, ϕ_j is the restriction of the identity map to $V_j \times Y$ for each $j \in J$, and the projection is the natural one. The bundle is trivial.

Now suppose that we are given the following: there is only one coordinate neighborhood $V = X$; there exists an effective topological tranformation group G of Y, an indexing set J, and a family of continuous nontrivial maps $g_{ji}: X \to G$ for each pair i, j of indices in J. The conditions of the Whitney–Steenrod construction are met. The construction gives us a bundle with total space B, and for each $j \in J$ a coordinate function

$$\phi_j: X \times Y \to B,$$

such that

$$\phi_{j,x}^{-1} \circ \phi_{i,x} = g_{ji}(x) \quad \forall x \in X.$$

The space B is homeomorphic, in a fibre-preserving manner, to $B_0 = X \times Y$, but there exists no natural or distinguished homeomorphism. The bundle obtained is trivializable but not trivial. The passage from one global trivialization to another is accomplished with the help of the group G of the bundle.

[1]Bundle representations were first defined in R.N. Sen, *Nonrelativistic Zero-Mass Systems*, Lecture Notes, Göttingen (1974); H.J. Borchers and R.N. Sen, Commun. Math. Phys. **42**, 101 (1975). The definition was strengthened, as regards continuity, in Ref. 2.

[2]R.N. Sen, Physica (Utrecht) A **94**, 39 (1978). We shall be working with this latter definition.

[3]We shall work exclusively with bundle homeomorphisms which are fibre presserving. In $B_0 = X \times \mathscr{H}$, X is the base space and $\mathscr{H}$ the fibre.

[4]N. Steenrod, *The Topology of Fibre Bundles* (Princeton U.P., Princeton, New Jersey, 1951, with an appendix added November 1956), see p. 16.

[5]J. Dieudonné, *Treatise on Analysis* (Academic, New York, 1972), Vol. III, see p. 78.

[6]See Ref. 4, p. 17.

[7]See Ref. 5, pp. 78 and 79.

[8]See Ref. 4, Secs. 4.3 and 4.4.

[9]In the geometrical formulations of gauge theories one works with differentiable rather than topological structures; however, this extra structure is not relevant in the present context.

[10]See, for example, W. Drechsler and M.E. Mayer, *Fibre Bundle Techniques in Gauge Theories*, Vol. 67 of *Lecture Notes in Physics* (Springer-Verlag, Berlin, 1977), and various articles in the following reference.

[11]*Differential Geometrical Methods in Mathematical Physics*, edited by K. Bleuler *et al.*, Vols. 570 (1977) and 676 (1978) of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin).

[12]In order to avoid any misunderstanding, we state explicitly that we use the term "fibre bundle" in the sense of Steenrod, Ref. 4. For various definitions of fibre bundles and fibre spaces, see, for example, W.S. Massey, "Some problems in algebraic topology and the theory of fibre bundles," Ann. Math. **62**, 327 (1955). D. Husemoller, *Fibre Bundles*, 2nd ed. (Springer-Verlag, GTM 20, 1975) calls a "bundle" what is called an *espace* découpé by R. Godement, *Théorie des faisceaux* (Hermann, Paris, 1958) and space over X by A. Dold, "Partitions of unity in the theory of fibrations," Ann. Math. **78**, 223 (1963).

[13]We call a bundle map *invertible* only if its inverse is also a bundle map.

[14]Coordinate transformations are discussed in detail in Ref. 2.

[15]See Ref. 4, Sec. 2.3.

[16]The concept of a G-structure is due to S.S. Chern. See his article "The geometry of G-structures." Bull. Am. Math. Soc. **72**, 167 (1966). See also S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliffs, New Jersey, 1964), Chap. VII.

[17]A. Böhm, "Relativistic rotators—a quantum mechanical de Sitter bundle," in: *Proceedings of the International Symposium on Mathematical Physics, 1976, Mexico City* (University of Mexico, Mexico City, 1977).

[18]See the contributions by W. Drechsler in Ref. 10 and in *Group Theoretical Methods in Physics*, Vol. 94 of *Lecture Notes in Physics*, edited by W. Beiglböck, A. Böhm, and E. Takasugi (Springer-Verlag, Berlin, 1979).

[19]P.A.M. Dirac, J. Math. Phys. **4**, 901 (1963).

[20]See the contribution by H. van Dam and L.C. Biedenharn, in Ref. 18.

[21]L. O'Raifeartaigh, Phys. Rev. B **139**, 1052 (1965).

JMAPA,800400,800433

An axiomatic system for Minkowski space–time

John W. Schutz
Bendigo College of Advanced Education, Bendigo, Victoria 3550, Australia [a]
Max-Planck-Institut für Physik und Astrophysik, Föhringer Ring 6, 8 München 40, Federal Republic of Germany [b]

(Received 13 August 1979; accepted for publication 17 December 1979)

Minkowski space–time is developed in terms of a set of undefined primitive elements called events, certain subsets of events called paths which correspond to the worldlines of free particles, and a temporal order relation on each path. Nine axioms describe the existence and uniqueness of paths, temporal order, connectedness, causality, collinearity, continuity, isotropy, and dimension.

PACS numbers: 02.40. + m, 03.30. + p

INTRODUCTION

Minkowski space–time will be described in terms of a set of undefined primitive elements called "events," certain subsets of events called "paths" which correspond to the worldlines of free particles, and a temporal order relation defined on the set of events of each path. The postulated relations of temporal order may be thought of as applying to freely moving observers who are capable only of distinguishing between events in their own local histories. The axioms permit the extension of these local temporal orderings defined on each path to a global temporal ordering defined on the set of all events.

In the present axiomatic system, all the axioms describe properties of paths. The existence of "light signals" is deduced rather than postulated axiomatically as in the preceding axiomatic systems of Walker,[1,2] Szekeres,[3] and Schutz.[4]

In all these axiomatic systems, the concept of a "coordinate frame" is developed from primitive notions which correspond to the temporal order properties and kinematic behavior of free particles and light signals. Other axiomatic systems, in which the existence of coordinate frames is assumed, have been briefly reviewed by Schutz.[4]

The form of the axiomatic system resembles both that of Hilbert[5] for Euclidean geometry and the previous system of Schutz[4] for Minkowski space–time. The axioms of the present system may be compared with those of Hilbert,[5] and it will be seen that all but one of the axioms have analogs within Hilbert's five axiom groups: the exception is the Axiom of Causality (Axiom V, 1.41) which expresses the essential character of a causal space–time and is therefore not required for a geometry. Whereas Hilbert's axiom system contained an entire group of axioms stating the properties of an undefined congruence relation, the present axiomatic system contains one symmetry axiom (the Axiom of Isotropy, Axiom VIII, 1.71) and does not require the introduction of a congruence axiom at all. Instead, it transpires that all the properties of a congruence relation can be deduced as theorems within the present axiomatic system.

Some comparisons can now be made with the previous axiomatic system of Schutz,[4] which was subtitled *Kinematic Axioms for Minkowski Space–Time* and will subsequently be referred to by the letters KA. The previous rather strong Axiom of Compactness of Bounded sub-SPRAY (KA Axiom XI, 2.13) is now replaced by the considerably weaker Axiom of Continuity (Axiom VII, 1.63) which is analogous to the geometric axiom of the same name.[5] This axiom applies to the set of events of a path rather than to a set of simultaneously coincident paths, so the axiom is not only weaker but might even be regarded as being more intuitively acceptable.

The present Axiom of Isotropy (Axiom VII, 1.71) is also weaker than its predecessor (KA Axiom VII, 2.9), the significant difference being that the present axiom does not assume that signals are mapped onto signals. This symmetry axiom may be interpreted intuitively as meaning that all "directions" are equivalent. Thus, it is this axiom which expresses an idea similar to the "Principle of Relativity" of Einstein,[6] who postulated that there should be no "preferred coordinate frames." The axiom does not assume the existence of coordinate frames, which will be developed from the axioms, nor does it assume that isotropy mappings are automorphisms. What is assumed, for a given isotropy mapping, is that the events on one path (an "observer's path") are invariant, that paths are mapped onto paths, and that the mapping is bijective on one subset of simultaneously coincident paths.

Two axioms are essentially different from those of the preceding axiomatic system (KA). The Axiom of Uniqueness of Paths (Axiom II, 1.12) is analogous to Hilbert's Axiom of Connection (Hilbert,[5] Axiom I, 2) and was previously deduced as a theorem (KA, Theorem 6, 2.9), while the Axiom of Collinearity (Axiom VI, 1.51) is analogous to Hilbert's Axiom of Order (Hilbert,[5] Axiom II, 4) and is introduced to replace the previous Axiom of Uniqueness of Extension of Optical Lines (KA, Axiom V, 2.7) and the Axiom of the Intermediate Path (KA, Axiom VI, 2.8).

Many properties which were taken as axiomatic in the previous system are now deduced as theorems. Thus the property stated in the Signal Axiom (KA Axiom I, 2.2) is now deduced as Theorem 1.72, the Triangle Inequality (KA Axiom IV, 2.4) as Theorem 1.66, the Uniqueness of Extension of Optical Lines (KA Axiom V, 2.7) as described in Sec. 2.D, the Existence of an Intermediate Path (KA Axiom VI,

[a] Permanent address.
[b] Visiting Scientist, January–August 1978.

 0022-2488/81/020293-10$1.00

2.8) as Theorem 1.56, while the roles of the Axioms of Incidence (KA Axiom XI, 2.11) and Connectedness (KA Axiom X, 2.12) are now subsumed in the axiom of Connectedness (Axiom IV, 1.31).

Within the present axiomatic system it is possible to prove several propositions which are assumed by many other writers. Thus, we do not assume the concept of a coordinate frame, we do not assume that the set of events of each path can be ordered by the real numbers, we do not assume that light signals exist, nor do we assume that paths and light signals have constant speed. These propositions all turn out to be provable within the system.

Three properties of Minkowski space–time are of central importance to the subsequent development. One-dimensional kinematics is in many ways analogous to plane absolute geometry, for it transpires that the concept of parallelism can be applied to paths and, furthermore, the corresponding question of uniqueness of parallelism is closely related to uniform motion along paths. Both Robb[7] and Szekeres[3] observed that uniform motion implies uniqueness of parallelism but, in the present axiomatic system as well as in KA, the uniqueness of parallelism is proved and then it is shown that this implies uniform motion along paths, so that Newton's first law of motion need not be assumed. The second important property is that, in contrast to the Euclidean velocity space of Newtonian kinematics, the velocity space associated with Minkowski space–time is hyperbolic, a property which is established by making use of a recent characterization of the elementary spaces by Tits.[8-10] The third important property is that space–time coordinates are related to homogeneous coordinates in a three-dimensional hyperbolic space. Consequently, there is an isomorphism between homogeneous Lorentz transformations and transformations of homogeneous coordinates in hyperbolic space.

In this paper many details of the proofs have been omitted in order to keep the paper to a reasonable length. Full details of the proofs are given in a research report[11] which is available on request from the Max-Planck-Institut für Astrophysik.

1. AXIOMS AND PRIMITIVE NOTIONS

Minkowski space–time will be described in terms of a set of undefined primitive elements called events, certain subsets of events called paths, and a *temporal order relation* $<$ defined on the set of events of each path.

The set of all paths will be denoted by $\mathscr{P}$ and the set of all events will be denoted by $\mathscr{E}$. Individual paths will be denoted by the symbols **Q**, **R**, **S**,···. Events belonging to a path, say **Q**, will be denoted by the path symbol together with a subscript, for example, Q_1, Q_a, Q_x; or by means of lower case letters such as a, b, c,···. If $a \in \mathbf{Q}$, we say that *the event a lies on the path* **Q** or more simply that *a lies on* **Q** or **Q** *passes through a*. An event may belong to two or more paths: if $\mathbf{Q} \cap \mathbf{R} = a$, we say that **Q** and **R** *coincide at a* or that **Q** *meets* **R** *at a*.

The temporal order relation applies only to pairs of events on the same path: The statement "$Q_1 < Q_2$" is to be read as "Q_1 is *before* Q_2" or "Q_2 is *after* Q_1".

Minkowski space–time will be defined to be the ordered triple $\mathscr{M} = (\mathscr{E}, \mathscr{P}, <)$.

A. The axioms of existence and uniqueness for paths

It is first necessary to ensure that $\mathscr{E}$ and $\mathscr{P}$ are nonempty and that there are at least two distinct paths (where the word "distinct" is used in the set-theoretic sense).

Axiom I, 1.11 (existence): There are at least two distinct paths.

Axiom II, 1.12 (uniqueness): There is at most one path passing through any two given distinct events.

Within Hilbert's[5] axiom system for geometry, these axioms have their analogues in the Axioms of Connection (I2 and I3).

The Axiom of Uniqueness is the first of three axioms which express different aspects of the "law of inertia" or "Newton's first law." The Axiom of Uniqueness expresses the idea that paths correspond to the possible trajectories of "free particles." Other aspects of the law of inertia are described by the Axiom of Collinearity (Axiom VI, 1.51) and the Axiom of Isotropy (Axiom VIII, 1.69).

B. The axiom of temporal order

This axiom ensures that there are no closed time-like world lines and is analogous to the axioms of order of geometry (see, for example, Hilbert's[5] Axioms of Order II 1–3).

Axiom III, 1.21 (temporal order): The events of each path are irreflexively and linearly ordered by a temporal order relation $<$.

The axiom also ensures that, in the present axiomatic system, we do not need to assume that the events on a path are order isomorphic with the real numbers, nor do we assume that a path may be prolonged (cf. Hilbert[5], Axiom II 2.): This proposition is proved as a theorem (Theorem 1.43).

Although the temporal order relation $<$ is clearly antisymmetric, there is no "anisotropy of time" within the axiomatic system since we can consistently interchange the symbols $<$ and $>$, the words "before" and "after", and the words "past" and "future" (see Sec. 1.C following). The choice of one direction rather than the other as a "forward direction of time" is simply a matter of convention within this axiomatic system.

C. The axiom of connectedness

This axiom also has an analog in the foundations of geometry [see, for example, Hilbert[5] Axiom (s) of Incidence I 1]. Before stating the axiom we make the following definitions: Given a path **R** and an event $e \notin \mathbf{R}$, we define

(i) the *future of e* in **R**:

$$\mathbf{R}(e, +) = \{g : g > e, g \in \mathbf{R}\};$$

(ii) the *past of e* in **R**:

$$\mathbf{R}(e, -) = \{d : d < e, d \in \mathbf{R}\};$$

(iii) the *unreachable* set of e in **R**:

$$\mathbf{R}(e, \phi) = \{f : f \not\gtrless e, f \in \mathbf{R}\}.$$

The future of e in **R** is the set of events in **R** which are

connected with e by paths directed from e to **R**, while similarly the past of e in **R** is the set of events in **R** which are connected with e by paths directed from **R** to e. Note that the unreachable set is the set of events in **R** which can not be connected with e by any path.

Axiom IV, 1.31 (connectedness): Given a path **R** and an event $e \notin \mathbf{R}$, the subsets $\mathbf{R}(e, -)$, $\mathbf{R}(e, \phi)$, $\mathbf{R}(e, +)$ each contain more that one event.

This axiom expresses the notion of the "boundedness of relative velocities" although it should be noted that the concept of velocity has not yet been developed. While this axiom has some resemblance to Hilbert's[5] first Axiom of Connectedness (Axiom I1), it also expresses an essential difference between a geometry and a space–time. The other essential difference is contained in the Axiom of Causality (Axiom V, 1.41). Note that, in Galilean space–time, the set $\mathbf{R}(e, \phi)$ contains exactly one event: Apart from this exception, all the other axioms are satisfied by Galilean space–time but additional axioms would be required for categoricity.

D. The axiom of causality

This axiom expresses the second essential difference between a causal space–time and a geometry. In effect it states that there is a temporal order relation on space–time permitting the usual interpretation of causality, namely that "a cause always precedes an effect in time." Another interpretation of the axiom is that there are no combinations of "forward-directed path segments" which permit "travel into the past." For a further discussion of causal space–times, see Kronheimer and Penrose[12] and Woodhouse.[13] Since the classical geometries have no causal properties, it is not surprising that the axiom of causality has no analog within Hilbert's axiom system for geometry. In the statement of the axiom, the temporal order relation $<$ has the following meaning for sets of events A, B:

$$[A < B] \Longleftrightarrow [a \in A,\ b \in B \Rightarrow a < b].$$

Axiom V, 1.41 (causality): Given a path **R** and an event $e \notin \mathbf{R}$, then

$$\mathbf{R}(e, -) < \mathbf{R}(e, \phi) < \mathbf{R}(e, +).$$

This axiom has, as a consequence, the following very useful lemma which is logically equivalent to the Axiom of Causality and which could have been used as an axiom in an alternative axiomatic system. The choice of one proposition, rather than the other, for an axiom is a matter of taste only and is arbitrary to that extent.

Lemma 1.42 (direct path): Given two paths **Q**, **R** with events a, $b \in \mathbf{Q}$ and b, $c \in \mathbf{R}$, respectively, such that

$$a < b \text{ and } b < c,$$

then (i) there is a path **S** which passes through the events a and c, and (ii) $a < c$.

Remarks: (i) This lemma also applies with time-reversed orderings. (ii) The lemma extends the transitivity of the temporal order relations from events on the same path to the set of all events $\mathscr{E}$. Thus, the "local" temporal orderings on paths can now be extended to a "global" temporal ordering on the set of all events $\mathscr{E}$. Given any three events a, b, c with $a < b$ and $b < c$ it follows that $a < c$ and we are therefore justified in using the more compact expression $a < b < c$ with the obvious meaning.

Proof: We are given that $b \in \mathbf{R}(a, +)$, that $c \in \mathbf{R}$, and that $c > b$, so by the Axiom of Causality (Axiom V) it follows that $c \in \mathbf{R}(a, +)$; thus, $a < c$ and there is a path **S** containing a and c.

It will now be shown that each path has no "first event" and no "last event."

Lemma 1.43 (prolongation): Given a path **Q** with events b, c such that $b < c$, then there are events a, $d \in \mathbf{Q}$ such that

$$a < b < c < d.$$

Proof: By the Axiom of Existence (Axiom I, 1.11) there is a path **R** distinct from **Q**, so by the Axioms of Uniqueness (Axiom II, 1.12) and Connectedness (Axiom IV, 1.31) there is a path **S** (which may be **R**) distinct from **Q** such that **S** meets **Q** at b.

Again by two applications of the Axiom of Connectedness, there is an event $e \in \mathbf{S}$ such that $e > c$ and an event $d \in \mathbf{Q}$ such that $d > e$, so by the Axiom of Causality (Axiom IV, 1.41), $d > c$. The existence of an event $a < b$ may be proved in a similar manner.

E. The axiom of collinearity

This axiom makes it possible to discuss "rectilinear motion" in terms of "collinear sets of events" so, in effect, the axiom asserts that "motion continues in a straight line" to paraphrase Newton's law of inertia. Whereas *local* versions of Axioms I, II, III, IV, V, and VII hold even in general relativistic space–times, there is no local analog of Axiom VI which applies, except in certain special cases. The nonvalidity of the corresponding statement expresses the presence of projective curvature. I am indebted to Professor J. Ehlers for pointing this out.

The axiom of collinearity is a kinematic analog of the geometric axiom of planarity given by Veblen[14] and, as with Veblen's axiom, it also makes a statement about intermediacy which has the consequence that each path is dense in itself (Lemma 1.52). Together with the Axiom of Isotropy (Axiom VIII, 1.69), this axiom of collinearity enables us to prove the important and subtle Signal Theorem (1.610) which asserts that "light signals exist and have the expected limiting properties of sequences of paths."

To obtain an appreciation of the kinematic ideas expressed by the axiom, the reader is urged to draw a diagram corresponding to the statement of the axiom, by sketching each path segment and then attaching a consistent temporal ordering scheme to the diagram. If a template is made by cutting a narrow slit in a sheet of paper, the "motions" of the paths may be observed by moving the template gradually across the diagram.

Before stating the axiom we make the following definitions: If a, b, c are three distinct events on one path such that either $a < b < c$ or $c < b < a$, we say that b is *between* a and c and we denote this by writing $[abc]$. If a, b, c are three distinct events, not all on the same path, such that each pair of events lies on some path, we say that abc is a *kinematic triangle*.

Axiom VI, 1.51 (collinearity): If *abc* is a kinematic triangle and *d*, *e* are events with [*bcd*] and [*cea*] and if there is a path *de*, then the path *de* meets the path *ab* at an event *f* such that [*afb*].

In Hilbert's[5] system of axioms for geometry, the analogous axiom is Axiom II 4. The ensuing discussion of collinear sets of events has much in common with the description of planes given by Veblen[14] although in the present case, the development of the kinematic theory is more complicated due to the existence of pairs of events which can not be connected by paths. Remember that a path corresponds to the "world line of a free particle."

Lemma 1.52 (each path is dense in itself): Given any path **Q** with distinct events $a, b \in \mathbf{Q}$ there is an event $f \in \mathbf{Q}$ such that [*afb*].

Proof: Full details of all proofs are included in Ref. 11.

Theorem 1.53: In the notation of the Axiom of Collinearity

[*def*],

i.e., if *abc* is a kinematic triangle and [*bcd*] and [*cea*] such that $d \lesssim e$, then on the path *de* there is an event *f* for which [*def*] (see Fig. 1).

Lemma 1.54

If *abc* is a kinematic triangle and [*afb*] and [*bcd*] with[15] $d \lesssim a$ and $d \lesssim f$, then on the path *df* there is an event *e* for which [*cea*].

Theorem 1.55 (ordered coincidence): Let **Q**, **R** be distinct paths which meet at *x*. Let $a, e \in \mathbf{Q}$ and let $b, d \in \mathbf{R}$. (i) If $x < a < b < d < e$, then the paths *ab* and *de* meet at an event *c* such that [*abc*] and [*cde*]. (ii) If $b < a < x < d < e$, then the paths *ab* and *de* meet at an event *c* such that [*bac*] and [*cde*].

Given three distinct paths **Q**, **R**, **S** which meet at an event *x*, we say that the event $b \in \mathbf{R}$ after *x* is between the paths **Q** and **S** if:

(i) for each event $c \in \mathbf{S}$ with $c > b$, the path *cb* meets **Q** at some event *a*, and

(i') for each event $e \in \mathbf{Q}$ with $e > b$, the path *eb* meets **S** at some event *d*.

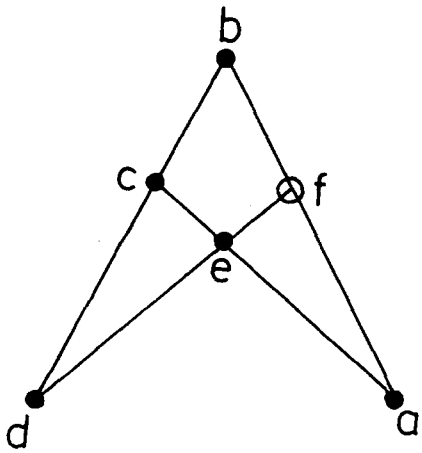


FIG. 1. An event *f* for which [*def*].

Note that (i') is similar to (i) but with the roles of **Q** and **S** reversed, so the definition is symmetric with respect to **Q** and **S**. An analogous definition applies for events before *x*, with reversed temporal orderings.

To extend the definition of "betweenness" to paths, we impose the further conditions (ii) and (ii') below. We say that the path **R** is between the paths **Q** and **S** and denote this by writing $\langle \mathbf{Q}, \mathbf{R}, \mathbf{S} \rangle$ if

(i) all events on **R** are between **Q** and **S**,

(ii) for any path **T** which meets **Q** and **S** at events after *x*, the path **T** meets **R** at some event, and

(ii') for any path **T** which meets **Q** and **S** at events before *x*, the path **T** meets **R** at some event.

Theorem 1.56 (intermediate path): Let **Q** and **S** be two distinct paths which meet at some event *x*, and let **T** be a path which meets **Q** and **S** at events *a* and *c*, respectively, after *x*. For each event $b \in \mathbf{T}$ such that [*abc*], (i) there is a path **R** which passes through *x* and *b*, and, furthermore, (ii) $\langle \mathbf{Q}, \mathbf{R}, \mathbf{S} \rangle$.

Corollary 1.57 (coincidence): Let **Q**, **R**, **S** be distinct paths which meet at *x* such that $\langle \mathbf{Q}, \mathbf{R}, \mathbf{S} \rangle$. Let **T** meet **Q** at *a* and **R** at *b*. If $a > b$, then **T** meets **S** at an event *c* such that $x < c < b < a$.

In the subsequent development we will frequently be discussing the properties of sets of paths which meet at a given event. As in KA we will call any one of these sets a *SPRAY* of paths, or more simply a *SPRAY*, where the upper case letters indicate that we are referring to a set of paths rather than a set of events. We define[16]

$$\mathrm{SPR}[Q_c] := \{\mathbf{R} : Q_c \in \mathbf{R},\ \mathbf{R} \in \mathscr{P}\}.$$

The corresponding set of events is called a *spray*, where the lower case letters indicate a set of events. We define

$$\mathrm{spr}[Q_c] := \{R_x : R_x \in \mathbf{R}, \mathbf{R} \in \mathrm{SPR}[Q_c]\}.$$

If **Q**, **S** are distinct paths which meet at an event *x*, we define the *collinear sub-spray*

$$\mathrm{CSP}\langle \mathbf{Q}, \mathbf{S} \rangle := \{\mathbf{R} : \langle \mathbf{R}, \mathbf{Q}, \mathbf{S} \rangle, \langle \mathbf{Q}, \mathbf{R}, \mathbf{S} \rangle \text{ or } \langle \mathbf{Q}, \mathbf{S}, \mathbf{R} \rangle;\ \mathbf{R} \in \mathrm{SPR}[x]\}.$$

Given four paths **Q**, **S**, **U**, **V** which meet at an event *x*, we write $\langle \mathbf{Q}, \mathbf{S}, \mathbf{U}, \mathbf{V} \rangle$ if $\langle \mathbf{Q}, \mathbf{S}, \mathbf{U} \rangle$, $\langle \mathbf{Q}, \mathbf{S}, \mathbf{V} \rangle$, $\langle \mathbf{Q}, \mathbf{U}, \mathbf{V} \rangle$, and $\langle \mathbf{S}, \mathbf{U}, \mathbf{V} \rangle$. The notation $\langle .,.,., \rangle$ may be extended in the obvious way to any set of paths provided that each ordered triple satisfies the definition of betweenness.

Theorem 1.58: Let **Q**, **S** be distinct paths which meet at *x*. Then $\mathrm{CSP}\langle \mathbf{Q}, \mathbf{S} \rangle$ is a simply ordered set.

Lemma 1.59 (generation): Let **Q**, **R** be distinct paths which meet at *x* and let **T** be a path which meets **Q** and **R** at events other than *x*. Then the sub-SPRAY of paths which join *x* to events of **T** is a subset of $\mathrm{CSP}\langle \mathbf{Q}, \mathbf{R} \rangle$.

Remark: It may appear that the set of paths generated by *x* and **T** is identical with $\mathrm{CSP}\langle \mathbf{Q}, \mathbf{R} \rangle$. It will turn out, as a consequence of the Uniqueness of Parallelism (KA, Theorem 46, 7.5) that this is not the case, for there is one path in $\mathrm{CSP}\langle \mathbf{Q}, \mathbf{R} \rangle$ through *x* "parallel to" **T** and this path is clearly not generated by *x* and **T**.

Corresponding to CSP $\langle \mathbf{Q}, \mathbf{R} \rangle$ we denote the *set of events belonging to the paths* of CSP $\langle \mathbf{Q}, \mathbf{R} \rangle$ by

$$\mathrm{csp}\langle \mathbf{Q}, \mathbf{R} \rangle := \{e : e \in \mathbf{S},\ \mathbf{S} \in \mathrm{CSP}\ \langle \mathbf{Q}, \mathbf{R} \rangle\},$$

and we denote the *set of events after the event of coincidence* by

$$\mathrm{csp}^+\langle \mathbf{Q}, \mathbf{R}\rangle := \{e{:}e > \mathbf{Q}\cap\mathbf{R},\ e\in\mathrm{csp}\langle \mathbf{Q}, \mathbf{R}\rangle\}.$$

Lemma 1.510 (containment): Let **Q**, **R**, **S** be distinct paths which meet at distinct events b, c, d where $b = \mathbf{S}\cap\mathbf{Q}$, $c = \mathbf{Q}\cap\mathbf{R}$, $d = \mathbf{R}\cap\mathbf{S}$ such that $b < c < d$. Then

$$\mathrm{csp}^+\langle \mathbf{R}, \mathbf{S}\rangle \subseteq \mathrm{csp}^+\langle \mathbf{Q}, \mathbf{R}\rangle \subseteq \mathrm{csp}^+\langle \mathbf{S}, \mathbf{R}\rangle.$$

We define a collinear set of events in a manner which is analogous to the definition of a plane set of points in absolute geometry. Let **R**, **S** be any two distinct paths which meet at some event. Then *the set of events collinear with* **R** and **S** is

$$\mathrm{col}[\mathbf{R}, \mathbf{S}] := \{W_x{:}W_x\in\mathbf{W} \text{ where } \mathbf{W} \text{ meets } \mathbf{R} \text{ and } \mathbf{S} \text{ at two distinct events}\}.$$

The next theorem states that collinear sets of events have properties which are analogous to the properties of coplanar sets of points in absolute geometry. Anticipating the result of part (i) of the theorem, we will define a *collinear set of paths*

$$\mathrm{COL}\ [\mathbf{R}, \mathbf{S}] := \{\mathbf{U}{:}\mathbf{U} \text{ meets col } [\mathbf{R}, \mathbf{S}] \text{ in two distinct events}\}.$$

Theorem 1.511 (collinear sets): (i) *Containment*: any path which coincides with two distinct events of a collinear set is contained in the collinear set. (ii) *Uniqueness*: a path and an event not on the path specify a unique collinear set.

Given two distinct paths **Q**, **S** which meet at an event x, the set of paths CSP $\langle \mathbf{Q}, \mathbf{S}\rangle$ is linearly ordered and, accordingly, we could assign a sense of direction[17] to them. The Intermediate Path Theorem (1.56) and its Corollary (1.57) show that, for any path $\mathbf{T}\in\mathrm{COL}\ [\mathbf{Q}, \mathbf{S}] - \mathrm{CSP}\langle \mathbf{Q}, \mathbf{S}\rangle$, the restriction

$$\mathbf{T}^+ = \{T_y{:}T_y > x,\ T_y\in\mathbf{T}\},$$

has the property that, for each $T_y\in\mathbf{T}^+$, there is some path $\mathbf{R}\in\mathrm{CSP}\langle \mathbf{Q}, \mathbf{S}\rangle$ such that $T_y\in\mathbf{R}$, and the events of $\mathbf{T}^+$ before T_y are on one side of **R** while the events of T^+ after T_y are on the other side of **R**: In this sense, $\mathbf{T}^+$ *crosses* **R** in $\mathrm{csp}^+\langle \mathbf{Q}, \mathbf{S}\rangle$.

Corollary 1.512 (crossing): A collinear set Σ can be given a sense of direction (left to right) which has the following properties: (i) Each path **S** in Σ separates the events of $\Sigma - \mathbf{S}$ into two disjoint sets—a *left side of* **S** and a *right side of* **S**; (ii) A path **T** which contains one event T_u from the left side of **S** and one event T_w from the right side of **S** meets **S** at an event v between T_u and T_w; thus, **T** *crosses* **S** at the event v; (iii) The sense of direction is consistent in the sense that:

(α) if (all events of) a path **W** is on the right side of a path **S**, then the right side of **W** is contained in the right side of **S** (and the left side of **S** is contained in the left side of **W**),

(β) if a path **T** is on the right side of a path **S** after (before) an event v, then the right side of **T** after (before) v is contained in the right side of **S** after (before) v; (iv) If two paths meet, then they cross each other.

F. Signal functions and the axiom of continuity

The Axiom of Continuity is analogous to the geometrical axiom of the same name as used by Hilbert[5] (Axiom V2) and ensures the "existence of light signals." Before stating this axiom it is useful to establish a lemma which may be scribed kinematically by the statement "given any path there is a faster path."

Lemma 1.61 (there is no fastest path): Given a path **Q** and an event $e\notin\mathbf{Q}$, (i) the set $\mathbf{Q}(e, +)$ has no first event, and (ii) the set $\mathbf{Q}(e, -)$ has no last event.

The Axiom of Continuity which follows asserts that each path has a completeness property which ensures that the concept of "fastest signal" can be defined. Before stating the axiom we need the following definition: A linearly ordered set is *conditionally complete* if each bounded subset has an infimum and a supremum.

Axiom VII, 1.62 (continuity): Each path is conditionally complete.

It is now possible to define two signal functions (see Fig. 2); the *forward signal function* is defined

$$f^+_{RQ}{:}\mathbf{Q}\to\mathbf{R},$$

$$Q_x \mapsto \begin{cases} \inf \mathbf{R}(Q_x, +), & \text{if } \mathbf{Q} \text{ and } \mathbf{R} \text{ do not meet at } Q_x \\ Q_x, & \text{if } \mathbf{Q} \text{ and } \mathbf{R} \text{ meet at } Q_x, \end{cases}$$

and the *reverse signal function*

$$f^-_{RQ}{:}\mathbf{Q}\to\mathbf{R},$$

$$Q_x \mapsto \begin{cases} \sup \mathbf{R}(Q_x, -), & \text{if } \mathbf{Q} \text{ and } \mathbf{R} \text{ do not meet at } Q \\ Q_x, & \text{if } \mathbf{Q} \text{ and } \mathbf{R} \text{ meet at } Q_x, \end{cases}$$

These signal functions correspond physically to light signals (see Fig. 2) and, as might be expected, it transpires (in the Signal Theorem 1.610) that pairs of functions, such as f^+_{RQ} and f^-_{QR}, are inverses of each other. However, this is not immediately apparent within the present axiom system since there is no axiom corresponding to the property which Kronheimer and Penrose[12] and Woodhouse[13] have called

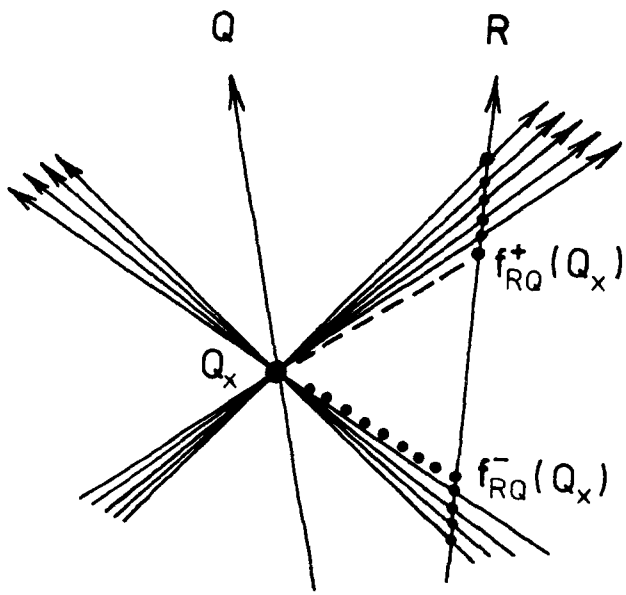


FIG. 2. Signal functions are defined by considering limiting properties of paths. Forward signal functions are illustrated by broken lines (----) and reverse signal functions are illustrated by dotted lines (·····).

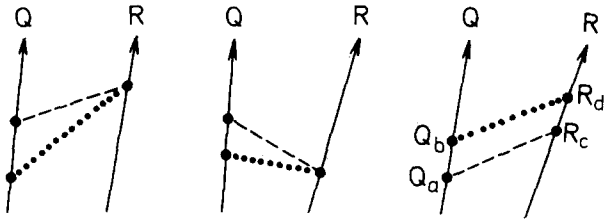


FIG. 3. The signal function inequalities.

"future and past distinguishing."

An immediate and useful consequence of the Axiom of Continuity is the following:

Lemma (monotonic sequence property): Any bounded monotone sequence of events on a path has a limit on the path (with respect to the order topology).

Lemma 1.64 (signal functions are weakly order preserving): Given two paths **Q**, **R** with events $Q_x, Q_z \in \mathbf{Q}$, then

(i) $Q_x < Q_z \Rightarrow f^{+}_{RQ}(Q_x) \leqslant f^{+}_{RQ}(Q_z)$,

(ii) $Q_x < Q_z \Rightarrow f^{-}_{RQ}(Q_x) \leqslant f^{-}_{RQ}(Q_z)$.

Lemma 1.65 (signal function inequalities): Given any two paths **Q** and **R**, then

(i) $f^{-}_{QR} \circ f^{+}_{RQ} \leqslant i_{QQ}$,

(ii) $f^{+}_{QR} \circ f^{-}_{RQ} \geqslant i_{QQ}$,

where i_{QQ} is the identity function on **Q**. Let $R_c := f^{+}_{RQ}(Q_a)$ and let $Q_b := f^{-}_{QR}(R_d)$. Then

(iii) $Q_a < Q_b \Longleftrightarrow R_c < R_d$

(see Fig. 3).

Theorem 1.66 (triangle inequalities): Let **Q**, **R**, **S** be three paths. Then

(i) $f^{+}_{SR} \circ f^{+}_{RQ} \geqslant f^{+}_{SQ}$,

(ii) $f^{-}_{SR} \circ f^{-}_{RQ} \leqslant f^{-}_{SQ}$.

Remark: In the previous axiomatic system (KA), similar properties were taken as axioms.

Given two events a, b and paths **Q**, **R** such that

$$f^{+}_{RQ}(a) = b,$$

we say that a and b are (*forward*) *signal related* and write $a \sigma b$. The binary relation σ is called the *signal relation* and an expression of the form $a \sigma b$ may be read as "a signal goes from a to b "or" a signal leaves a and arrives at b." Given three events a, b, c such that

$$a \sigma b \text{ and } b \sigma c \text{ and } a \sigma c,$$

we say that the ordered set of events $|a, b, c\rangle$ *is in optical line.*[18] A similar definition applies to events which are (reverse) signal related.

Lemma 1.67 (optical line): If $\langle \mathbf{Q}, \mathbf{R}, \mathbf{S} \rangle$ then, for all $Q_u \in \mathbf{Q}$,

(i) the events Q_u, $f^{+}_{RQ}(Q_u)$, $f^{+}_{SQ}(Q_u)$ are in optical line;

(ii) the events Q_u, $f^{-}_{RQ}(Q_u)$, $f^{-}_{SR}(Q_u)$ are in optical line.

Theorem 1.68 (existence): Let **Q**, **U** be distinct paths which meet at some event x. Let Q_a, $Q_b \in \mathbf{Q}$ be distinct events such that

$$x < Q_a < Q_b \quad \text{and} \quad f^{+}_{UQ}(Q_a) > f^{-}_{UQ}(Q_b).$$

Then there exists a path **S**, between **Q** and **U** and distinct from both, such that

$$f^{+}_{SQ}(Q_a) = f^{-}_{SQ}(Q_b)$$

(see Fig. 4).

Remark: At this stage we can not assert that **S** is unique. The uniqueness of **S** can, however, be asserted after the Signal Theorem (1.72).

G. The axiom of isotropy

The Principle of Relativity of Einstein[6] states, in effect, that there is no preferred reference frame. In the present axiomatic system, the Axiom of Isotropy asserts that "all directions are equivalent" and so it is this axiom which performs a similar role to Einstein's Principle of Relativity.

Axiomatic systems for the classical geometries usually have a set of "congruence" axioms (e.g., Hilbert[5]) or several symmetry axioms (e.g., Redei[19]). In the present axiomatic system for Minkowski space–time we do not need any congruence axioms, and the Axiom of Isotropy is the only symmetry axiom required for categoricity. By means of "comparisons" using light signals and isotropy mappings, it is possible to compare intervals of distance and time, and so the concept of congruence can be developed without any further axioms.

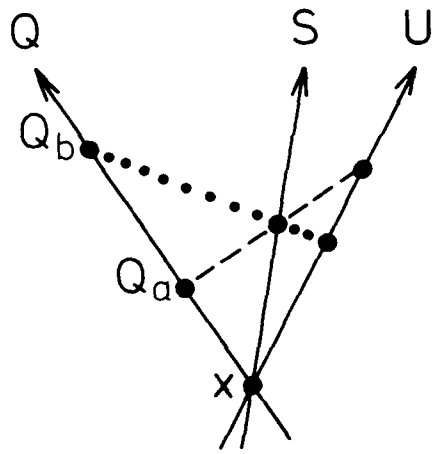


FIG. 4. The existence theorem.

Although the question of "relative simplicity" of axiomatic structures can not be defined or answered in any precise way and is ultimately a matter of aesthetics or fashion, it appears to the author that the additional structure of Minkowski space–time, compared with the classical geometries, enables it to be categorized with fewer and simpler axioms.

Axiom VIII, 1.71 (isotropy): Let **Q**, **R**, **S** be distinct paths meeting at x. If, for some event $Q_y \in \mathbf{Q}$ with $Q_y \neq x$,

$$f^+_{QR} \circ f^+_{RQ}(Q_y) = f^+_{QS} \circ f^+_{SQ}(Q_y),$$

then there exists a map θ of $\mathscr{E}$ into itself such that

(i) θ maps $\mathscr{P}$ into itself,

(ii) for $Q_z \in \mathbf{Q}$, $\theta(Q_z) = Q_z$,

(iii) θ induces a bijection on the set SPR[x] of paths,

(iv) $\theta(\mathbf{R}) = \mathbf{S}$.

It may seem surprising that this axiom is sufficient for our purpose.[20] It transpires that θ is an order-preserving bijective mapping which sends signals onto signals as one would expect (and require), and this is demonstrated in Theorem 1.610 for the restriction to "one-dimensional motion" and in Theorem 2.31 for the general case.

An immediate consequence of this axiom and the Collinear Sets Theorem (1.511) is that θ sends collinear sets onto collinear sets. Furthermore if **Q**, **R**, **S** belong to a collinear set Σ, then θ induces a mapping of Σ onto itself and the induced mapping has the same properties as stated in the axiom.

Theorem 1.72 (signal): (i) Signal functions have inverses and, for any paths **Q** and **R**,

$$f^-_{RQ} = (f^+_{QR})^{-1} \text{ and } f^+_{RQ} = (f^-_{QR})^{-1}.$$

(ii) Signal functions are one to one onto mappings. (iii) Signal functions are strictly order preserving. (iv) Signal functions are continuous (with respect to the order topology on each path).

Remarks: Thus, both forward and reverse signal functions may be expressed in terms of the (forward) signal function

$$f_{RQ} := f^+_{RQ},$$

which corresponds physically to a light signal: The reverse signal function is related to its inverse, although the order of the path subscripts must be reversed; thus,

$$f^-_{QR} = f^{-1}_{RQ}.$$

In Sec. 2, superscript $+$ and $-$ signs will be used to indicate "modified signal functions" as in Walker[1,2] and Schutz.[4] After the proof of this theorem and throughout the rest of this chapter, we will dispense with the symbols f^+ and f^- entirely.

If three events, a, b, c are signal related as follows:

$a \sigma b$ and $b \sigma c$ and $a \sigma c$,

we say, as before, that the events are in optical line and denote this by $|a, b, c\rangle$ where the notation indicates the direction of the (forward) signal relation. Similarly, a set of events $\{Q^{(1)}_1, Q^{(2)}_2, \cdots, Q^{(n)}_n\}$ is *in optical line* if and only if, for all a, b, c with $1 \leqslant a \leqslant b \leqslant c \leqslant n$,

$$|Q^{(a)}_a, Q^{(b)}_b, Q^{(c)}_c\rangle.$$

Physically, a set of events would be in optical line if they lay on a "light ray." The existence of such maximal sets of signal related events is proved much later within the present axiomatic system (see Sec. 2.D) since, at this stage, we can not yet assert that $|a, b, c\rangle$ implies the collinearity of the events a, b, c. Thus, it would be premature to speak of a "light ray." However, useful results can be established if we restrict our attention to collinear sets. Anticipating the results of the following theorem we will define an *optical line* to be a maximal subset of a collinear set such that any three events are in optical line.

Theorem 1.73 (optical line): Let Σ be a collinear set of paths and events belonging to the paths, let **Q** be any path in Σ, and let Q_y be any event of **Q**. (i) There are two distinct optical lines, each containing Q_y and exactly one event from each path in Σ (other than **Q**). The optical line which contains events in the order from left to right is called a *right-directed optical line* or simply a *right optical line*. The other optical line is called a *left optical line*. (ii) Each optical line is simply ordered, has no first or last event, and is dense in itself.

Given a collinear set Σ and a path $\mathbf{Q} \in \Sigma$, we see that the events e_1 and e_2 are *reflections* of each other in **Q** if they are on opposite sides of **Q** in Σ and if there are two events Q_a, $Q_b \in \mathbf{Q}$ which are signal related to both e_1 and e_2, i.e., if $Q_a \sigma e_1 \sigma Q_b$ and $Q_a \sigma e_2 \sigma Q_b$. We say that two paths **S**, **T** are *reflections* of each other in **Q** if their sets of events are reflections in **Q** and we indicate this by writing $\mathbf{S} = \mathbf{T}_Q$ or $\mathbf{T} = \mathbf{S}_Q$.

Lemma 1.74 (reflection mapping): Given any collinear set Σ and any path **Q** in the collinear set, there is a *reflection mapping* ϕ of the collinear set onto itself such that

(i) $\mathbf{T} \in \Sigma \Rightarrow \phi(\mathbf{T}) \in \Sigma$,

(ii)for any paths **T**, **U** $\in \Sigma$ and any event $T_x \in \mathbf{T}$,

$$f_{TU} \circ f_{UT}(T_x) = T_z \Rightarrow f_{\phi(T)\phi(U)} \circ f_{\phi(U)\phi(T)}(\phi(T_x)) = \phi(T_z),$$
$$f_{UT}(T_x) = U_y \Rightarrow f_{\phi(U)\phi(T)}(\phi(T_x)) = \phi(U_y).$$

H. Properties of collinear subSPRAYs

Given a path **Q** and an event $Q_c \in \mathbf{Q}$ and a sub-SPRAY $\mathscr{B}[Q_c] \subset \text{SPR}[Q_c]$, we say that $\mathscr{B}[Q_c]$ is a *bounded sub-SPRAY* if there are events Q_d, Q_e after Q_c such that, for all paths $\mathbf{R} \in \mathscr{B}[Q_c]$,

$$f_{QR} \circ f_{RQ}(Q_d) < Q_e.$$

The next theorem establishes an important completeness property of bounded collinear sub-SPRAYs.

Theorem 1.81 (collinearity of the limit path): Let Σ be a collinear subset, let $\mathbf{Q} \in \Sigma$ be a path with an event $Q_c \in \mathbf{Q}$, and let $\{\mathbf{R}^{(n)}: n = 1, 2, \cdots; \mathbf{R}^{(n)} \in \Sigma\}$ be a bounded set of paths which meet at Q_c. If $\langle \mathbf{Q}, \mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(n)}, \mathbf{R}^{(n+1)}, \cdots \rangle$, there is a unique path **S** through Q_c such that

(i) $\langle \mathbf{Q}, \ldots, \mathbf{R}^{(n)}, \mathbf{R}^{(n+1)}, \ldots \mathbf{S} \rangle$ and

(ii) for any event $Q_x \in \mathbf{Q}$ with $Q_x > Q_c$,

$$f_{QS} \circ f_{SQ}(Q_x) = \sup\{f_{QR^{(n)}} \circ f_{R^{(n)}Q}(Q_x)\},$$
$$f^{-1}_{SQ} \circ f^{-1}_{QS}(Q_x) = \inf\{f^{-1}_{R^{(n)}Q} \circ f^{-1}_{QR^{(n)}}(Q_x)\}.$$

We call **S** the *limit path* of the sequence of paths $(\mathbf{R}^{(n)})$.

I. The kinematics of collinear sets of paths

It has now been shown that collinear sets of paths have properties corresponding to those which were taken as axioms in the previous axiomatic system (KA), with two exceptions. The first exception is that the property of collinearity of the limit path (Theorem 1.81) has been established from the Axiom of Continuity (Axiom VII, 1.62) rather than from an Axiom of Compactness (Axiom XI, 2.13, KA). The second exception is that an axiom of dimension has not yet been stated.

Neither of these exceptions affects the subsequent development which leads up to the kinematics of one-dimensional motion. All the details may be found in KA Chaps. 4–8. Those kinematic relations which will be needed to fully develop Minkowski space–time will now be briefly reviewed for the sake of completeness of the present exposition.

Given a collinear set Σ, each path belongs to an equivalence class of paths which are described as being *parallel* since they never meet (KA, 7.1). Whereas a "parallel postulate" is required to distinguish between the geometries of Euclid and Bolyai–Lobachevsky, the uniqueness of parallelism may be deduced as a theorem (KA, Theorem 46, 7.5) for the present kinematic axiom system.

The description of one-dimensional motion can be simplified with the use of "modified signal functions" which are defined in the following way (KA, 43): Given any two paths $\mathbf{Q}, \mathbf{R} \in \Sigma$, the *modified signal function* f^{+}_{RQ} which is related to right optical lines is defined

$$f^{+}_{RQ}(Q_x) = \begin{cases} f_{RQ}(Q_x), & \text{if } f_{RQ}(Q_x) \text{ is to the right of } Q_x, \\ Q_x, & \text{if } \mathbf{R} \text{ meets } \mathbf{Q} \text{ at } Q_x, \\ f^{-1}_{QR}(Q_x), & \text{if } f^{-1}_{QR}(Q_x) \text{ is to the left of } Q_x, \end{cases}$$

and the *modified signal function* f^{-}_{RQ}, which is related to left optical lines, is defined in a similar manner. In the remaining part of this exposition, the symbols f^{+} and f^{-} will designate modified signal functions and not the forward and reverse signal functions of Secs. 1.A–1.G.

The events on any path can be indexed with the real numbers, to within an arbitrary strictly increasing linear transformation (KA, 7.3–7.5), in such a way that, for any two (collinear) paths $\mathbf{Q}$ and $\mathbf{T}$, the modified signal functions have the form (KA, Theorem 49, 7.5 and Theorem 51, 8.1)

$$f^{+}_{QS}(S_t) = Q_{\alpha_{QS}t+\gamma} \quad \text{and} \quad f^{-}_{QS}(S_t) = Q_{\beta_{QS}t+\delta},$$

where $\alpha_{QS}, \beta_{QS}, \gamma, \delta$ are real-valued constants.

The *directed rapidity* of $\mathbf{Q}$ with respect to $\mathbf{S}$ is defined to be

$$r_{QS} = 1/2\log(\alpha_{QS}\beta_{QS}),$$

(KA, 8.1) and for any three collinear paths $\mathbf{Q}, \mathbf{S}, \mathbf{T}$ it is shown (KA, Theorem 51, 81.) that

$$r_{QT} = r_{QS} + r_{ST},$$

so that rapidity may be regarded as a "natural measure for speed."

By considering signal functions between parallel paths, it is possible to represent an equivalence class of parallel paths as

$$\{\mathbf{S}^x : x \text{ real}, \mathbf{S}^x \in \Sigma\},$$

with events

$$\{S^x_t : t \text{ real}, S^x_t \in \Sigma\},$$

where the coefficients of the ordered pair $(x; t)$ of reals are called *position–time* coordinates of the event S^x_t (KA, 8.4). The set of all events in Σ, indexed by the corresponding position–time coordinates, is called a *coordinate frame* in Σ; the event $(0; 0)$ is called the *origin in position–time* of the coordinate frame; and the set of events $\{(0; t)\ t \text{ real}\}$ is called the *origin in position* of the coordinate frame.

Finally, any path through the origin in position–time has coordinates which satisfy the equation

$$x/t = \tanh r = v,$$

where r is the directed rapidity of the path and v is the *velocity* of the path with respect to the coordinate frame (KA, Theorem 56, 8.4).

The reader is referred to KA Chaps. 4–8 for complete details of the definitions and theorems which have been briefly reviewed in this section.

2. THREE-DIMENSIONAL KINEMATICS

A. The axiom of dimension

The axiom of dimension, which is required to specify the dimension of space–time could have been stated in several places within this axiomatic system and could even have been combined with the axiom of existence. However, this would have necessitated slight modifications[21] to preceding axioms and theorems, and it therefore seemed more natural to regard the properties of existence and dimension as being independent.

Before stating the axiom it is necessary to define a concept which is closely related to the concept of dimension. If three distinct paths $\mathbf{Q}, \mathbf{R}, \mathbf{S}$ of a SPRAY can be joined by a path $\mathbf{W}$ which is not in the SPRAY, we write $\overline{\mathbf{Q}, \mathbf{R}, \mathbf{S}}$ and say that the three paths are *dependent* (Fig. 5).

We say that $\mathbf{S}$ is *dependent* on $\mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}$ if $\overline{\mathbf{Q}^{(1)}\ \mathbf{Q}^{(2)}\ \mathbf{S}}$; otherwise, $\{\mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \mathbf{S}\}$ is an *independent set* of paths. Similarly, T is *dependent* on $\mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \mathbf{Q}^{(3)}$ if there are paths $\mathbf{S}^{(1)}, \mathbf{S}^{(2)}$ each dependent on two of the three paths $\mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \mathbf{Q}^{(3)}$

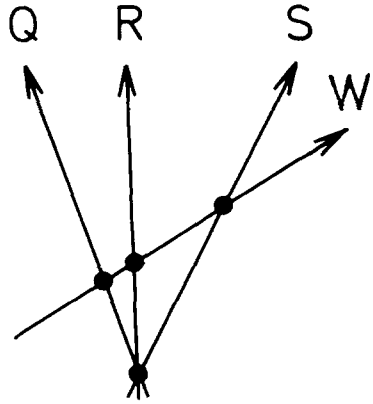


FIG. 5. The paths $\mathbf{Q}, \mathbf{R}, \mathbf{S}$ are dependent.

such that $\overline{S^{(1)}S^{(2)}T}$; otherwise, $\{Q^{(1)}, Q^{(2)}, Q^{(3)}, T\}$ is an *independent* set. Finally, U is *dependent* on $Q^{(1)}, Q^{(2)}, Q^{(3)}, Q^{(4)}$ if there are paths $T^{(1)}, T^{(2)}$ each dependent on three of the four paths $Q^{(1)}, Q^{(2)}, Q^{(3)}, Q^{(4)}$ such that $\overline{T^{(1)}T^{(2)}U}$.

The following definition enables us to specify the dimension of space–time. A SPRAY is a 3-SPRAY if (i) it contains four independent paths $Q^{(1)}, Q^{(2)}, Q^{(3)}, Q^{(4)}$ and (ii) all paths of the SPRAY are dependent on $Q^{(1)}, Q^{(2)}, Q^{(3)}, Q^{(4)}$.

Axiom IX, 2.11 (dimension): If the set of SPRAYs is nonempty, then there is at least one 3-SPRAY.

B. Geometric properties of simultaneously coincident subsets

Lemma 2.21: Each 3-SPRAY is a three-dimensional ordered geometry.

Proof: The proof consists in showing that the axioms for an ordered geometry, as given by Veblen,[14,22] are satisfied.

Lemma: 2.31: Each 3-SPRAY is a locally compact projective metric space.

Lemma: 2.32: (i) Isotropy mappings are order-preserving bijections on $\mathscr{M}$ and map signals onto signals. (ii) Relative rapidity is an invariant under isotropy mappings.

Theorem 2.33: Each 3-SPRAY is a hyperbolic space of three dimensions with curvature of -1: the "points" of the space are the paths of the 3-SPRAY and relative rapidity is an intrinsic metric.

C. Characterization of Minkowski space–time

The usual characterization of Minkowski space–time in terms of coordinate frames and the inhomogenous Lorentz transformations between them, as well as the kinematic description of the trajectories of paths, now follows according to the treatment given in KA[23] (9.1–9.7, Appendix 2).

The existence of optical lines may be established by considering any two given signal-related events and the set of all events which are signal related to both of them. Each of these events must satisfy two equations similar to Eq. (2) of Theorem 61 (KA, 9.5). For any given value of coordinate time these equations describe sets of events lying on nonconcentric spheres (in position space) with one point of tangency, so the set of all signal-related events is a linearly ordered set as described in Theorem 65 (KA, 9.7).

Minkowski space–time is now fully characterized in terms of the conventional coordinate frames and the Lorentz transformations between them, together with the trajectories of free particles and light signals which correspond, respectively, to paths and optical lines.

3. CONCLUSION

Minkowski space–time has been developed from nine axioms which describe the kinematic properties of free particles. Thus, it has been shown that the axiomatic system is categoric. The question of independence of the axioms has not been discussed.

Many of the proofs[11] involve methods which are essentially global in character : for example, the theory of collinearity developed in Sec. 1.E and the theory of parallelism which leads to the kinematics of one-dimensional motion in KA.[11] It is conceivable that global axioms, similar to those used here, might be sufficient to categorise other space–times with special symmetry properties, such as the de Sitter universe and the Robertson-Walker metrics, for example.

Einstein's theory of General Relativity describes space–times which need not have even local symmetries. Thus, some of the methods used here would require considerable modification before they could even be considered for an axiomatic system to describe general relativity. It appears that the pseudo-Riemannian space–times have somewhat independent substructures, namely, the causal, differential–topological, conformal, projective, and metric structures. Causal and topological properties have been discussed by Kronheimer and Penrose[12] while Castagnino,[24] Kundt and Hoffman,[25] Marzke and Wheeler,[26] Pirani,[27] and Synge[28] have considered questions relating to time measurement and the definition of the metric tensor. Axiomatic systems for general relativity have been proposed by Reichenbach[29] and Weyl,[30] who showed how pseudo-Riemannian space–times could be developed from differentiable manifolds by considering the paths of freely falling particles and light signals. This approach has been futher developed by Castagnino,[24] and by Ehlers, Pirani, and Schild[31] and Woodhouse,[13] who show how the different mathematical structures can be successively developed on a space–time manifold by considering axioms which describe the kinematic behavior of free particles and light signals.

ACKNOWLEDGMENTS

I wish to express my gratitude to Professor J. Ehlers for his hospitality at the Max-Planck-Institut für Astrophysik, for many stimulating discussions, and for useful criticisms of the manuscript. I would like to thank the members of the Institut for discussions, especially Dr. R.A. Breuer, who made several suggestions about the presentation of the axioms. Thanks are also due to Dr. S.T. Tsou for critical comments. I express my appreciation to Dr. C.B.G. McIntosh and Professor G. Szekeres for continued interest, useful suggestions, and encouragement. This work was commenced at theBendigo College of Advanced Education and completed at the Max-Planck-Institut für Astrophysik.

[1] A.G. Walker, Proc. R. Soc. Edinburgh Sect. A **62**, 319 (1948).

[2] A.G. Walker, "Axioms for Cosmology," in *The Axiomatic Method*, edited by L. Henkin, P. Suppes, and A. Tarski (North-Holland, Amsterdam, 1959).

[3] G. Szekeres, J. Aust. Math. Soc. **8**, 134 (1968).

[4] J.W. Schutz, *Foundations of Special Relativity: Kinematic Axioms for Minkowski Space–Time*, Lecture Notes in Mathematics (Springer, Berlin-Heidelberg-New York, 1973), Vol. 361. This will be referred to subsequently as KA.

[5] D. Hilbert, *Grundlagen der Geometrie* (Teubner, Leipzig, 1913). Hilbert describes plane geometry in terms of two sets of undefined elements, points and lines, and an undefined relation of incidence. An alternative formal approach which describes a geometry as a set of points, with certain subsets called lines, is given by K. Borsuk and W. Szmielew, *Foundations*

of Geometry (North-Holland, Amsterdam, 1960). In the present axiomatic system, we regard Minkowski space–time as a set of paths, each path being a set of undefined elements called "events."

[6]A. Einstein, Ann. Phys. (Leipzig) **17**, 891 (1905).

[7]A.A. Robb, *Geometry of Time and Space* (Cambridge University, Cambridge, England, 1936).

[8]J. Tits, Bull. Soc. Math. Belg. **4**, 44 (1952).

[9]J. Tits, Bull. Soc. Math. Belg. **5**, 126 (1953).

[10]J. Tits, Mem. Acad. R. Belg. Sci. **29**, 166 (1955).

[11]Details of all proofs are given in the following research report: J.W. Schutz, "An Axiomatic System for Minkowski Space–Time," MPI-PAE/Astro 181, April 1979 published by (and available on request from) Max-Planck-Institut für Physik und Astrophysik, Föhringer Ring 6, 8 München 40, Federal Republic of Germany. In this report the temporal order relation has been defined as a relation on the set of events, rather than as a relation on each path as in the present paper, and the axiom of connectedness (axiom IV, 1.31) is stated differently. The axiom of uniqueness (axiom II, 1.12) ensures the consistency of the two alternative definitions of temporal order.

[12]E.H. Kronheimer and R. Penrose, Proc. Cambridge Philos. Soc. **63**, 481 (1967).

[13]N.M.J. Woodhouse, J. Math. Phys. **14**, 495 (1973).

[14]O. Veblen, Trans. Am. Math. Soc. **5**, 343 (1904).

[15]The significance of the statements "$d \lesssim a$" and "$d \lesssim f$" is that there should be paths joining the pairs of events d, a and d, f.

[16]The term SPRAY in this context should not be confused with the different meaning assigned to the same word in the theory of second order differential equations as given by, for example, S. Lang, *Introduction to Differentiable Manifolds* (Wiley, New York, 1962).

[17]CSP$\langle \mathbf{R}, \mathbf{S} \rangle$ is a linearly ordered set, so any of its members separates it into two sides. A more detailed definition of the concept of left and right sides is given in Theorem 1.512.

[18]Also see Theorem 1.73 for a more complete definition.

[19]L. Redei, "Foundation of Euclidean and Non-Euclidean Geometries According to F. Klein," *International Series of Monographs in Pure and Applied Mathematics* (Pergamon, Oxford-New York-Toronto, 1968), Vol. 97.

[20]It would be sufficient to consider a mapping ϕ of spr[x] into itself. Then the statement of (ii) would have to be modified because ϕ would, in general, send subsets of paths onto subsets of paths.

[21]The axiomatic system could be restated by rephrasing each theorem with the conditional statement "If there are at least two distinct paths..." prior to the statement of the theorem as given. Then with a simple restatement of the axiom of dimension (axiom IX, 2.1) a separate axiom of existence would not be required and the axiomatic system would appear to have one less axiom.

[22]Veblen also states an axiom of uniqueness of parallels which is required for Euclidean geometry, but not for ordered geometry.

[23]With the following substitutions for quoted propositions, where the first-mentioned 2 is to KA and the second is to the present treatment: [Axiom I, 2.2 |1.72], [Theorem 1, 2.5 |1.72], [Axiom, VIII, 2.10 |Axiom IX (2.1)], [Axiom X, 2.12 |Axiom IV (1.31)], [Corollary 1 to Theorem 33, 6.4 |Corollary 1.512], [Theorem 56, 8.4 |1.9], [Theorem 57, 9.1 |2.33].

[24]M.A. Castagnino, J. Math. Phys. **12**, 2203 (1971).

[25]W. Kundt and B. Hoffmann, "Determination of Gravitational Standard Time," in *Recent Developments in General Relativity*, (Warszawa, Warszaw, 1962), p. 63.

[26]R.F. Marzke and J.A. Wheeler, "Gravitation as Geometry I," in *Gravitation and Relativity*, edited by H.Y. Chiu and W.F. Hoffman (Benjamin, New York, 1964).

[27]F.A.E. Pirani, *Building Space–Time from Light Rays and Free Particles*, Symposia Mathematica (Academic, London, 1973), Vol. XII, p. 67.

[28]J.L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960).

[29]H. Reichenbach, *Axiomatik der relativistischen Raum–Zeit* (Vieweg, Braunschweig, 1924).

[30]H. Weyl, *Raum, Zeit, Materie* (Springer, Berlin, 1918); and H. Weyl, *Mathematische Analyse des Raumproblems* (Springer, Berlin, 1913).

[31]J. Ehlers, F.A. Pirani, and A. Schild, "The Geometry of Free Fall and Light Propagation," in *General Relativity, Papers in Honour of J.L. Synge* (Oxford University, Oxford, 1972).

The spreading of free wave packets and the entropy of position

Marian Grabowski
Institute of Physics, Nicholas Copernicus University, Toruń, Poland

(Received 15 August 1979; accepted for publication 7 December 1979)

The spreading of free wave packets is expressed by means of the entropy of position for a certain class of states. Connection between such formulation and the usual treatment is discussed.

PACS numbers: 03.65.Bz

1. INTRODUCTION

In quantum mechanics we may define a notion which is a measure of uncertainty of the localization of a particle in the space. If ψ is a wavefunction which describes the state of the system then $|\psi|^2$ is the probability density on the position space. We may consider the quantity

$$S(|\psi|^2) = -\int_{\mathbb{R}^1} |\psi(x)|^2 \ln|\psi(x)|^2\, dx.$$

It has the properties of an entropy and we call it the entropy of position. The uncertainty relation for position and momentum and spreading of a free wave packet expresses in principle the uncertainty of localization of a particle. The uncertainty of outcome of simultaneous measurement of position and momentum and increasing of uncertainty of position during the free movement of the particle would be expressed by means of such entropy. In Refs. 1 and 2 it was shown that the sum of these entropies for position and momentum in a given pure state is bounded from below. Using this inequality in the papers quoted above the authors derived the uncertainty relation in the usual form.

This article is an attempt to express the fact of spreading of a wave packet for a free particle by means of entropy of position. We also compare such formulation with the usual one.

2. ENTROPY OF POSITION

We consider a moving particle on a real line $\mathbb{R}^1$. (An extension of our result to higher dimensions is obvious.) The Hilbert space corresponding to this system is $L^2(\mathbb{R},dx)$—the space of square-integrable functions with the Lebesgue measure on a line. The position operator of the particle is defined as follows $(Q\psi)(x) = x\psi(x)$ for every $x\in\mathbb{R}^1$ and all wave functions $\psi\in D(Q)$—the dense domain of Q in $L^2(\mathbb{R}^1,dx)$. From the spectral theorem[3] we know that for every Borel set $\Delta\in\mathscr{B}(\mathbb{R}^1)$ [$\mathscr{B}(\mathbb{R}^1)$—$\sigma$-algebra of Borel sets on $\mathbb{R}^1$] a projector $E(\Delta)$ corresponding to Q is defined by the equality

$$(E(\Delta)\psi)(x) = \chi_\Delta(x)\psi(x),$$

for every $\psi\in L^2(\mathbb{R}^1,dx)$, where $\chi_\Delta(x)$ is a characteristic function of Δ.

The probability that the measurement of Q in a state ψ gives a result in the Borel set Δ, is $\mu_Q^\psi(\Delta) = \int_\Delta |\psi(x)|^2\, dx$. Thus for position measurement in a pure state the Radon–Nikodym derivative is $p(x) = |\psi(x)|^2$ and the entropy of position in a state ψ is equal to

$$S(|\psi|^2) = -\int |\psi(x)|^2 \ln|\psi(x)|^2\, dx. \tag{1}$$

The possibilities of defining and the general properties of the entropy of an observable were investigated for observables with purely discrete spectrum,[4] for observables with absolutely continuous spectrum[5] and for generalized observables.[6] In Refs 5 and 6 we obtained some inequalities for a von Neumann measurement of an observable with continuous spectrum and for a fuzzy observable which confirm our interpretation of (1) as an uncertainty of an outcome of an observable measurement.

3. ENTROPY OF POSITION AND A FREE WAVE PACKET

The dynamics of the particle is described by the group of unitary operators $U(t) = e^{-itH_0}$ where $H_0 = p^2/2m$, p—the operator of momentum, m—mass of the particle.

We show, that for a certain class of wavefunctions the effect of spreading of the wave packets may be expressed by means of the entropy of position.

Let $\psi\in L^2(\mathbb{R}^1)\cap L^1(\mathbb{R}^1)$ and $\|\psi\|_2 = 1$, then for $t>0$ we have the following inequality:

$$S(|\psi|^2) + S(|\psi(t)|^2) > \ln(et\pi/m), \tag{2}$$

where $\psi(t) = e^{-itH_0}\psi$. To show (2) we need the following result of Ref. 1.

Let $\psi\in L^p(\mathbb{R}^n)$— the space of p-integrable functions with norm for $1\leqslant p\leqslant 2$

$$\|\psi\|_p = \left(\int |\psi(x)|^p\, d^n x\right)^{1/p},$$

and

$$\hat\psi(k) = \frac{1}{\sqrt{(2\pi)^n}}\int e^{-ikx}\psi(x)\, d^n x$$

is a Fourier transform of ψ, then

$$\|\hat\psi\|_{p'} \leqslant C(p',p)\|\psi\|_p, \tag{2a}$$

where

$$C(p',p) = \left(\frac{2\pi}{p'}\right)^{n/2p'}\left(\frac{2\pi}{p}\right)^{-n/2p}, \quad \text{and} \quad \frac{1}{p'} + \frac{1}{p} = 1.$$

We take an explicit form $U(t) = e^{-itH_0}$ for $t\in\mathbb{R}^1$ and $\psi\in L^2(\mathbb{R}^1)$ [7]:

$$\psi(x,t) = (e^{-itH_0}\psi)(x) = \underset{R\to\infty}{\text{l.i.m.}}\left(\frac{m}{2\pi it}\right)^{1/2}\int_{|y|<R} e^{im|x-y|^2/2t}\psi(y)\, dy, \tag{3}$$

where l.i.m. denotes the limit in the L^2-norm. Using

 0022-2488/81/020303-03$1.00

$\psi \in L^2(\mathbb{R}^1) \cap L^1(\mathbb{R}^1)$ and

$$\exp\left(\frac{mi|x-y|^2}{2t}\right) = \exp\left(\frac{mix^2}{2t}\right)\exp\left(-\frac{imxy}{t}\right)\exp\left(\frac{miy^2}{2t}\right),$$

we have

$$\psi(x,t) = \left(\frac{m}{it}\right)^{1/2}\exp\left(\frac{imx^2}{2t}\right)\hat{F}_t\left(\frac{mx}{t}\right),$$

where

$$F_t(y) = \exp\left(\frac{imy^2}{2t}\right)\psi(y).$$

Now we obtain

$$\begin{aligned}\|\psi(t)\|_{p'} &= \left(\frac{m}{t}\right)^{1/2}\left\|\hat{F}_t\left(\frac{m\cdot}{t}\right)\right\|_{p'} \\ &= \left(\frac{t}{m}\right)^{-[(1/2)-(1/p')]}\|\hat{F}_t\|_{p'} \\ &\leqslant \left(\frac{t}{m}\right)^{-[(1/2)-(1/p')]}C(p',p)\|F_t\|_p \\ &= \left(\frac{t}{m}\right)^{-[(1/2)-(1/p')]}C(p',p)\|\psi\|_p, \qquad (4)\end{aligned}$$

where in the second equality we have changed the variable in the integral, then we have used the mentioned inequality (2a) for the Fourier transform. Using $(1/p') + (1/p) = 1$ we may write (4) in the form

$$f(p') = \left(\frac{t}{m}\right)^{-[(1/2)-(1/p')]}C(p',p)\|\psi\|_p - \|\psi(t)\|_{p'} \geqslant 0,$$

for $p' \geqslant 2$, The function f is equal to zero at $p' = 2$ because $\|\psi\|_2 = \|\psi(t)\|_2$ and $C(2,2) = 1$. Thus the right derivative of $f(p')$ at $p' = 2$ must be nonnegative. Because $\psi \in L^2 \cap L^1$, $\exp(-itH_0)$ extends uniquely to a map from $L^p(\mathbb{R}^1)$ to $L^{p'}(\mathbb{R}^1)$ by the Riesz-Thorin theorem[8] for $p' \in [2,\infty]$ (for details see Ref. 8, proof of Theorem IX. 30). The function $f(p')$ is determined on the interval $[2,\infty]$ and we may take such derivatives. For $\|\psi\|_2 = 1$, the right derivative of $f(p')$ at $p' = 2$ reduces to (2).

Equality in (2) holds if

$$\psi(y)\exp\left(\frac{imy^2}{2t}\right) = \exp(-ay^2).$$

Such a function ψ does not exist because ψ is not a function of time. We have sharp inequality but $\ln(\pi et/m)$ is the best possible lower bound for the sum $S(|\psi|^2)$ and $S(|\psi(t)|^2)$.

Using the concavity of entropy we may extend our inequality to mixed states, which have the following spectral decomposition $\rho = \Sigma p_i |\psi_i\rangle\langle\psi_i|$, where $|\psi_i\rangle$ is the basis from $L^2 \cap L^1$. Extension to $L^2(\mathbb{R}^n)$ is trivial and yields $n\ln(\pi et/m)$ in (2).

From (2) we see that the sum of uncertainties of localization of a particle at $t = 0$ and $t > 0$ is bounded from below by an increasing function of time (logarithm). For two instants of time $t_2 > t_1$ such that

$$\ln(\pi et_2/m) \geqslant S(|\psi(t_1)|^2) + S(|\psi|^2),$$

we have

$$S(|\psi(t_2)|^2) \geqslant S(|\psi(t_1)|^2).$$

If we wait sufficiently long then the uncertainty $S(|\psi(t)|^2)$ will increase. Thus (2) in a certain manner reflects the spreading of the free wave packets.

4. DISCUSSION

Usually in textbooks on quantum mechanics (for example Ref. 9) the spreading of free wave packets is expressed by means of the quantity

$$\Delta Q^2 = \int (x - \langle Q\rangle)^2 |\psi(x)|^2\,dx, \quad \text{where } \langle Q\rangle = (\psi, Q\psi).$$

Further for $\psi \in D(Q) \cap D(P)$, where $D(P)$ is a domain of momentum P we obtain

$$\Delta Q_t^2 = \Delta Q^2 + 2\left[\tfrac{1}{2}\langle QP + PQ\rangle - \langle Q\rangle\langle P\rangle\right]\frac{t}{m} + \Delta P^2\frac{t^2}{m^2},$$

where

$$\Delta P^2 = \int (p - \langle P\rangle)^2 |\hat{\psi}(p)|^2\,dp.$$

Thus the standard deviation of position ΔQ_t^2 is an increasing function of time at least starting from a certain point at which it takes minimum. This means that the uncertainty (according to this measure) of particle position will systematically increase if we wait sufficiently long. When $\psi \in D(Q)$, then $\Delta Q^2 = \infty$, and it is not a good measure of spreading of the free packets. Inequality (2) enlarges the set of wavefunctions for which we consider the effect of spreading. In $L^1 \cap L^2$ there exist the functions for which $\Delta Q^2 = \infty$ but the entropy of position is finite. The following function is a simple example

$$x = \sqrt{2}\,\frac{1}{x^{3/2}}\chi_{[1,\infty]}(x),$$

where $\chi_{[1,\infty]}(x)$ is a characteristic function of interval $[1,\infty]$. If $\psi \in D(Q) \cap D(P)$ and is integrable then of course the entropy is also finite.

We will show how (2) is connected with the usual description of spreading of wave packets for the free evolution. Let

$$\mathfrak{M}_0 = \left\{|\psi(x)|^2;\ \int (x - \langle Q\rangle)^2 |\psi(x)|^2\,dx = \Delta Q^2\right\}$$

$$\mathfrak{M}_t = \left\{|\psi(x,t)|^2;\ \int (x - \langle Q\rangle_t)^2 |\psi(x,t)|^2\,dx = \Delta Q_t^2\right\}$$

be two sets of wavefunctions (the functions are determined up to a factor $e^{if(x)}$, f—real function on $\mathbb{R}^1$) with a finite variance at $t = 0$ and $t > 0$. Only for such sets does the position observable exists.[10] Moreover, for every function belonging to these sets $S(|\psi|^2)$ and $S(|\psi(t)|^2)$ are finite. No difficulties with the interpretation of $S = \pm\infty$ arise. The states $\psi \in \mathfrak{M}_0$ with maximal entropy of position has the following density of probability distribution

$$|\psi(x)|_G^2 = (2\pi\Delta Q^2)^{-1/2}\exp\left(-\frac{(x - \langle Q\rangle)^2}{2\Delta Q^2}\right),$$

and similarly

$$|\psi(x,t)|_G^2 = (2\pi\Delta Q_t^2)^{-1/2}\exp\left(-\frac{(x - \langle Q\rangle_t)^2}{2\Delta Q_t^2}\right).$$

Thus we obtain

$$S(|\psi|^2)+S(|\psi(t)|^2)\leqslant S(|\psi|_G^2)+S(|\psi(t)|_G^2)$$
$$=\tfrac{1}{2}\ln 4\pi^2 e^2 \Delta Q^2 \Delta Q_t^2.$$

Using (2) and the last inequality we have

$$\Delta Q^2 \Delta Q_t^2 > (t^2/4m^2).$$

We see that for $t > 2\Delta Q^2 m$

$$\Delta Q_t^2 > \Delta Q^2$$

holds. For a time smaller than $2\Delta Q^2 m$ from (2) we do not obtain any information above the behavior of the variance. Equation (2) becomes an equality on the Gaussian state for an asymptotic formula of e^{-itH_0}.

$$(e^{-itH_0}\psi)(x)\to\left(\frac{m}{it}\right)^{1/2}\hat{\psi}\left(\frac{mx}{t}\right)$$

in the sense that the difference goes to zero in the L^2-norm at $t\to\infty$. It is a slight modification of IX.33 in Ref. 7. For long times, factors of the form $\exp(imx^2/2t)$ may be neglected. A simple calculation for the asymptotic formula and $\psi\in D(Q)\cap D(P)$ shows that

$$\Delta Q_t^2=\left(\frac{t^2}{m^2}\right)\Delta P^2, \quad \text{and} \quad \Delta Q_t^2 \Delta Q^2 \geqslant \left(\frac{t^2}{4m^2}\right).$$

The last inequality becomes an equality also on the Gaussian state. We see that our result agrees with the usual formulation for great values of time.

Using (3) simple calculations show that for $\psi\in L^1\cap L^2$ a more elementary notion also expresses the effect of spreading

$$\mu_Q^{\psi(t)}([a,b\,]) = \int_a^b |\psi(x,t)|^2\,dx\to 0,$$

when $t\to\infty$.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor R.S. Ingarden for his kind interest and discussion about this work. I am very indebted to Professor A. Wehrl (Vienna) for reading this paper and for his comments and remarks.

[1]W. Beckner, Ann. Math. **102**, 159 (1975).
[2]I. Białynicki-Birula and J. Mycielski, Commun. Math. Phys. **44**, 129 (1975).
[3]B. Simon and M. Reed, *Methods of Modern Mathematical Physics*, (Academic, New York, 1972), Vol. 1.
[4]A. Wehrl, Rep. Math. Phys. **12**, 386 (1977).
[5]M. Grabowski, Int. J. Theor. Phys. **17**, 635 (1978).
[6]M. Grabowski, Rep. Math. Phys. **14**, 377 (1978).
[7]B. Simon and M. Reed, *Method of Modern Mathematical Physics*, (Academic, New York, 1976), Vol. 2.
[8]See Ref. 7.
[9]E. Durand, *Mécanique Quantique*, (Masson, Paris, 1970), Vol. 1.
[10]K. Kraus and J. Schröter, Int. J. Theor. Phys. **7**, 431 (1973).

Nonrelativistic Coulomb Green's function in parabolic coordinates

S. M. Blinder
Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109

(Received 5 December 1979; accepted for publication 25 January 1980)

The nonrelativistic Coulomb Green's function $G^{(+)}(\mathbf{r}_1,\mathbf{r}_2,k)$ is evaluated by explicit summation over discrete and continuum eigenstates in parabolic coordinates. This completes the derivation of Meixner, who was able to obtain only the $\mathbf{r}_1 = 0$ and $\mathbf{r}_2 \to \infty$ limiting forms of the Green's function. Further progress is made possible by an integral representation for a product of two Whittaker functions given by Buchholz. We obtain the closed form for the Coulomb Green's function previously derived by Hostler, via an analogous summation in spherical polar coordinates. The Rutherford scattering limit of the Green's function is also demonstrated, starting with an integral representation in parabolic coordinates.

PACS numbers: 03.65.Db, 02.30.Hq, 02.30.Gp

1. INTRODUCTION

The nonrelativistic Coulomb Green's function $G(\mathbf{r}_1,\mathbf{r}_2,k)$ [$G(1,2,k)$ for short] is the solution under specified boundary conditions of the equation

$$(k^2 + \nabla_1^2 + 2Z/r_1)G(\mathbf{r}_1,\mathbf{r}_2,k) = \delta(\mathbf{r}_1 - \mathbf{r}_2). \tag{1.1}$$

Atomic units: $h = m = e = 1$, are employed for convenience. Any Green's function can, in principle, be constructed from its spectral representation

$$G(1,2,k) = \sum_n \frac{\psi_n(1)\psi_n^*(2)}{k^2 - \epsilon_n}, \tag{1.2}$$

the summation running over the complete set of discrete and continuum eigenstates. Meixner[1] in 1933 attempted to evaluate the Coulomb Green's function by explicit summation over eigenfunctions in parabolic coordinates. He was able, however, to obtain closed forms only in the special cases $\mathbf{r}_1 = 0$ and $\mathbf{r}_2 \to \infty$. Hostler[2] first worked out the general closed-form expression for $G(1,2,k)$ by summing over Coulomb eigenfunctions in spherical polar coordinates. A key element in Hostler's derivation was an integral representation for a product of two Whittaker functions given by Buchholz.[3]

We shall demonstrate in this paper that Hostler's result can also be derived by working in parabolic coordinates. We will thus explicitly complete the work of Meixner. In addition, we shall obtain in straightforward fashion the Rutherford scattering limit of the Green's function and also a possible starting point for a compact treatment of the Stark effect.

2. COULOMB EIGENFUNCTIONS IN PARABOLIC COORDINATES

Parabolic coordinates (ξ,η,ϕ) can be defined in terms of spherical polar coordinates (r,θ,ϕ) and Cartesian coordinates (x,y,z) by the relations

$$\begin{aligned} \xi &= r(1 + \cos\theta) = r + z, \\ \eta &= r(1 - \cos\theta) = r - z, \\ \phi &= \phi = \tan^{-1}(y/x). \end{aligned} \tag{2.1}$$

Conversely

$$x = (\xi\eta)^{1/2}\cos\phi, \quad y = (\xi\eta)^{1/2}\sin\phi,$$

$$z = \tfrac{1}{2}(\xi - \eta), \quad r = \tfrac{1}{2}(\xi + \eta). \tag{2.2}$$

The volume element is given by

$$d^3\mathbf{r} = \tfrac{1}{4}(\xi + \eta)\,d\xi\,d\eta\,d\phi \quad (0 \leqslant \xi,\eta \leqslant \infty, 0 \leqslant \phi \leqslant 2\pi), \tag{2.3}$$

and the Laplacian operator by

$$\nabla^2 = \frac{4}{\xi + \eta}\left(\frac{\partial}{\partial\xi}\xi\frac{\partial}{\partial\xi} + \frac{\partial}{\partial\eta}\eta\frac{\partial}{\partial\eta}\right) + \frac{1}{\xi\eta}\frac{\partial^2}{\partial\phi^2}. \tag{2.4}$$

The Coulomb Schrödinger equation

$$(\epsilon + \nabla^2 + 2Z/r)\psi = 0, \quad \epsilon = 2E = k^2 \tag{2.5}$$

is separable in parabolic coordinates as well as spherical polar coordinates. The factorization

$$\psi(\xi,\eta,\phi) = f_1(\xi)f_2(\eta)e^{im\phi}, \quad m = 0, \pm 1, \pm 2, \cdots, \tag{2.6}$$

leads to the ordinary differential equations[4]

$$\begin{aligned} \frac{d}{d\xi}\left(\xi\frac{df_1}{d\xi}\right) + \left(Z_1 + \frac{k^2\xi}{4} - \frac{m^2}{4\xi}\right)f_1(\xi) = 0, \\ \frac{d}{d\eta}\left(\eta\frac{df_2}{d\eta}\right) + \left(Z_2 + \frac{k^2\eta}{4} - \frac{m^2}{4\eta}\right)f_2(\eta) = 0, \end{aligned} \tag{2.7}$$

where

$$Z_1 + Z_2 = Z. \tag{2.8}$$

Either Z_1 or Z_2 labels the one-parameter family of degenerate eigenstates for each value of $\epsilon = k^2$.

The subtitutions

$$\begin{aligned} &f(x) = x^{-1/2}M(-ikx), \\ &x = \xi,\eta, \quad z \equiv -ikx, \quad \nu_{1,2} \equiv Z_{1,2}/k, \end{aligned} \tag{2.9}$$

bring (2.7) into the form of Whittaker's differential equation

$$M''(z) + \left(-\frac{1}{4} + \frac{i\nu}{z} + \frac{1 - m^2}{4z^2}\right)M(z) = 0. \tag{2.10}$$

The solutions to (2.10) regular at $x = 0$ are the Whittaker functions $M^{m/2}_{\pm i\nu}(\mp ikx)$.[5] For $m \geqslant 0$

$$\begin{aligned} &M^{m/2}_{\pm i\nu}(\mp ikx) \\ &\quad = (m!)^{-1}(\mp ikx)^{(m+1)/2}\,e^{\mp ikx/2} \\ &\qquad \times {}_1F_1((m+1)/2 \pm i\nu;\, m+1;\, \pm ikx), \end{aligned} \tag{2.11}$$

where ${}_1F_1$ is a confluent hypergeometric function. For $m < 0$,

 0022-2488/81/020306-06$1.00

the corresponding Whittaker functions are given through the identity

$$\Gamma((1-m)/2-i\nu)\,M_{i\nu}^{-m/2}(-ikx) = \Gamma((1+m)/2-i\nu)\,M_{i\nu}^{m/2}(-ikx). \tag{2.12}$$

The functions (2.11) with alternative choice of signs are related by

$$M_{-i\nu}^{m/2}(ikx) = e^{i\pi(m+1)/2}\,M_{i\nu}^{m/2}(-ikx), \tag{2.13}$$

which shows incidentally that $e^{i\pi(m+1)/4}M_{i\nu}^{m/2}(-ikx)$ is a real function.

The asymptotic form for M as $x\to\infty$ is given by[6]:

$$M_{i\nu}^{m/2}(-ikx)\sim e^{-\pi\nu/2}\left[\frac{(kx)^{-i\nu}e^{-ikx/2}}{\Gamma((m+1)/2-i\nu)} + e^{-i\pi(m+1)/2}\frac{(kx)^{i\nu}e^{ikx/2}}{\Gamma((m+1)/2+i\nu)}\right]. \tag{2.14}$$

For positive energy, the wavenumber k is real; to avoid divergences in the wavefunctions (2.9) we must require that

$$|\mathrm{Im}\nu|\leqslant\tfrac{1}{2}. \tag{2.15}$$

Positive-energy eigenstates can be specified by the two continuous quantum numbers

$$\nu_1\equiv Z_1/k \text{ and } \nu_2\equiv Z_2/k. \tag{2.16}$$

Thus, Eq. (2.8) implies

$$k = Z/(\nu_1+\nu_2). \tag{2.17}$$

It is sufficient to assume $k\geqslant 0$ and to choose real values for ν_1 and ν_2. Then both ν_1 and ν_2 can run over the range $(-\infty,+\infty)$ but, by virtue of (2.17), their sum must be nonnegative:

$$\nu_1+\nu_2\geqslant 0. \tag{2.18}$$

For compactness, we shall continue to write k in the arguments of Whittaker functions, understanding k to be a function of ν_1 and ν_2 via (2.17).

The foregoing considerations lead to the following positive-energy Coulomb eigenfunctions in parabolic coordinates:

$$\psi_{\nu_1\nu_2 m}(\xi,\eta,\phi) = e^{i\pi(m+1)/2}(2\pi)^{-3/2}Z^{-1/2}k e^{\pi(\nu_1+\nu_2)/2} \times|\Gamma((m+1)/2-i\nu_1)|\,|\Gamma((m+1)/2-i\nu_2)|(\xi\eta)^{-1/2} \times M_{i\nu_1}^{m/2}(-ik\xi)\,M_{i\nu_2}^{m/2}(-ik\eta)e^{im\phi}. \tag{2.19}$$

The phase factor $e^{i\pi(m+1)/2}$ is retained for later convenience. These continuum eigenfunctions are orthonormalized according to

$$\int_0^\infty\int_0^\infty\int_0^{2\pi}\psi^*_{\nu_1\nu_2 m}(\xi,\eta,\phi)\psi_{\nu_1'\nu_2' m'}(\xi,\eta,\phi)\tfrac{1}{4}(\xi+\eta)\,d\xi\,d\eta\,d\phi = \delta(\nu_1-\nu_1')\,\delta(\nu_2-\nu_2')\delta_{mm'}. \tag{2.20}$$

Meixner[1] and other authors employed eigenfunctions normalized to $\delta(k-k')\delta(\zeta-\zeta')$, in which ζ corresponds to our variable $(\nu_1-\nu_2)/2$. The more symmetrical normalization scheme we have introduced will facilitate evaluation of the Green's function.

Equation (2.20) can be demonstrated with the help of the following integrals over Whittaker functions:

$$\int_0^\infty M_{i\nu}^{m/2}(-ikx)M_{i\nu}^{m/2}(-ik'x)\,dx = 4\pi e^{-i\pi(m+1)/2}e^{-\pi\nu}\delta(k-k')\Big/\left|\Gamma\left(\frac{m+1}{2}-i\nu\right)\right|^2 \tag{2.21}$$

and

$$\int_0^\infty M_{i\nu}^{m/2}(-ikx)M_{i\nu'}^{m/2}(-ikx)x^{-1}\,dx = 4\pi e^{-i\pi(m+1)/2}e^{-\pi\nu}\delta(\nu-\nu')\Big/\left|\Gamma\left(\frac{m+1}{2}-i\nu\right)\right|^2. \tag{2.22}$$

Equation (2.21) also occurs in the normalization of spherical eigenfunctions. Both (2.21) and (2.22) can be demonstrated from integral representations of Whittaker functions in terms of Bessel functions[7] with use of an integral given by Watson.[8] More simply, by virtue of the fact that the principal contributions to (2.21) and (2.22) come from the asymptotic region $x\to\infty$, it suffices to approximate the Whittaker functions by their asymptotic forms (2.14). Using (2.16),

$$\delta(\nu_1-\nu_1')\delta(k-k') = (Z_2/k^2)\delta(\nu_1-\nu_1')\delta(\nu_2-\nu_2'),$$
$$\delta(\nu_2-\nu_2')\delta(k-k') = (Z_1/k^2)\delta(\nu_1-\nu_1')\delta(\nu_2-\nu_2'), \tag{2.23}$$

which completes the required normalization.

The negative-energy parabolic eigenfunctions are quite standard.[9] Expressed in terms of Whittaker functions[10]:

$$\psi_{n_1n_2m}(\xi,\eta,\phi) = \frac{Z^{1/2}}{\pi^{1/2}n}\left[\frac{(|m|+n_1)!(|m|+n_2)!}{n_1!n_2!}\right]^{1/2}(\xi\eta)^{-1/2} \times M_{n_1+(|m|+1)/2}^{|m|/2}(Z\xi/n)M_{n_2+(|m|+1)/2}^{|m|/2}(Z\eta/n)e^{im\phi}, \tag{2.24}$$
$$n_1,n_2 = 0,1,2,\cdots;\ m = 0,\pm1,\pm2,\cdots;$$
$$n = n_1+n_2+|m|+1 = 1,2,3,\cdots.$$

We shall also require Whittaker functions of the second kind, $W_{i\nu}^{m/2}(-ikx)$, which represent solutions to (2.10) having the form of outgoing waves. Specifically we note the transformation[11]

$$M_{i\nu}^{m/2}(-ikx) = e^{-\pi\nu}\left[\frac{W_{-i\nu}^{m/2}(ikx)}{\Gamma((m+1)/2-i\nu)} + e^{-i(m+1)\pi/2}\frac{W_{i\nu}^{m/2}(-ikx)}{\Gamma((m+1)/2+i\nu)}\right], \tag{2.25}$$

the identity[12]

$$W_{i\nu}^{-m/2}(-ikx) = W_{i\nu}^{m/2}(ikx), \tag{2.26}$$

and the asymptotic form as $x\to\infty$[13]

$$W_{i\nu}^{m/2}(-ikx)\sim(-ikx)^{i\nu}e^{ikx/2}. \tag{2.27}$$

Note that Eq. (2.14) also follows from (2.25) with (2.27).

For values of ν occurring in the discrete spectrum [cf. (2.24)] the two types of Whittaker functions become proportional.[14] Specifically

$$W_{n'+(m+1)/2}^{m/2}(z) = (-)^{n'}(n'+m)!M_{n'+(m+1)/2}^{m/2}(z). \tag{2.28}$$

A key result, both in Hostler's derivation and in the present work, is an integral representation for a product of two Whittaker functions given by Buchholz.[3] With appropriate specialization of the variables, we write

$$\Gamma((m+1)/2 - i\nu)M_{i\nu}^{m/2}(-iky)W_{i\nu}^{m/2}(-ikx)$$
$$= (-i)^{m+1}k(xy)^{1/2}\int_0^\infty ds \exp\left[+\frac{i}{2}k(x+y)\cosh s\right]$$
$$\times J_m\left(k\sqrt{xy}\sinh s\right)[\coth(s/2)]^{2i\nu},$$
$$\mathrm{Re}((m+1)/2 - i\nu) > 0, \quad \mathrm{Im}k > 0, \quad x > y. \tag{2.29}$$

3. EVALUATION OF THE GREEN'S FUNCTION

The summation (1.2) explicitly written out in terms of discrete and continuum parabolic quantum numbers becomes

$$G(1,2,k) = \sum_{m=-\infty}^{\infty}\left[\sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty}\left(k^2 + \frac{Z^2}{n^2}\right)^{-1}\psi_{n_1n_2m}(1)\right.$$
$$\times\psi^*_{n_1n_2m}(2) + \int_{-\infty}^{\infty}d\nu_1\int_{-\infty}^{\infty}d\nu_2\,\theta(\nu_1+\nu_2)$$
$$\left.\times(k^2-\kappa^2)^{-1}\psi_{\nu_1\nu_2m}(1)\psi^*_{\nu_1\nu_2m}(2)\right]. \tag{3.1}$$

The Heaviside function

$$\theta(x) \equiv \begin{cases} 1, & \text{for } x \geqslant 0, \\ 0, & \text{for } x < 0. \end{cases} \tag{3.2}$$

has been introduced to take account of the condition (2.18). The wavenumber in the eigenfunctions has been redesignated κ, to reserve k for the Green's function. Putting in the eigenfunctions (2.19) and (2.24), we obtain

$$G(1,2,k) = \sum_{m=-\infty}^{\infty}\frac{e^{im(\phi_1-\phi_2)}}{(\xi_1\xi_2\eta_1\eta_2)^{1/2}}$$
$$\times\left\{\sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty}\left(k^2+\frac{Z^2}{n^2}\right)^{-1}\frac{Z}{\pi n^2}\frac{(|m|+n_1)!(|m|+n_2)!}{n_1!n_2!}\right.$$
$$\times M^{|m|/2}_{n_1+(|m|+1)/2}(Z\xi_1/n)M^{|m|/2}_{n_1+(|m|+1)/2}(Z\xi_2/n)$$
$$\times M^{|m|/2}_{n_2+(|m|+1)/2}(Z\eta_1/n)M^{|m|/2}_{n_2+(|m|+1)/2}(Z\eta_2/n)$$
$$+ e^{i\pi(m+1)}\frac{Z}{8\pi^3k^2}\int_{-\infty}^{\infty}d\nu_1\int_{-\infty}^{\infty}d\nu_2\theta(\nu_1+\nu_2)$$
$$\times[(\nu_1+\nu_2)^2 - Z^2/k^2]^{-1}e^{\pi(\nu_1+\nu_2)}$$
$$\times\left|\Gamma\left(\frac{m+1}{2}-i\nu_1\right)\right|^2\left|\Gamma\left(\frac{m+1}{2}-i\nu_2\right)\right|^2$$
$$\times M^{m/2}_{i\nu_1}(-i\kappa\xi_1)M^{m/2}_{i\nu_1}(-i\kappa\xi_2)$$
$$\left.\times M^{m/2}_{i\nu_2}(-i\kappa\eta_1)M^{m/2}_{i\nu_2}(-i\kappa\eta_2)\right\}. \tag{3.3}$$

The M function of the argument $\xi_>$ (the greater of ξ_1, ξ_2) can be transformed to a sum of W functions using (2.25). We find

$$e^{im(m+1)/2}e^{\pi\nu_1}\left|\Gamma\left(\frac{m+1}{2}-i\nu_1\right)\right|^2$$
$$\times M^{m/2}_{i\nu_1}(-i\kappa\xi_1)M^{m/2}_{i\nu_1}(-i\kappa\xi_2)$$
$$= \Gamma\left(\frac{m+1}{2}+i\nu_1\right)M^{m/2}_{-i\nu_1}(i\kappa\xi_<)W^{m/2}_{-i\nu_1}(i\kappa\xi_>)$$
$$+\Gamma\left(\frac{m+1}{2}-i\nu_1\right)M^{m/2}_{i\nu_1}(-i\kappa\xi_<)W^{m/2}_{i\nu_1}(-i\kappa\xi_>). \tag{3.4}$$

We have made use of (2.13) to get $M^{m/2}_{-i\nu_1}$ to multiply $W^{m/2}_{-i\nu_1}$. Now the first term ΓMW in (3.4) can be transformed into the second by the substitutions

$$\nu_1 \to -\nu_1, \quad \nu_2 \to -\nu_2. \tag{3.5}$$

By applying this in the continuum integral and noting the identity

$$\theta(\nu_1+\nu_2) + \theta(-\nu_1-\nu_2) = 1, \tag{3.6}$$

the Heaviside function is eliminated.

The functions of η_1 and η_2 are transformed in an analogous way. Under the double integration, the two terms ΓMW make equal contributions. The continuum part of the Green's function thus reduces to the compact form

$$\frac{Z}{4\pi^3k^2}\int_{-\infty}^{\infty}d\nu_1\int_{-\infty}^{\infty}d\nu_2\,[(\nu_1+\nu_2)^2 - Z^2/k^2]^{-1}$$
$$\times\Gamma\left(\frac{m+1}{2}-i\nu_1\right)M^{m/2}_{i\nu_1}(-i\kappa\xi_<)W^{m/2}_{i\nu_1}(-i\kappa\xi_>)$$
$$\times\Gamma\left(\frac{m+1}{2}-i\nu_2\right)M^{m/2}_{i\nu_2}(-i\kappa\eta_<)W^{m/2}_{i\nu_2}(-i\kappa\eta_>). \tag{3.7}$$

The integrals in (3.7) can be most readily evaluated by interpreting them as contour integrals in the complex planes of ν_1 and ν_2. The integrand is an analytic function of each variable in its lower half-plane with the exception of a set of simple poles. One should verify this, in particular, for $\nu_2 = 0$ and $\nu_1 \to 0$ with $\mathrm{Im}\nu_1 < 0$. This corresponds to $\kappa = Z/\nu_1 \to \infty$ with $\mathrm{Im}\kappa > 0$. From (2.23) and (2.27) we find the relevant asymptotic dependence

$$\Gamma((m+1)/2 - i\nu)M^{m/2}_{i\nu}(-i\kappa x_<)W^{m/2}_{i\nu}(-i\kappa x_>)$$
$$\sim e^{i\kappa(x_>-x_<)/2}, \tag{3.8}$$

which approaches zero as $|\kappa| \to \infty$ with $\mathrm{Im}\kappa > 0$. By virtue of (2.12) and (2.26), m in the functions ΓMW can be replaced by $|m|$. This will make more explicit the poles of the integrand.

We note also the asymptotic behavior as $\nu_1 \to \infty$ with ν_2 fixed. As $k \to 0$[15]

$$\Gamma((m+1)/2 - i\nu)M^{m/2}_{i\nu}(-iky)W^{m/2}_{i\nu}(-ikx) \sim k. \tag{3.9}$$

Thus each factor $\Gamma MW \sim \nu_1^{-1}$. Including the energy denominator ($\sim\nu_1^{-2}$), the entire integrand behaves as ν_1^{-4} when $|\nu_1| \to \infty$.

Evidently, the ν_1 integral can be evaluated by application of the residue theorem after the contour is closed from below with a semicircle at infinity. As $|\nu_1| \to \infty$, the contribution from the semicircle approaches zero as a result of the asymptotic behavior discussed above. The singular points in the integrand arise from the factors $\Gamma((|m|+1)/2 - i\nu_1)$ and $[(\nu_1+\nu_2)^2 - Z^2/k^2]^{-1}$. The gamma function has poles at the points

$$\nu_1 = -i((|m|+1)/2 + n_1), \quad n_1 = 0,1,2,\cdots, \tag{3.10}$$

with the corresponding residues $i(-)^{n_1}/n_1!$. The energy de-

nominator has a pole in the lower half-plane at $\kappa = k + i\delta$ or

$$\nu_1 = -\nu_2 + Z/k - i\delta', \tag{3.11}$$

with residue $k/2Z$. $\mathrm{Im}k > 0$ is taken, such that the resulting Green's function will correspond to $G^{(+)}(1,2,k)$.

The ν_1 integral thereby reduces to $(-2\pi i)$ times the sum of the residues in the lower half-plane. The continuum integral (3.7) can thereby be expressed as a sum of two contributions: (3.12) plus (3.13). From the poles of the gamma function [cf. (3.10)]:

$$-2\pi i \frac{Z}{4\pi^3 k^2} \sum_{n_1=0}^{\infty} \frac{i(-)^{n_1}}{n_1!} \times \int_{-\infty}^{\infty} d\nu_2 \left[(\nu_1' + \nu_2)^2 - \frac{Z^2}{k^2}\right]^{-1} \times M_{i\nu_1'}^{|m|/2}(-i\kappa'\xi_<) W_{i\nu_1'}^{|m|/2}(-i\kappa'\xi_>) \Gamma\left(\frac{|m|+1}{2} - i\nu_2\right) \times M_{i\nu_2}^{|m|/2}(-i\kappa'\eta_<) W_{i\nu_2}^{|m|/2}(-i\kappa'\eta_>), \tag{3.12}$$

$$\nu_1' \equiv -i\left(\frac{|m|+1}{2} + n_1\right), \quad \kappa' \equiv Z/(\nu_1' + \nu_2).$$

From the energy factor [cf. (3.11)]:

$$-2\pi i \frac{Z}{4\pi^3 k^2} \frac{k}{2Z} \int_{-\infty}^{\infty} d\nu_2 \Gamma\left(\frac{|m|+1}{2} - i\nu + i\nu_2\right) \times M_{i\nu - i\nu_2}^{|m|/2}(-ik\xi_<) W_{i\nu - i\nu_2}^{|m|/2}(-ik\xi_<) \times \Gamma\left(\frac{|m|+1}{2} - i\nu_2\right) M_{i\nu_2}^{|m|/2}(-ik\eta_<) \times W_{i\nu_2}^{|m|/2}(-ik\eta_>) \quad (\nu \equiv Z/k). \tag{3.13}$$

The second integration in (3.12) can be carried out in an exactly analogous way. The ν_2 contour, again closed by an infinite semicircle in the lower half-plane, encloses only the poles of $\Gamma((|m|+1)/2 - i\nu_2)$, at the points

$$\nu_2 = -i((|m|+1)/2 + n_2), \quad n_2 = 0,1,2,\cdots. \tag{3.14}$$

Thus, in Eq. (3.12),

$$\nu' = -i(n_1 + n_2 + |m| + 1) \equiv -in, \quad n = 1,2,3,\cdots,$$
$$\kappa' = Z/\nu' = iZ/n, \tag{3.15}$$

and (3.12) becomes

$$(-2\pi i)^2 \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \frac{i^2(-)^{n_1+n_2}}{n_1! n_2!} \frac{Z}{4\pi^3 k^2} \left(-n^2 - \frac{Z^2}{k^2}\right)^{-1} \times M_{n_1+(|m|+1)/2}^{|m|/2}(Z\xi_</n) W_{n_1+(|m|+1)/2}^{|m|/2}(Z\xi_>/n) \times M_{n_2+(|m|+1)/2}^{|m|/2}(Z\eta_</n) W_{n_2+(|m|+1)/2}^{|m|/2}(Z\eta_>/n). \tag{3.16}$$

Application of (2.28) shows now that (3.16) exactly cancels the sum over the discrete spectrum in (3.3). The Green's function is thus reduced to the contribution containing (3.13). Writing λ in place of ν_2 and reintroducing m in place of $|m|$:

$$G^{(+)}(1,2,k) = \frac{1}{4\pi^2 ik} \sum_{m=-\infty}^{\infty} \frac{e^{im(\phi_1 - \phi_2)}}{(\xi_<\xi_>\eta_<\eta_>)^{1/2}} \times \int_{-\infty}^{\infty} d\lambda \left[\Gamma\left(\frac{m+1}{2} - i\nu + i\lambda\right) M_{i\nu - i\lambda}^{m/2}(-ik\xi_<) W_{i\nu - i\lambda}^{m/2}(-ik\xi_>)\right] \times \left[\Gamma\left(\frac{m+1}{2} - i\lambda\right) M_{i\lambda}^{m/2}(-ik\eta_<) W_{i\lambda}^{m/2}(-ik\eta_>)\right]. \tag{3.17}$$

This can also be expressed in the form

$$G^{(+)}(1,2,k) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} e^{im(\phi_1 - \phi_2)} \left(-\frac{1}{2\pi i}\right) \int_{-\infty}^{\infty} dZ_2\, g_m^{(+)}(\xi,\xi',Z - Z_2) g_m^{(+)}(\eta,\eta',Z_2), \tag{3.18}$$

in which

$$g_m^{(+)}(x,x',Z_{1,2}) = (ikx)^{-1/2}(ikx')^{-1/2} \Gamma\left(\frac{m+1}{2} - i\nu_{1,2}\right) M_{i\nu_{1,2}}^{m/2}(-ikx_<) W_{i\nu_{1,2}}^{m/2}(-ikx_>) \quad (Z_{1,2} = k\nu_{1,2}). \tag{3.19}$$

The convolution integral in (3.17) or (3.18) is standard for Green's functions of separable operators. In the present case, a contour can be closed by an infinite semicircle in the lower half-plane such as to enclose the poles of $\Gamma((m+1)/2 - i\lambda)$ but exclude those of $\Gamma((m+1)/2 - i\nu + i\lambda)$. Equation (3.19) represents the Green's function for the differential equation (2.7) obtained after separation of variables, viz.,

$$\left(Z_1'^2 + \frac{\partial}{\partial x} x \frac{\partial}{\partial x} + \frac{k^2 x}{4} - \frac{m^2}{4x}\right) g_m^{(+)}(x,x',Z_1'^2) = \delta(x - x'). \tag{3.20}$$

Note that $Z_{1,2}$ rather than E now plays the role of eigenvalue. This formulation of the separated Schrödinger equation is convenient for treatment of the Stark effect.[16] Reduced Green's functions derived from (3.19) can provide an elegant alternative for computation of Stark-effect perturbation energies.

Returning now to Eq. (3.17), Buchholz's integral representation (2.29) can be applied to each factor ΓMW giving, after some rearrangement:

$$G^{(+)}(1,2,k) = \frac{ik}{4\pi^2} \int_{-\infty}^{\infty} d\lambda \int_0^{\infty} ds \int_0^{\infty} dt\, e^{-ik/2(\xi_1 + \xi_2)\cosh s} e^{-ik/2(\eta_1 + \eta_2)\cosh t} [\coth(s/2)]^{2i\nu - 2i\lambda} [\coth(t/2)]^{2i\lambda} \times \sum_{m=-\infty}^{\infty} e^{im(\phi_1 - \phi_2)} J_m(k(\xi_1\xi_2)^{1/2} \sinh s) J_m(-k(\eta_1\eta_2)^{1/2} \sinh t). \tag{3.21}$$

We have now been able to revert to the original parabolic coordinates $\xi_1,\eta_1,\xi_2,\eta_2$. In the sum over Bessel functions we have noted that $J_{-m}(z) = J_m(-z) = (-)^m J_m(z)$. The integral over λ gives a delta function:

$$\int_{-\infty}^{\infty} d\lambda\, [\coth(s/2)]^{-2i\lambda}\,[\coth(t/2)]^{2i\lambda}$$
$$= \pi\delta(\ln\coth(t/2) - \ln\coth(s/2))$$
$$= \pi \sinh s\, \delta(t-s). \tag{3.22}$$

The integral over t is thus immediate. The sum is in the form of Graf's addition theorem[17]:

$$J_0(r) = \sum_{m=-\infty}^{\infty} J_m(p) J_m(q) e^{im\phi},$$
$$r = (p^2 + q^2 - 2pq\cos\phi)^{1/2}, \tag{3.23}$$

where we identify

$$p = k(\xi_1\xi_2)^{1/2}\sinh s,$$
$$q = -k(\eta_1\eta_2)^{1/2}\sinh s, \quad \phi = \phi_1 - \phi_2,$$
$$r = k[\xi_1\xi_2 + \eta_1\eta_2 + 2(\xi_1\xi_2\eta_1\eta_2)^{1/2}\cos(\phi_1-\phi_2)]^{1/2}\sinh s. \tag{3.24}$$

The Green's function thus reduces to

$$G^{(+)}(1,2,k) = \frac{ik}{4\pi}\int_0^{\infty} ds \sinh s\, e^{iku\cosh s} \times J_0(ku\sinh s)[\coth(s/2)]^{2i\nu}, \tag{3.25}$$

where [cf. (2.1)]

$$v \equiv \tfrac{1}{2}(\xi_1+\xi_2+\eta_1+\eta_2) = r_1 + r_2 = \tfrac{1}{2}(x+y),$$
$$u \equiv [\xi_1\xi_2 + \eta_1\eta_2 + 2(\xi_1\xi_2\eta_1\eta_2)^{1/2}\cos(\phi_1-\phi_2)]^{1/2}$$
$$= (2\mathbf{r}_1\cdot\mathbf{r}_2 + 2r_1r_2)^{1/2} = (xy)^{1/2}, \tag{3.26}$$

in terms of the variables

$$x \equiv r_1 + r_2 + r_{12}, \quad y \equiv r_1 + r_2 - r_{12}. \tag{3.27}$$

To complete the derivation, we make use of the identity

$$J_0(ku\sinh s) = \frac{1}{ku\sinh s}\frac{\partial}{\partial u} uJ_1(ku\sinh s), \tag{3.28}$$

in conjunction with the integral representation (2.29) with $m = 1$. We obtain thereby

$$G^{(+)}(1,2,k) = \frac{1}{4\pi iku}\frac{\partial}{\partial u}\Gamma(1-i\nu)M_{i\nu}^{1/2}(-iky) \times W_{i\nu}^{1/2}(-ikx). \tag{3.29}$$

Noting, finally, that [cf. (3.26), (3.27)]

$$\frac{1}{u}\frac{\partial}{\partial u} = -\frac{2}{x-y}\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right) = -\frac{1}{r_{12}}\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right), \tag{3.30}$$

we obtain Hostler's expression for the Coulomb Green's function [2]

$$G^{(+)}(1,2,k) = -\frac{1}{4\pi ikr_{12}}\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right)\Gamma(1-i\nu) \times M_{i\nu}^{1/2}(-iky)W_{i\nu}^{1/2}(-ikx). \tag{3.31}$$

The Green's function (3.31) applies to both attractive and repulsive Coulomb interactions. Most generally one can redefine

$$\nu \equiv -ZZ'/k, \tag{3.32}$$

in which Z and Z' are the charges (in atomic units) of the interacting particles. For an electron interacting with a nucleus, the problem we have considered explicitly, $Z' = -1$.

4. RUTHERFORD SCATTERING LIMIT

The continuum eigenfunction (2.19) with quantum numbers $m = 0$, $\nu_1 = -i/2$, $\nu_2 = \nu + i/2$ ($\nu = -ZZ'/k$) represents scattering of an incident plane wave by a nucleus (Rutherford scattering). With use of (2.11) we obtain[18]:

$$\psi_R^{\mathrm{RUTH}}(\mathbf{r}) = e^{ik\xi/2}e^{-ik\eta/2}\,{}_1F_1(i\nu;1;ik\eta)$$
$$= e^{ikz}\,{}_1F_1(i\nu;1;ik\eta) \tag{4.1}$$

normalized such that $\psi(0) = 1$. It is of interest to obtain this Rutherford scattering limit in an alternative way, by reduction of the Coulomb Green's function. $G^{(+)}(\mathbf{R},\mathbf{r},k)$ can be interpreted as the amplitude at point $\mathbf{r}$ for scattering of a spherical wave originating at $\mathbf{R}$ by the nucleus at $\mathbf{r} = 0$. As $\mathbf{R}$ is moved to infinity along the negative z axis, the spherical wave approaches modified plane-wave behavior in the vicinity of the origin. The Rutherford scattering wave function (4.1) can thereby be represented as a limiting form of the Green's function as follows:

$$\psi_k^{\mathrm{RUTH}}(\mathbf{r}) = \lim_{R\to\infty}\frac{G^{(+)}(\mathbf{R},\mathbf{r},k)}{G^{(+)}(\mathbf{R},\mathbf{0},k)}. \tag{4.2}$$

The denominator $G^{(+)}(\mathbf{R},\mathbf{0},k)$ is obtained readily from (3.31) with $x = 2R$, $y = 0$. From the limiting forms of $M_{i\nu}^{1/2}(-iky)$ and its derivative[15] we obtain

$$G^{(+)}(\mathbf{R},\mathbf{0},k) = -\frac{1}{4\pi R}\Gamma(1-i\nu)W_{i\nu}^{1/2}(-2ikR). \tag{4.3}$$

To evaluate the limiting form of $G^{(+)}(\mathbf{R},\mathbf{r},k)$ we make use of the integral formula (3.17) with the following specialization of the parabolic coordinates:

$$\xi_< = \xi_1 = r_1 + z_1 = R - R = \epsilon \to 0,$$
$$\eta_> = \eta_1 = r_1 - z_1 = 2R \to \infty,$$
$$\xi_> = \xi_2 = \xi, \quad \eta_< = \eta_2 = \eta. \tag{4.4}$$

For the factor in (3.17) containing $\xi_< = \epsilon$[15]:

$$\lim_{\epsilon\to 0}\epsilon^{-1/2}M_{i\nu-i\nu_2}^{m/2}(-ik\epsilon) = (-ik)^{1/2}\delta_{m,0}, \tag{4.5}$$

which eliminates all but the $m = 0$ term of the summation. With use of (4.3)–(4.5), the wavefunction (4.2) reduces to

$$\psi_k^{\mathrm{RUTH}}(\mathbf{r}) = \lim_{R\to\infty}\frac{1}{2\pi}\int_{-\infty}^{\infty} d\lambda \frac{\Gamma(\frac{1}{2}-i\nu+i\lambda)\Gamma(\frac{1}{2}-i\lambda)}{\Gamma(1-i\nu)} \times \frac{M_{i\lambda}^0(-ik\eta)}{(-ik\eta)^{1/2}}\frac{W_{i\nu-i\lambda}^0(-ik\xi)}{(-ik\xi)^{1/2}} \times \left[(-2ikR)^{1/2}\frac{W_{i\lambda}^0(-2ikR)}{W_{i\nu}^0(-2ikR)}\right]. \tag{4.6}$$

From the asymptotic form of the W functions [cf. (2.27)], the bracket in (4.6) can be reduced to

$$(-2ikR)^{1/2-i\nu+i\lambda}. \tag{4.7}$$

Note that (4.7) approaches zero as $R\to\infty$ for $\mathrm{Im}\lambda > 0$. Thus the integral (4.6) can be evaluated by closing a contour with a semicircle in the upper half of the complex λ plane. The integrand, specifically $\Gamma(\frac{1}{2} - i\nu + i\lambda)$, possesses poles in the upper half-plane at the points where

$$\tfrac{1}{2} - i\nu + i\lambda = -n, \quad n = 0,1,2,\cdots. \tag{4.8}$$

However, the limit of the factor (4.7) as $R\to\infty$ will approach zero unless $n = 0$. Thus (4.6) reduces to $2\pi i$ times the residue at $\lambda = \nu + i/2$ ($i\lambda = i\nu + \frac{1}{2}$), viz,

$$\psi_k^{\mathrm{RUTH}}(\mathbf{r}) = \frac{M^0_{i\nu-1/2}(-ik\eta)}{(-ik\eta)^{1/2}} \frac{W^0_{1/2}(-ik\xi)}{(-ik\xi)^{1/2}}. \tag{4.9}$$

Now[19]

$$W^0_{1/2}(-ik\xi) = (-ik\xi)^{1/2} e^{ik\xi/2}, \tag{4.10}$$

while [cf. (2.11)]

$$M^0_{i\nu-1/2}(-ik\eta) = (-ik\eta)^{1/2} e^{-ik\eta/2}\, {}_1F_1(i\nu;1;ik\eta), \tag{4.11}$$

which results in the Rutherford eigenfunction (4.1).

The Rutherford scattering cross section follows from the asymptotic form (2.14):

$$\begin{aligned}\psi_k^{\mathrm{RUTH}}(\mathbf{r}) &= e^{ik\xi/2} \frac{M^0_{i\nu-1/2}(-ik\eta)}{(-ik\eta)^{1/2}} \\ &\sim e^{-\pi\nu/2} e^{ik\xi/2} \left[\frac{(k\eta)^{-i\nu}}{\Gamma(1-i\nu)} e^{-ik\eta/2} \right. \\ &\quad \left. - i \frac{(k\eta)^{i\nu-1}}{\Gamma(i\nu)} e^{ik\eta/2} \right] \\ &= \frac{e^{-\pi\nu/2}}{\Gamma(1-i\nu)} \left[e^{ikz - i\nu \log[k(r-z)]} \right. \\ &\quad \left. - \frac{i}{k} \frac{\Gamma(1-i\nu)}{\Gamma(i\nu)} \frac{e^{ikr + i\nu\log[k(r-z)]}}{(r-z)} \right].\end{aligned} \tag{4.12}$$

This corresponds to a scattering amplitude

$$f(\theta) = \frac{-i\Gamma(1-i\nu)}{k\Gamma(i\nu)(1-\cos\theta)}, \tag{4.13}$$

which leads to the famous Rutherford formula

$$\sigma(\theta) = |f(\theta)|^2 = \frac{\nu^2}{k^2(1-\cos\theta)^2} = \frac{Z^2 Z'^2 e^4}{16E^2} \csc^4\frac{\theta}{2}. \tag{4.14}$$

[1] J. Meixner, Math. Z. **36**, 677 (1933).
[2] L. Hostler, J. Math. Phys. **5**, 591 (1964).
[3] H. Buchholz, *The Confluent Hypergeometric Function* (Springer, New York, 1969), p. 86, Eq. (5c).
[4] H.A. Bethe and E.E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic, New York, 1957), p. 27.
[5] Ref. 3, pp. 11 *ff*. We follow throughout the notation of Buchholz except that we write, for compactness, $M_\kappa^{\mu/2}(z)$ in place of $\mathcal{M}_{\kappa,\mu/2}(z)$ and $W_\kappa^{\mu/2}(z)$ in place of $W_{\kappa,\mu/2}(z)$.
[6] Ref. 3, p. 91, Eq. (3).
[7] Ref. 3, p. 82, Eq. (1).
[8] G.N. Watson, *Theory of Bessel Functions*, 2nd. ed. (Cambridge U.P., Cambridge, 1966), p. 395, Eq. (1).
[9] Ref. 4, p. 29.
[10] Ref. 3, p. 214, Eq. (1a).
[11] Ref. 3, p. 19, Eq. (20a).
[12] Ref. 3, p. 19, Eq. (19).
[13] Ref. 3, p. 90, Eq. (1b).
[14] Ref. 3, p. 214, Eq. (1c).
[15] Ref. 3, p. 28.
[16] Ref. 4, p. 229 *ff*.
[17] Ref. 8, p. 359, Eq. (1).
[18] Ref. 4, p. 30.
[19] Ref. 3, p. 207.

On the Abel summability of partial wave amplitudes for Coulomb-type interactions

F. Gesztesy[a] and C. B. Lang
Institut für Theoretische Physik, Universität Graz, A-8010 Graz, Austria

(Received 19 November 1979; accepted for publication 7 December 1979)

We prove that the sum of partial wave amplitudes for Coulomb-type potentials [e.g., $V(r) = \gamma/r + O(r^{-2})$] is convergent in the Abel sense although it diverges in the ordinary sense. The method of Abel summation is a generalization of the ordinary summation and allows one to sum certain divergent series explicitly. It is closely connected with analytic continuation; with the help of optimal conformal mappings the convergence of the Abel sum (for long- and short-range interactions) can be improved substantially. This enables us to obtain values of the scattering amplitude for each scattering angle (except forward direction). In particular, we show that the screened scattering amplitude converges in the Abel sense up to a phase factor to the unscreened one if the screening is removed.

PACS numbers: 03.65.Nk, 02.30.Lt

I. MOTIVATION AND INTRODUCTION

The problem of whether a series obtained for a physical quantity is summable in the ordinary sense is a notorious one. The intention in all perturbative approaches is of course to get rapidly converging expressions that allow the determination of the result to arbitrary accuracy. Increasing complexity of the problems, however, often makes it hardly possible to make sure that the perturbative scheme applied is actually adequate for the specific case and does lead to a convergent series. A classical example for a divergent partial wave expansion series is that of the nonrelativistic Coulomb amplitude. The sum of the partial wave amplitudes, determined from the asymptotic phases of the partial waves, does not converge to the full Coulomb scattering amplitude in the ordinary sense. The partial sums diverge oscillatorily and have no limit.

On the other hands, naively one should expect that on the way to the series (i.e., by the manipulations required by the formal expansion algorithm) no information should be lost. That means that it should be possible to obtain the desired results even from the divergent series.[1] It is well known[2] by now[3] that there are techniques to obtain from certain divergent series finite values. From the abstract point of view this is quite clear since it is just a convention (i.e., a definition) how to give an infinite series a value, as long as the internal consistency of the postulates is guaranteed. The basic idea is to reconstruct from the series the value of a function that leads, when expanded, to the series in question. One thereby assumes that the divergent series is just the result of the polynomial expansion of a function with certain singularity structure applied outside its region of convergence—the result of a method of analytic continuation that is not appropriate for the problem. For asymptotic series additional assumptions on the function are necessary in order to guarantee uniqueness. For less divergent series other resummation (i.e., reconstruction) procedures are possible.

We discuss only the case of partial wave expansions of nonrelativistic scattering amplitudes of long and short range interactions. For Coulomb-type potentials [$V(r) = \gamma/r + O(r^{-2})$] the sum

$$\sum_l (2l+1)(e^{2i\delta_l}-1)P_l(x),$$

is divergent in the ordinary sense, the partial sums oscillate and grow without limit, and converge only in a distributional sense,[4] i.e., when first multiplied with a certain test function in $x = \cos\theta$ and integrated over x. For short-range potentials [$V(r) = O(r^{-3-\epsilon})$ as $r \to \infty$] the corresponding sum is even absolutely convergent in the ordinary sense but not the separate sums

$$\sum_l (2l+1)e^{2i\delta_l}P_l(x) \quad \text{and} \quad \sum_l (2l+1)P_l(x).$$

We show that all these sums do exist as Abel sums, i.e., sums in a modified sense. Abel summation[2,5,6] is a direct generalization of ordinary summation and allows the summation of the above-mentioned partial wave amplitudes for each $x \in [-1, 1]$. For ordinary convergent series Abel summation reduces to the ordinary one. We therefore propose to treat the sums of partial wave amplitudes as Abel sums.

The Abel sum can be considered as the Abel limit of the partial sums. The Abel limit again is a generalization of the ordinary limit and reduces to it when applied to quantities limitable in the ordinary sense. This Abel limit will be useful in demonstrating how the scattering amplitude for a screened Coulomb potential is related to the normal Coulomb amplitude.

In some respect Abel summability was already noted by Mott[7]—although he was not aware of Abel's work. For pure Coulomb scattering he reconstructed the correct amplitude from the divergent partial wave series in exactly the required way of Abel summation. We show how to use Abel summation in practical cases where no closed form of the result is accessible and where one accelerates convergence with conformal mapping techniques. In a certain sense the existence of the Abel sum (which we prove for Coulomb-type interac-

[a]Supported by Fonds zur Förderung der Wissenschaftlichen Forschung in Österreich.

 0022-2488/81/020312-08$1.00

tions) justifies *a posteriori* other practical techniques to obtain from the divergent series a convergent one, e.g., multiplying with a singularity reducing function in x which leads to a resummation,[8] "punctual" Padé approximation,[9] or the Legendre Padé approximation.[10]

II. THE SCATTERING AMPLITUDE FOR COULOMB-TYPE POTENTIALS

In this section let us briefly review what is known about nonrelativistic scattering amplitudes for spherically symmetric Coulomb-type potentials. We discuss first the case of pure Coulomb interactions and denote the potential by

$$V^c(r) = \gamma/r, \quad \gamma \in R, \quad e = \hbar = 2m = 1. \tag{2.1}$$

The first rigorous approach for this type of long-range potentials within the framework of time-dependent scattering theory was initiated by Dollard[11] and relied on an appropriate modification of the free time evolution operator (cf. also Ref. 12). It can be shown that Dollard's scattering operator actually coincides with the usual Coulomb S matrix.[13] An alternative description of the Coulomb scattering operator S^c, based on an algebraic approach to scattering theory by Combes[14] and Lavine[15] (cf. also Ref 16), has been adoped by Grosse *et al.*[17], exploiting the SO(3,1) symmetry of the problem. As a result S^c in the interaction picture can be written[17,18]

$$S^c = \frac{\Gamma\left(\tfrac{1}{2} + (L^2 + \tfrac{1}{4})^{1/2} + i\gamma/2\sqrt{-\Delta}\,\right)}{\Gamma\left(\tfrac{1}{2} + (L^2 + \tfrac{1}{4})^{1/2} - i\gamma/2\sqrt{-\Delta}\,\right)}, \tag{2.2}$$

where $-\Delta$ denotes the kinetic energy operator and L^2 the square of the angular momentum operator.

In momentum space S^c acts on states $\Phi \in C^1(R^3) \cap L^2(R^3)$ as

$$\begin{aligned}(S^c\Phi)(k) = \lim_{\epsilon \to 0_+} & \frac{\gamma}{2\pi i|k|} \frac{\Gamma(1 + i\gamma/2|k|)}{\Gamma(1 - i\gamma/2|k|)} \\ & \times \int_{R^3} d^3k'\, \delta(|k|^2 - |k'|^2) \\ & \times \left[\frac{4|k|^2}{(k-k')^2}\right]^{1 - \epsilon + i\gamma/2|k|} \Phi(k').\end{aligned} \tag{2.3}$$

For $\Phi_l(k) = Y_l^m(k/|k|)\phi(|k|)$, Eq. (2.3) implies

$$(S^c\Phi_l)(k) = \exp(2i\delta_l^c(|k|))\Phi_l(k), \tag{2.4}$$

where $\delta_l^c(|k|)$ are the usual Coulomb phase shifts

$$\delta_l^c(|k|) = \arg\Gamma(1 + l + i\gamma/2|k|). \tag{2.5}$$

Further on we abbreviate $x = \cos\theta = k\cdot k'/|k||k'|$, where θ denotes the scattering angle and, without loss of generality, put the incident momentum $|k| = 1/2$.

The Coulomb scattering amplitude[19] then reads

$$f^c(x) = -\gamma 2^{1+i\gamma} \frac{\Gamma(1+i\gamma)}{\Gamma(1-i\gamma)} (1-x)^{-1-i\gamma}, \quad -1 \leqslant x < 1, \tag{2.6}$$

and Eqs. (2.3)–(2.5) imply

$$f^c(x) = \frac{1}{i} \sum_{l=0}^{\infty} (2l+1)(e^{2i\delta_l} - 1) P_l(x), \tag{2.7}$$

in the sense of distributions[4,20,21]; this means that both sides have to be integrated with a twice continuously differentiable test function $\psi(x)$ vanishing at $x = 1$ in order to obtain convergence of the sum over l. Considering $f^c(x)$ as a mere function the corresponding expansion into a Legendre series is divergent for all x. This is of course common to amplitudes with certain singularities at $x = 1$. A Legendre series for a function converges pointwise in the interior of the largest ellipse with foci ± 1 where it is holomorphic, but is divergent outside.[22] On the boundary there may be regions of conditional convergence depending on the order of the singularity. Since $f^c(x)$ has a pole (and a branch point) at $x = 1$, the sum in Eq. (2.7) diverges in the ordinary sense for all $x \in C$.

However, one may expand $f^c(x)$ in a Taylor series at, for example, $x = 0$ or $x = -1$ with radii of convergence 1 or 2, respectively. Alternatively, introducing the conformal variable

$$w = \frac{1 - (1-x)^{1/2}}{1 + (1-x)^{1/2}}, \tag{2.8}$$

which maps the x plane, cut from 1 to infinity, into the interior of the unit circle in the w plane, produces an optimal converging series in w^n. (For a review on this type of techniques see Ref. 23.)

Let us now discuss the general case of Coulomb-type potentials

$$V(r) = V^c(r) + V^s(r), \tag{2.9}$$

where V^s denotes the short range part of the potential and fulfills

$$\int_0^\infty dr\, r|V^s(r)| < \infty. \tag{2.10}$$

If one decomposes the total phase shift δ_l, which is defined in terms of the asymptotic form of the radial scattering wave function, into

$$\delta_l = \delta_l^c + \mu_l, \tag{2.11}$$

and assumes that

$$V^s(r) = O(r^{-3-\epsilon}), \quad \epsilon > 0 \quad \text{as} \quad r \to \infty, \tag{2.12}$$

then a result by Semon and Taylor[24] can be stated as follows: The sums in

$$\begin{aligned} f(x) &= \frac{1}{i} \sum_{l=0}^{\infty} (2l+1)(e^{2i\delta_l} - 1) P_l(x) \\ &= f^c(x) + \frac{1}{i} \sum_{l=0}^{\infty} (2l+1) e^{2i\delta_l^c}(e^{2i\mu_l} - 1) P_l(x), \end{aligned} \tag{2.13}$$

are both convergent in the sense of distributions. From Eq. (2.12) one concludes[25]

$$|e^{2i\mu_l} - 1| = O(l^{-2-\epsilon}), \tag{2.14}$$

and thus the second sum in Eq. (2.13) is not only convergent in the sense of distributions, but converges uniformly and absolutely for all $x \in [-1, 1]$.

Another result by Semon and Taylor[24] concernes screened Coulomb potentials. Let

$$V^R(r) = V^c(r)\alpha^R(r) + V^s(r), \tag{2.15}$$

where the screening function $\alpha^R(r)$ obeys[4,24]

$O \leqslant \alpha^R(r) \leqslant 1$, for R fixed, $\alpha^R(r) \overset{r\to\infty}{\to} 0$ monotonically like $O(r^{-2-\epsilon})$ for some $\epsilon > 0$, for r fixed, $\alpha^R(r) \overset{R\to\infty}{\to} 1$, (2.16)

and $V^s(r)$ obeys Eq. (2.12). Using the method of variable phase[26] the authors show that for large values of the screening radius R the phase shift δ_l^R that corresponds to the short range potential $V^R(r)$ behaves like

$$\delta_l^R - \omega(R) \overset{R\to\infty}{\to} \delta_l, \tag{2.17}$$

where

$$\omega(R) = -\gamma \int_1^\infty dr\, r^{-1} \alpha^R(r). \tag{2.18}$$

Since $\omega(R)$ does not depend on l, this shows that $f_R(x)$, the scattering amplitude corresponding to $V^R(r)$, converges in the distributional sense to $f(x)$ up to a phase factor[24]

$$\exp[-2i\omega(R)]\, f_R(x) \overset{R\to\infty}{\to} f(x) \quad \text{(distributional)}. \tag{2.19}$$

Considered as a function, $\exp[-2i\omega(R)] f_R(x)$ does not converge to $f(x)$ in general.[27] For an alternative approach to screening where the screened Møller operators converge to the unscreened ones (up to R-dependent phase factors) if the screening is removed see Ref. 28.

III. ABEL SUMMABILITY OF SCATTERING AMPLITUDES

In this section we first introduce the concept of Abel summability and then we prove that the partial wave expansion of the scattering amplitude for Coulomb-type potentials, although divergent in the ordinary sense, has a well defined Abel sum.

A sum over a sequence of complex numbers $\Sigma_{n=0}^\infty c_n$ is called convergent in the ordinary sense (or short: convergent) with sum C if and only if

$$\lim_{N\to\infty} \Big| \sum_{n=0}^N c_n - C \Big| = 0.$$

Then we write $\Sigma_{n=0}^\infty c_n = C$.

Let us now introduce the more general concept of an Abel summable series.[2,6] Let $\{c_n\}_{n=0}^\infty$ be a sequence of complex numbers and suppose that the power series $\Sigma_{n=0}^\infty c_n y^n$ is converging at least for $|y| < 1$ and define $f(y) = \Sigma_{n=0}^\infty c_n y^n$. Then $\Sigma_{n=0}^\infty c_n$ is called Abel convergent (Abel summable) with Abel sum C if and only if

$$\lim_{y\uparrow 1} |f(y) - C| = 0.$$

In this case we write $\text{A-}\Sigma_{n=0}^\infty c_n = C$. The Abel sum therefore is the value of a function which is known only from its power series. This value can be obtained by analytic continuation to $y = 1$ from the values of the function inside the unit disc. Since the power series (the asymptotic behavior of the c_n) allows one to determine the position and type of the closest singularities, the analytic continuation can be performed, for example, by a change of variables through a conformal mapping.

Abel's limit point theorem[6] implies that the Abel summation fulfills the so-called consistency condition, i.e., if a sum exists in the ordinary sense then the Abel sum exists with the same value too:

$$\sum_{n=0}^\infty c_n = C \quad \text{implies} \quad \text{A-}\sum_{n=0}^\infty c_n = C.$$

Obviously, this generalized summation procedure is, similar to other related ones like Padé or Borel summation, closely linked to the problem of analytic continuation. Whereas $f(y)$ in the case of Abel summation should be holomorphic in the unit disc, this is not necessary for Borel summation, which applies to asymptotic series. There, however, certain supplementary conditions need to be fulfilled in order to guarantee uniqueness of the sum.[2,3] We also note the great importance of such generalized summations in the theory of Fourier series (Fejér's theorem).[6]

Before we state our main result concerning Abel summability of the partial wave expansion of Coulomb-type scattering amplitudes we introduce a somewhat more general class of potentials including $1/r^2$ potentials. Let

$$V(r) = \frac{\gamma}{r} + \frac{\alpha - 1/4}{r^2} + V^s(r), \quad \gamma \in R, \quad \alpha \geqslant 0, \tag{3.1}$$

with

$$\int_0^\infty dr\, r |V^s(r)| < \infty. \tag{3.2}$$

The total phase shift[29] corresponding to $V(r)$ is denoted by δ_l.

Then the scattering amplitude $f(x)$ defined by the Abel sum

$$f(x) = \text{A-}\sum_{l=0}^\infty \frac{1}{i} (2l+1)(e^{2i\delta_l} - 1) P_l(x), \quad -1 \leqslant x < 1, \tag{3.3}$$

exists. This means that, although the partial wave expansion for Coulomb-type scattering amplitudes diverges when summed in the ordinary sense, it is convergent when summed as an Abel series.

To prove this we show that

$$f(x,y) = \sum_{l=0}^\infty \frac{1}{i} (2l+1)(e^{2i\delta_l} - 1) P_l(x) y^l \equiv \sum_{l=0}^\infty c_l(x) y^l, \tag{3.4}$$

has at least radius of convergence one in the y plane and that the limit $\lim_{y\uparrow 1} f(x,y)$ exists for $x \in [-1, 1)$.

We decompose

$$\delta_l = \sigma_l + \mu_l, \tag{3.5}$$

where

$$\sigma_l = \arg\{\Gamma[\tfrac{1}{2} + (l^2 + l + \alpha)^{1/2} + i\gamma]\} + [l + \tfrac{1}{2} - (l^2 + l + \alpha)^{1/2}]\pi/2, \tag{3.6}$$

is the phase shift[30] associated with the potential

$$\frac{\gamma}{r} + \frac{\alpha - 1/4}{r^2}, \quad \gamma \in R, \quad \alpha \geqslant 0, \tag{3.7}$$

which of course reduces to the pure Coulomb potential for $\alpha = 1/4$. We have to study the asymptotic behavior of $c_l(x)$ for $|x| < 1$. For this purpose we note[31]

$$P_l(x) = 2(2\pi l)^{-1/2}(1-x^2)^{-1/4} \times \cos\left[(l+\tfrac{1}{2})\arccos x - \frac{\pi}{4}\right](1+O(l^{-1})), \quad |x|<1, \tag{3.8}$$

and

$$\exp(2i\sigma_l) = l^{2i\gamma}(1+O(l^{-1})). \tag{3.9}$$

If $V^s(r) = O(r^{-\beta})$ with $\beta > 2$ as $r \to \infty$, then[24,25]

$$\exp(2i\mu_l) = 1 + O(l^{1-\beta}), \tag{3.10}$$

and finally we obtain asymptotically

$$c_l(x) = c(x) l^{1/2}(l^{2i\gamma}-1) \times \exp(\pm il\arccos x)(1+O(l^{-1})), \quad |x|<1, \tag{3.11}$$

where $c(x)$ depends only on x.

From this asymptotic behavior one may determine type and position of the nearest singularities. If one expands $(y+a)^{-b-1}$ in a Taylor series

$$(y+a)^{-b-1} = \sum_{l=0}^{\infty} d_l y^l,$$

one gets asymptotically

$$d_l = (a^{b+1}\Gamma(b+1))^{-1}(-a)^{-l} l^b (1+O(l^{-1})). \tag{3.12}$$

By comparison of Eqs. (3.11) and (3.12) we conclude that $f(x,y)$, $|x|<1$, has singularities in the y plane of the type

$$(y - \exp(\pm i\arccos x))^{-3/2-2i\gamma}, \quad |x|<1. \tag{3.13}$$

The same analysis performed for $x = -1$ shows that $f(-1, y)$ has a singularity of the type

$$(y+1)^{-2-2i\gamma}. \tag{3.14}$$

Note that only the kind but not the position of the singularity in the y plane is affected by $\exp(2i\delta_l)$. Its position $\exp(\pm i\arccos x)$ is entirely determined by $P_l(x)$ [see also Eq. (3.21)] and is not changed by higher order terms in Eq. (3.8).[31]

Equations (3.13) and (3.14) show that $f(x,y)$ converges for $|y|<1$ and can be analytically continued to $y=1$ for $-1 \leqslant x < 1$. This proves the existence of the Abel sum of partial wave amplitudes

$$f(x) = \lim_{y\to 1-} f(x,y) = \text{A-}\sum_{l=0}^{\infty} \frac{1}{i}(2l+1)(e^{2i\delta_l}-1)P_l(x), \quad -1 \leqslant x < 1. \tag{3.15}$$

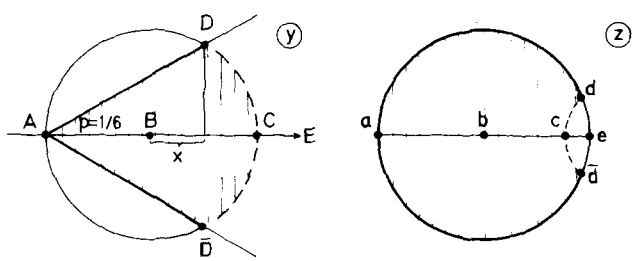


FIG. 1. This figure exhibits the mapping $y \to z$ [Eq. (3.16)] for a special value of $p = 1/6$. The points in the y plane denoted by capital letters are mapped to the z plane points denoted by the corresponding small letters. If the singularities closest to the origin B are the points D and $\bar{D}$, then an expansion in polynomials in z will be convergent at the point c in the z plane. This practically allows the analytic continuation to $y=1$ (point C) required in the Abel summation.

It remains to show how the continuation may be done in order to obtain numerical results. We introduce a conformal mapping $y \to z$ (cf. Fig. 1):

$$z = \frac{(y+1)^{1/2p}-1}{(y+1)^{1/2p}+1}, \quad y = \left(\frac{1+z}{1-z}\right)^{2p} - 1,$$
$$\left(y = -1,0,1,\infty \to z = -1,0,\frac{2^{1/2p}-1}{2^{1/2p}+1},1\right). \tag{3.16}$$

This mapping maps the interior of the sector generated by the points $y = -1$, $y = \exp(\pm i\arccos x)$ with apex $y = -1$ in the y plane into the interior of the unit circle in the z plane (the opening angle of the sector is $2\pi p = \arccos x$).

Writing

$$f(x,y) = \sum_{l=0}^{\infty} c_l(x) y^l = \sum_{l=0}^{\infty} d_l(x) z^l, \quad |y|<1, \tag{3.17}$$

one infers

$$d_0(x) = c_0(x) = \frac{1}{i}(e^{2i\delta_0}-1),$$

$$d_l(x) = \sum_{m=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{j} \frac{p^m}{m!}\binom{l+m-1}{2m-1}\binom{j}{k}(-1)^{j-k} \times k\left(k-\frac{1}{p}\right)\cdots\left(k-\frac{m-1}{p}\right)c_j(x), \quad l = 1,2,3,\cdots. \tag{3.18}$$

The conformal mapping was chosen in such a way (i.e., $y=0$ mapped to $z=0$) that each $d_l(x)$ can be evaluated from the knowledge of $c_k(x)$, $k \leqslant l$, i.e., from the knowledge of $\delta_0,\ldots,\delta_l$.

Therefore it is now possible to determine the Abel sum $f(x,1)$ from Eq. (3.17) by evaluating the ordinary convergent sum over z^l at $z(y=1)$,

$$f(x) = \sum_{l=0}^{\infty} d_l(x)\left(\frac{2^{1/2p}-1}{2^{1/2p}+1}\right)^l. \tag{3.17'}$$

For numerical results compare Sec. IV.[32]

Mapping techniques like the one discussed are necessary for Abel summation of Coulomb-type partial wave amplitudes. For short-range potentials [$V(r) = O(r^{-3-\epsilon})$, $\epsilon > 0$, as $r \to \infty$] where the partial wave expansion converges already in the ordinary sense one may also use these conformal mappings in order to accelerate the convergence. From the consistency condition discussed before we conclude that the natural way of summing partial wave expansions for scattering amplitudes is to perform Abel summations. For slowly convergent partial wave series it is possible to use rational approximants[9] to determine the value of the Abel sum at $y=1$. For functions with singularities at points $|y|=1$ like the Coulomb partial waves sum it is not clear whether analytic continuation with Padé approximants (that approximate the cut structure by poles) is adequate.

Let us discuss two examples for which the Abel sum may be evaluated in closed form. We start with

$$\delta(x,y) = \sum_{l=0}^{\infty} \frac{1}{i}(2l+1)(-1)\,P_l(x)y^l, \quad |x|\leqslant 1, \quad |y|<1. \tag{3.19}$$

with the help of the generating function of the Legendre

polynomials

$$(y^2-2xy+1)^{-1/2}=\sum_{l=0}^{\infty}P_l(x)y^l,$$
$$|x|\leqslant 1,\quad |y|<1, \tag{3.20}$$

we obtain

$$\delta(x,y)=i\left[2y\frac{d}{dy}(y^2-2xy+1)^{-1/2}+(y^2-2xy+1)^{-1/2}\right]$$
$$=i(1-y^2)(y^2-2xy+1)^{-3/2}. \tag{3.21}$$

Thus,

$$A\text{-}\sum_{l=0}^{\infty}(2l+1)P_l(x)$$
$$=\lim_{y\to 1}\delta(x,y)=\begin{cases}0, & x\neq 1,\\ \infty, & x=1.\end{cases} \tag{3.22}$$

Remembering the closure relation for $P_l(x)$, i.e.,

$$\sum_{l=0}^{\infty}(l+\tfrac{1}{2})P_l(x)P_l(x')=\delta(x-x'),$$

in the sense of distributions, the result Eq. (3.22) looks familiar $[P_l(1)=1]$.

Equation (3.22) shows that one may use

$$\hat{c}_l(x)=\frac{1}{i}(2l+1)\,e^{2i\delta_l}P_l(x) \tag{3.23}$$

instead of $c_l(x)$ in order to evaluate the Abel sum of the scattering amplitude $f(x)$ for $x\neq 1$. It also shows that for ordinary convergent partial wave amplitudes the sum over $(2l+1)e^{2i\delta_l}P_l(x)$ can only be Abel convergent. For these the reduction (3.23) is of course not advisable.

The second example concerns the pure Coulomb scattering amplitude $f^c(x)$; it has been evaluated by Mott[7] who apparently was unaware of Abel's work. His procedure is actually Abel summation of the Coulomb partial wave amplitudes but seems to have been overlooked in most publications on that subject. His method is based on the formula[31]

$$\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}=\frac{-\exp[-i\pi(x+y)]}{4\sin\pi x\sin\pi y}$$
$$\times\oint_{\Gamma}dt\,t^{x-1}(1-t)^{y-1},\qquad x,y \text{ not integer}. \tag{3.24}$$

Here the closed path Γ starts from a point on the real t axis between 0 and 1, encircles $t=1$ in the positive sense (counterclockwise), $t=0$ in the positive sense, then $t=1$ in the negative sense, and $t=0$ in the negative sense, before it returns to the starting points. Thus,

$$\hat{f}^c(x,y)$$
$$=\sum_{l=0}^{\infty}\frac{1}{i}(2l+1)\frac{\Gamma(1+l+i\gamma)}{\Gamma(1+l-i\gamma)}P_l(x)y^l$$
$$=\sum_{l=0}^{\infty}\frac{1}{i}(2l+1)[\Gamma(-2i\gamma)(1-\exp(4\pi\gamma))$$
$$\times(1-\exp(-2\pi\gamma))]^{-1}$$
$$\times\oint_{\Gamma}dt\,t^{l+i\gamma}(1-t)^{-2i\gamma-1}P_l(x)y^l$$
$$=-[\Gamma(-2i\gamma)(1-\exp(4\pi\gamma))(1-\exp(-2\pi\gamma))]^{-1}$$
$$\times\oint_{\Gamma}dt\,t^{i\gamma}(1-t)^{-2i\gamma-1}\delta(x,yt),\quad |x|\leqslant 1,\quad |y|<1, \tag{3.25}$$

and

$$f^c(x)=A\text{-}\sum_{l=0}^{\infty}\frac{1}{i}(2l+1)\frac{\Gamma(1+l+i\gamma)}{\Gamma(1+l-i\gamma)}P_l(x)$$
$$=[\Gamma(-2i\gamma)(1-\exp(4\pi\gamma))(1-\exp(-2\pi\gamma))]^{-1}$$
$$\times\oint_{\Gamma}dt\,t^{i\gamma}(1-t)^{-2i\gamma-1}\delta(x,t),\quad x\neq 1. \tag{3.26}$$

The last integral can be evaluated explicitly,[7] yielding exactly the result (2.6)

Finally, let us turn to screened Coulomb potentials and prove that the screened Coulomb-type scattering amplitudes converge to the unscreened ones (up to certain phase factors). We first introduce the concept of an Abel limit.[2,6]

A sequence of complex numbers $\{d_n\}_{n=0}^{\infty}$ is called convergent in the Abel sense with Abel limit D if and only if the sum $\Sigma_{n=m}^{\infty}d_ny^n$ (for any $m\geqslant 0$) converges for at least $|y|<1$ and

$$\lim_{y\to 1}(1-y)\sum_{n=m}^{\infty}d_ny^n=D. \tag{3.27}$$

Then we write $A\text{-}\lim_{n\to\infty}d_n=D$. It is obvious that a sequence limitable in the ordinary sense is also Abel limitable to the same limit. It can be easily shown that the Abel sum is just the Abel limit of the partial sums

$$A\text{-}\lim_{N\to\infty}\sum_{n=0}^{N}c_n=A\text{-}\sum_{n=0}^{\infty}c_n. \tag{3.28}$$

Let

$$V(r)=\frac{\gamma}{r}+\frac{\alpha-1/4}{r^2}+V^s(r),\quad \gamma\in R,\quad \alpha\geqslant 0,$$

and let Eqs. (2.12) and (3.2) be valid for $V^s(r)$. We then introduce the screened potential

$$V^n(r)=\left(\frac{\gamma}{r}+\frac{\alpha-1/4}{r^2}\right)\alpha^n(r)+V^s(r),$$
$$n=1,2,3,\cdots, \tag{3.29}$$

where $\alpha^n(r)$ fulfills Eqs. (2.16).

If $f(x)$ and $f_n(x)$ denote the scattering amplitudes for $V(r)$ and $V^n(r)$, respectively, then we can prove the following analog to Eq. (2.19). Up to a factor of modulus one, the Abel limit of $f_n(x)$ as n tends to infinity equals $f(x)$:

$$A\text{-}\lim_{n\to\infty}\exp[-2i\omega(n)]\,f_n(x)=f(x),\quad x\neq 1, \tag{3.30}$$

where $\omega(n)$ is defined in Eq. (2.18).

Before proving this let us discuss the special situation of a sharp cutoff function

$$\alpha^n(r)=\theta(n-r),\quad n=1,2,3,\cdots. \tag{3.31}$$

This case was analyzed by Semon and Taylor[27] using Born and Eikonal approximations for the pure Coulomb potential ($\alpha=1/4$, $V^s=0$). In the Born approximation $f^c(x)$ and $f_n^c(x)$ are given by

$$f^{c,B}(x)=\frac{2\gamma}{1-x},$$
$$f_n^{c,B}(x)=\frac{2\gamma}{1-x}\{1+\cos[n((1-x)/2)^{1/2}]\}, \tag{3.32}$$

and obviously $f_n^{c,B}(x)$ does not approach $f^{c,B}(x)$ in the ordinary sense as $n\to\infty$. The same holds in the Eikonal approximation.[27] It is clear from the Riemann–Lebesgue Lemma[33] that, when smeared with test functions, the screened amplitudes converge to the unscreened ones, i.e., distributional convergence[4,24] holds. The same is however true if we take the Abel limit instead of an ordinary limit. In fact,

$$\begin{aligned}\text{A-}\lim_{n\to\infty}\cos(nz) &= \lim_{y\to 1-}(1-y)\sum_{n=0}^{\infty}y^n\cos(nz)\\ &= \lim_{y\to 1-}(1-y)\frac{1-y\cos z}{1-2y\cos z+y^2}=0,\\ &\quad \text{for } z\neq 2k\pi,\ k=0,\pm 1,\pm 2,\cdots,\end{aligned}\tag{3.33}$$

and thus the screened amplitude converges in this example in the Abel sense to the unscreened one.

To prove the general case in Eq. (3.30) we write

$$\begin{aligned}&\text{A-}\lim_{n\to\infty}\exp[-2i\omega(n)]f_n(x)\\ &\quad=\lim_{y\to 1-}(1-y)\sum_{n=1}^{\infty}e^{-2i\omega(n)}f_n(x)y^n\\ &\quad=\lim_{y\to 1-}(1-y)\sum_{n=1}^{\infty}y^n e^{-2i\omega(n)}\\ &\quad\times\lim_{z\to 1-}\sum_{l=0}^{\infty}\frac{1}{i}(2l+1)(e^{2i\delta_l^n}-1)P_l(x)z^l\\ &\quad=\lim_{y\to 1-}\lim_{z\to 1-}(1-y)\sum_{n=1}^{\infty}y^n e^{-2i\omega(n)}\\ &\qquad\times\sum_{l=0}^{\infty}\frac{1}{i}(2l+1)e^{2i\delta_l^n}P_l(x)z^l\\ &\quad=\lim_{y\to 1-}\lim_{z\to 1-}\sum_{l=0}^{\infty}\frac{1}{i}(2l+1)e^{2i\delta_l}P_l(x)z^l\\ &\qquad\times(1-y)\sum_{n=1}^{\infty}\exp[2i(\delta_l^n-\omega(n)-\delta_l)]y^n,\quad x\neq 1.\end{aligned}\tag{3.34}$$

Here we interchanged Σ_n and $\lim_{z\to 1-}$ with the help of bound (2.14) and Weierstrass criterion. For $x\neq 1$ the Abel sum (3.19) vanishes and therefore the contribution from the (-1) in the sum over l can be dropped. We subsequently interchanged Σ_n and Σ_l since $|y|<1$ and $|z|<1$. From[4,24]

$$\lim_{n\to\infty}(\delta_l^n-\omega(n))=\delta_l,\tag{3.35}$$

and therefore

$$\begin{aligned}&\text{A-}\lim_{n\to\infty}\exp[2i(\delta_l^n-\omega(n)-\delta_l)]\\ &\quad=\lim_{y\to 1-}(1-y)\sum_{n=1}^{\infty}y^n\exp[2i(\delta_l^n-\omega(n)-\delta_l)]=1,\end{aligned}\tag{3.36}$$

we finally obtain the desired result

$$\begin{aligned}&\text{A-}\lim_{n\to\infty}\exp(-2i\omega(n))f_n(x)\\ &\quad=\lim_{z\to 1-}\sum_{l=0}^{\infty}\frac{1}{i}(2l+1)e^{2i\delta_l}P_l(x)z^l\\ &\quad=\text{A-}\sum_{l=0}\frac{1}{i}(2l+1)e^{2i\delta_l}P_l(x)=f(x),\quad x\neq 1.\end{aligned}\tag{3.37}$$

This means that the specific form of the screening function $\alpha^n(r)$ is irrelevant for the differential cross section if the Abel limit $n\to\infty$ is applied.

IV. NUMERICAL ILLUSTRATIONS AND CONCLUSION

Having proved that Abel summation is appropriate for the partial wave expansions of scattering amplitudes for long- and short-range interactions, we now discuss a few numerical examples. For all our examples we choose a typical value of the Coulomb parameter $\gamma=0.1$. In the figures we plot only the real parts of the quantities for convenience.

In Fig. 2 we compare the partial sums of the diverging partial wave summation for pure Coulomb scattering ($\alpha=1/4$) and $x=-1$

$$f^K(x=-1)=\frac{1}{i}\sum_{l=0}^{K}(2l+1)\left[\frac{\Gamma(1+l+i\gamma)}{\Gamma(1+l-i\gamma)}-1\right](-1)^l\tag{4.1}$$

with the partial sums of the converging Abel summation in Eq. (3.17):

$$f^K(x=-1,y=1)=\sum_{l=0}^{K}d_l(x=-1)\,(z(y=1))^l,\tag{4.2}$$

which converge for $K\to\infty$ to the Abel sum of the partial wave amplitudes. For $x=-1$ we have $p=1$ and $z(y=1)=(\sqrt{2}-1)/(\sqrt{2}+1)=0.171573$ for the mapping (3.16). The correct value of the Coulomb scattering amplitude (2.6) at $x=-1$ is $(-0.099344+0.011439\,i)$ which is obtained with four digit accuracy from the 10th partial sum onwards (six digits for the 15th partial sum). An equivalent figure can be drawn for the case of $V(r)=\gamma/r+(\alpha-\frac{1}{4})/r^2$, where the phase shifts are explicitly known from Eq. (3.6); the convergence properties are very similar.

The rate of convergence is satisfactory for $x\lesssim 0$ but gets worse as x approaches one. In Fig. 3 we give the partial sum

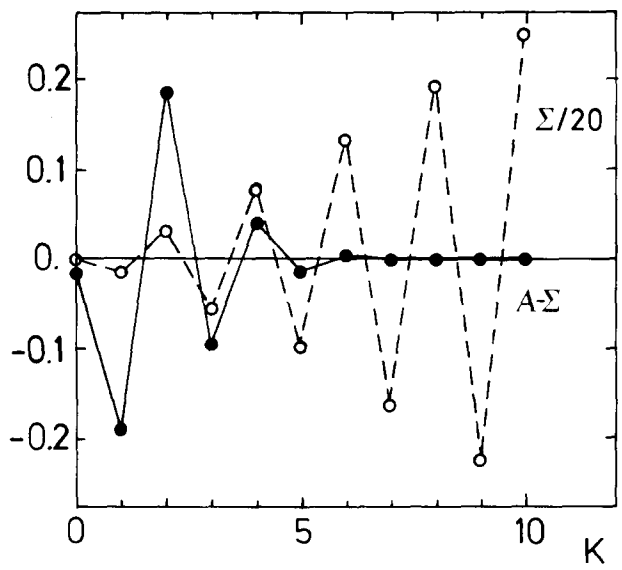


FIG. 2. We compare the difference between the correct value of the real part of the Coulomb amplitude and the partial sums for the ordinary partial wave summation [denoted by $\Sigma\equiv f^K(x=-1)-f^c(x=-1)$, cf. Eq. (4.1)] and the partial Abel sums [denoted by $\text{A-}\Sigma\equiv f^K(x=-1,\ y=1)-f^c(x=-1)$, cf. Eq. (4.2)]. The normal partial sums diverge oscillatorily and are therefore scaled down by a factor of 20; the partial Abel sums converge quickly and are correct to four decimals for $K>10$.

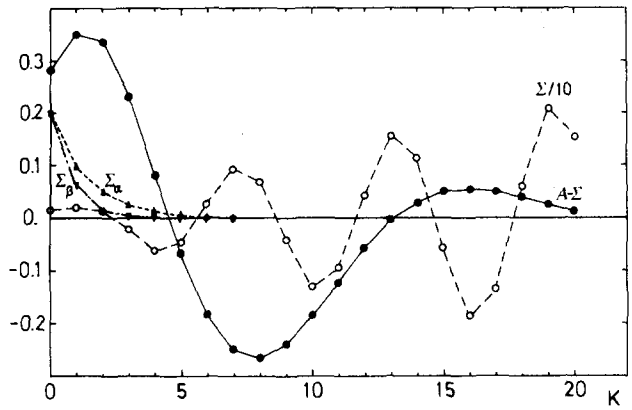


FIG. 3. For $x = 0.5$ the convergence of the Abel sum is slow; the difference between the correct value and the divergent ordinary partial sums is denoted by $\Sigma \equiv f^K(x = 0.5) - f^c(x = 0.5)$ and is scaled down by a factor of 10 in the figure. The difference between the correct value and the partial Abel sums is denoted by $\text{A-}\Sigma \equiv f^K(x = 0.5, y = 1) - f^c(x = 0.5)$. The convergence of the sum can be improved substantially if one sums the series in x' [Eq. (4.3)], where the expansion coefficients α_l are themselves obtained from Abel sums. The difference between the correct value $f^c(x = 0.5)$ and the partial sums in x' is denoted by Σ_α and vanishes quickly. Even faster convergence is achieved if one sums the corresponding series in $w(x)^l$, the optimal conformal variable. The expansion coefficients β_l can be computed from the α_l. The difference between the correct value and the partial sums is denoted by Σ_β. At $K = 7$ the partial sum in w^l is already correct for six digits. (Only the real parts are plotted.)

(4.1) and the partial Abel sum (4.2) for pure Coulomb scattering at $x = 0.5$; the corresponding values for p and z are $p = 1/6$ and $z = 7/9$. The correct value of the scattering amplitude is $(-0.399885 - 0.009592\,i)$ and the deviation of the 20th partial Abel sum is still about 5%.

A way out of the convergence problems for $x \gtrsim 0$ is offered by the expansion in x^l or even $(w(x))^l$ [cf. Eq. (2.8)] rather than $P_l(x)$. Although this is a technical detail, let us discuss this approach since it is helpful in practical summations. For simplicity we treat only the pure Coulomb potential, but the formulas derived may be easily generalized.

For this purpose we expand $\hat{f}^c(x, y)$ [cf. Eq. (3.23)] by expanding the Legendre polynomials[31] into a series in x^l.

$$\hat{f}^c(x,y) = \sum_{l=0}^{\infty} \frac{1}{i}(2l+1)\frac{\Gamma(1+l+i\gamma)}{\Gamma(1+l-i\gamma)} P_l(x) y^l$$
$$= \sum_{l=0}^{\infty} \alpha_l(y) x^l, \quad -1 \leqslant x < 1, \quad |y| < 1,$$

$$\alpha_l(v = y^2) = \frac{(2y)^l}{\Gamma(l+1)\Gamma(1/2)} \sum_{k=0}^{\infty} (-1)^k (4k + 2l + 1) \times \frac{\Gamma(2k+l+1+i\gamma)\Gamma(k+l+1/2)}{\Gamma(2k+l+1-i\gamma)\Gamma(k+1)} v^k$$
$$\equiv \sum_{k=0}^{\infty} \alpha_{lk} v^k, \quad |v| < 1. \tag{4.3}$$

The coefficients $\alpha_l(y = 1)$ are again given by Abel summable series; the limit $\lim_{v \to 1-} \alpha_l(v)$ exists since the large k behavior implies a singularity of the type

$$(v+1)^{-3/2 - 2i\gamma - l}$$

for $\alpha_l(v)$ (the factor y^l may be omitted since $y \to 1$). In order to determine $\alpha_l(1)$ we apply the technique discussed in Sec. III with the mapping $v \to z$ [Eq. (3.16) and $p = 1$], i.e.,

$$\alpha_l(v) = \sum_{k=0}^{\infty} \alpha_{lk} v^k = \sum_{k=0}^{\infty} \eta_{lk} z^k, \tag{4.4}$$

where the η_{lk} are determined from the α_{lk} like the d's from the c's in Eq. (3.18).

In the special case of the Coulomb potential we can determine the correct values of α_l by expanding $f^c(x)$ in a binomial series

$$f^c(x) = -\gamma 2^{1+i\gamma} \frac{\Gamma(1+i\gamma)}{\Gamma(1-i\gamma)} (1-x)^{-1-i\gamma}$$
$$= \sum_{l=0}^{\infty} \alpha_l x^l, \quad |x| < 1,$$

$$\alpha_l = \binom{-1-i\gamma}{l} (-1)^{l+1} \gamma 2^{1+i\gamma} \frac{\Gamma(1+i\gamma)}{\Gamma(1-i\gamma)}, \quad l = 0,1,2,\cdots. \tag{4.5}$$

The partial Abel sums [cf. Eq. (4.4)] for $v = 1$ are rapidly approaching these correct values. In Table I we give the values of the first seven coefficients as obtained from the30th partial Abel sum

$$\tilde{\alpha}_l \equiv \sum_{k=0}^{30} \eta_{lk} (z(v=1))^k, \tag{4.6}$$

which are correct for all six decimals. For positive x the rate of convergence is still improved if one introduces the conformal mapping

$$w = \frac{1-(1-x)^{1/2}}{1+(1-x)^{1/2}}, \quad x = \frac{4w}{(1+w)^2}, \tag{4.7}$$
$$(x = -\infty, -1, 0, 1 \to w = -1,$$
$$(1-\sqrt{2})/(1+\sqrt{2}), 0, 1).$$

Then

$$\hat{f}^c(x,y) = \sum_{l=0}^{\infty} \alpha_l(y) x^l = \sum_{l=0}^{\infty} \beta_l(y) w^l, \tag{4.8}$$

$$\beta_l(y) = \sum_{k=0}^{l} (-1)^{l+k} 4^k \frac{\Gamma(l+k)}{\Gamma(2k)\Gamma(l-k+1)} \alpha_k(y). \tag{4.9}$$

In Table I we give $\tilde{\beta}_l$ $(y = 1)$ that are determined from the $\tilde{\alpha}_l(y = 1)$. In Fig. 3 we also exhibit the partial sums of (4.8) for the expansion in x (Σ_α) and in $w(\Sigma_\beta)$ at $x = 0.5$; the comparison with the slowly convergent Abel sum (A-Σ) clearly demonstrates the strongly improved rate of convergence. Note however that polynomial expansions have stability problems; small errors in the coefficients may produce large deviations in the results. For the related questions of

TABLE I. The values of the first seven coefficients $\tilde{\alpha}_l$ [Eq. (4.6)] and $\tilde{\beta}_l$ [determined from the $\tilde{\alpha}_l$ with Eq. (4.9)] for the expansion of the Coulomb amplitude in x^l [Eq. (4.3)] and in $w(x)^l$ [Eq. (4.8)].

l	$\text{Re}\tilde{\alpha}_l$	$\text{Im}\tilde{\alpha}_l$	$\text{Re}\tilde{\beta}_l$	$\text{Im}\tilde{\beta}_l$
0	−0.199795	0.009063	−0.199795	0.009063
1	−0.200701	−0.010916	−0.802804	−0.043664
2	−0.200155	−0.020951	−1.596872	−0.247888
3	−0.199457	−0.027623	−2.363740	−0.558000
4	−0.198766	−0.032610	−3.106192	−0.918432
5	−0.198114	−0.036585	−3.830596	−1.317072
6	−0.197504	−0.039887	−4.532568	−1.750256

stability of analytic continuations we refer to Ciulli *et al.*[23]

Concluding, we have shown that partial wave expansions of scattering amplitudes for Coulomb-type potentials, although divergent when summed in the ordinary sense, converge when treated as Abel sums. Thus, from the theoretical point of view Abel's summation procedure is well suited to sum partial wave expansions of scattering amplitudes for long-range as well as short-range interactions. We have also demonstrated that physical quantities for screened Coulomb-like potentials are in the Abel limit independent of the specific form of screening. Finally, we have shown how to obtain numerical results with the help of conformal mapping techniques and the construction of optimal converging series.

[1]Abel, although contributing fundamental theorems and thus preparing the basis of nowadays limitation theory, condemns divergent series in his letter to Holmboe (1826): "Les séries divergentes sont, en général, quelque chose de bien fatal, et c'est une honte qu'on ose y fonder aucune démonstration" [Hardy's translation: "Divergent series are the invention of the devil and it is shameful to base on them any demonstration whatsoever"]!

[2]G.H. Hardy, *Divergent Series* (Clarendon, Oxford, England, 1963).

[3]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. IV: Analysis of Operators* (Academic, New York, 1978).

[4]J.R. Taylor, Nuovo Cimento B **23**, 313 (1974).

[5]K. Zeller and W. Beekmann, *Theorie der Limitierungsverfahren* (Springer, New York, 1970).

[6]K. Knopp, *Theorie und Anwendungen der Unendlichen Reihen* (Springer, New York, 1964).

[7]N.F. Mott, Proc. R. Soc. (London) Ser. A **118**, 542 (1928).

[8]D.R. Yennie, D.G. Ravenhall, and R.N. Wilson, Phys. Rev. **95**, 500 (1954).

[9]C.R. Garibotti and F.F. Grinstein, J. Math. Phys. **19**, 821, 2405 (1978); **20**, 141 (1979).

[10]A.K. Common and T.W. Stacey, J. Phys. A **11**, 259, 275 (1978).

[11]J.D. Dollard, J. Math. Phys. **5**, 729 (1964); Rocky Mountain J. Math. **1**, 5 (1971).

[12]V.S. Buslaev and V.B. Matveev, Theor. Math. Phys. **2**, 266 (1970). E.B. Manoukian and E. Prugovecki, Can. J. Phys. **49**, 102 (1971); E. Prugovecki, Nuovo Cimento B **4**, 105 (1971); D. Marchesin and M.L. O'Carroll, J. Math. Phys. **13**, 982 (1972).

[13]A.M. Veselova, Theor. Math. Phys. **13**, 1200 (1972); H. Grosse, H. Narnhofer, and W. Thirring, in Proceedings of the Adriatic Summer Meeting on Particle Physics (Rovinj, 1973).

[14]J.M. Combes, "An Algebraic Approach to Quantum Scattering Theory," Marseille, 1969, unpublished.

[15]R.B. Lavine, J. Funct. Anal. **5**, 368 (1970); **12**, 30 (1973); Commun. Math. Phys. **20**, 301 (1971); Indiana Univ. Math. J. **21**, 643 (1972); J. Math. Phys. **14**, 376 (1973).

[16]W.O. Amrein, Ph. A. Martin, and B. Misra, Helv. Phys. Acta **43**, 313 (1970); J.V. Corbett, Phys. Rev. D **1**, 3331 (1970); L.E. Thomas, J. Funct. Anal. **15**, 364 (1974); W.O. Amrein, J.M. Jauch, and K.B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, London, 1977); M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. III: Scattering Theory* (Academic, New York, 1979).

[17]H. Grosse, H.R. Grümm, H. Narnhofer, and W. Thirring, Acta Phys. Austriaca **40**, 97 (1974).

[18]W. Thirring, *Lehrbuch der Mathematischen Physik* (Springer, Wien, 1979), Vol. III.

[19]This form was originally derived by Gordon; W. Gordon, Z. Phys. **48**, 180 (1928).

[20]I.W. Herbst, Commun, Math. Phys. **35**, 181 (1974).

[21]H. van Haeringen, J. Math. Phys. **17**, 995 (1976).

[22]Se, for example, B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, 1971).

[23]S. Ciulli, C. Pomponiu, and I.-S. Stefanescu, Phys. Rep. C **17**, 133 (1975).

[24]M.D. Semon and J.R. Taylor, Nuovo Cimento A **26**, 48 (1975).

[25]S. Klarsfeld, Nuovo Cimento A **43**, 1077 (1966).

[26]F. Calogero, *The Variable Phase Approach to Potential Scattering* (Academic, New York, 1967).

[27]M.D. Semon and J.R. Taylor, J. Math. Phys. **17**, 1366 (1976); Phys. Rev. A **16**, 33 (1977).

[28]E. Prugovecki and J. Zorbas, J. Math. Phys. **14**, 1398 (1973); Nucl. Pys. A **213**, 541 (1973); J. Zorbas, J. Phys. A **7**, 1557 (1974); H. Narnhofer, Nuovo Cimento B **30**, 254 (1975).

[29]For $0 \leqslant \alpha < 1$ the radial Schrödinger operator for $l = 0$, i.e.,

$$-\frac{d^2}{dr^2} + \frac{\alpha - \frac{1}{4}}{r^2} + \frac{\gamma}{r} + V^s(r),$$

is not essentially self-adjoint on $C_0^\infty(0, \infty)$, so we take the phase shift corresponding to its Friedrichs extension (cf. also Ref. 30).

[30]F. Gesztesy, W. Plessas, and B. Thaller, J. Phys. A (to appear).

[31]W. Magnus, F. Oberhettinger, and R.P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966).

[32]An alternative mapping suited especially for the case $x > 0$ would be to map the strip $|\mathrm{Im} y| \leqslant \sqrt{1 - x^2} \equiv a/2$ into the unit circle $|z| < 1$:

$$z = \frac{\exp(\pi y/a) - 1}{\exp(\pi y/a) + 1}.$$

A relation for the coefficients of the expansion in z similar to Eq. (3.18) can be derived.

[33]For example, E. Hewitt and K. Stromberg, *Real and Abstract Analysis* (Springer, New York, 1969).

Source equivalency based on the energy specification of order /

Stanislaw J. Dolzycki[a)]
Department of Electrical Engineering, Marquette University, Milwaukee, Wisconsin 53233

(Received 12 June 1979; accepted for publication 29 February 1980)

It is shown that the multipole expansion of electrostatic energy can be expressed in the form of energy specification $S\{W_e\} = \langle [\mathbf{P}]_l, \nabla^{(l)}\phi \rangle$, where $\nabla^{(l)}$ is a differential operator, whereas ϕ and $[\mathbf{P}]_l$ represent an arbitrary test potential and equivalent reduced volume multipole density, respectively. Two electrostatic sources are l-equivalent if their energy specifications are identical. The formalism by means of which electrostatic multipole sources can be effectively handled is developed.

PACS numbers: 41.10.Dq

I. INTRODUCTION

On reviewing problems regarding the foundation of macroscopic electrodynamics, it appears that certain of these demand further elaboration. One of these problems is concerned with the energy of an electrostatic system, and this may be analyzed with emphasis on two particular aspects. First, the procedure of integration by parts commonly employed in energy-integrals becomes progressively awkward as multipoles of higher order are considered. The second aspect is the physical basis of the macroscopic electrodynamics. It may happen that, due to the mathematical idealization involved, calculations including self-energy give rise to ill-defined results. This is an extremely serious issue to be settled in elementary-particles physics. In macroscopic electrodynamics, however, it may be reasonable to avoid the self-energy of charged particles completely in energy calculation, maintaining that the interaction energy of the system is fundamental to the entire theory. In other words, it is conjectured that any part of macroscopic electrodynamics may be developed by starting with the interaction energy of the source distribution of a system and arbitrary external field.

It is the purpose of this paper to develop a formalism that refines and generalizes the concept of a potential equivalence[1] of electrostatic sources along the lines discussed above. This formalism will allow multiple densities of an arbitrary high order to be included explicitly in the course of the analysis. Needless to say, the multipole properties of atoms and molecules are well defined and therefore it is very natural and important to consider an equivalency relation of sources including their arbitrary polarities. In the following, any source of electromagnetic field will be placed in a family of potential fields produced by external agents. Furthermore, besides the field itself, the field gradient and its tensorial generalizations of higher orders are considered to be significant probing agents that test the characteristics of a given source distribution.

The main result established in this paper is the formalism for handling a multipole of arbitrarily high order, based on the concept of a source body as defined in Def. 1 and its application to the multipole expansion of electrostatic energy. Subsequently, it will be shown that multipole densities may always be considered equivalent to appropriate densities of lower orders. Both the aforementioned formalism and the concept of equivalency are based on the interaction between a given source body $[\mathbf{P}]$ and an externally applied field. To develop a consistent formalism that conforms with these objectives, the concept of energy specification[2] is made precise by defining it as a set of all possible values of the functional $\langle [\mathbf{P}], L\phi \rangle$, where L is a linear differential operator and ϕ is an arbitrary test potential. In view of this, the theory of generalized functions appears to be a desirable framework with which to work. The merit of the distributional formalism is clearly presented when one deals with discontinuities and other singularities of physical quantities. Thus a considerable amount of applications of the generalized functions to macroscopic electrodynamics have been given by now. In particular, the description of electromagnetic sources has been discussed in this context. Mazur[3] and de Groot and his school[4] have developed the generalized description of the multipole electric and magnetic sources with polarity of an arbitrary high order by employing the Taylor series-type expansion of the δ function. The generalized formulation of Maxwell's equations has been studied by several authors.[5,6,7] Some applications of the vector generalized functions are also discussed by Gagnon.[8] In the present paper, the merit of the distributional formalism will be emphasized in a broader sense in that, by its virtue, the field-gradient qualities of an external field may be related straightforwardly with the boundary effects that appear upon forcing the support of a source density to vanish outside a finite space.

In Sec. II a notation will be introduced in which various tensor operations are incorporated with the distributional formalism. In particular, the formula that generalizes a classical identity $\nabla\cdot(\phi\mathbf{a}) = \nabla\phi\cdot\mathbf{a} + \phi\nabla\cdot\mathbf{a}$ is of special interest. In Sec. III the concepts of the source body and the energy specification of order l are defined. It is shown in Sec. IV how to transform an energy integral of a system into another one which is more readily ameanable to physical interpretation by systematically using the procedure

$$\langle [\mathbf{P}]_m, L\phi \rangle \cong \langle \nabla^{(m-l)}[\mathbf{P}]_m, \nabla^{(l)}\phi \rangle = \langle [\mathbf{P}]_{l,eq}, \nabla^{(l)}\phi \rangle, \tag{1}$$

where the original differential operator L is factored like $L = \nabla^{(m-l)}\nabla^{(l)}$ while $[\mathbf{P}]_m$ and $[\mathbf{P}]_{l,eq}$ are a given multipole source body of the order m and reduced multipole source body of order l, respectively. Furthermore, the above

[a)]On leave of absence from the Technical University of Wroclaw, Poland.

 0022-2488/81/020320-05$1.00

procedure yields an additive version of the multipole expansion of interaction energy. Finally, in Sec. V the importance of the formalism and the consequences for its application are pointed out.

II. DIVERGENCE ITERATIONS OF CONTRACTION

Let R^3 be a real three-dimensional Euclidean space with a fixed base and let $\mathbf{F} = \{F_{j_1\cdots j_k}\} = \{F_{J_k}\}$, $\mathbf{a} = \{a_{i_1\cdots i_m}\} = \{a_{I_m}\}$, respectively, denote sufficiently smooth tensor functions defined on R^3. The k-element (m-element) set of indices j_α (i_β) is denoted by J_k (I_m). The sth tensorial contraction of $\mathbf{F}$ and $\mathbf{a}$ is defined as

$$\mathbf{C}^{(s)}(\mathbf{F},\mathbf{a}) = \left\{\sum_{J_s} F_{J_k\cdots J_s} a_{J_s I_{m-s}}\right\} = \{F_{J_k\cdots J_s} a_{J_s I_{m-s}}\}, \tag{2}$$

where the contraction in the second member is performed over arbitrary s-element subset J_s chosen as common subset from J_k and I_m. It is assumed in the following that a contraction involved in (2) starts with the last available index of $\mathbf{F}$ and the first available index of $\mathbf{a}$, whereupon the available pair of indices will be subsequently contracted; such a procedure continues s times. The l th iterations of gradient and divergence are defined, respectively, by

$$\begin{aligned} &\nabla^{(l)}\mathbf{F} = D_{h_l}\cdots D_{h_1} F_{j_1\cdots j_k}, \\ &\nabla^{(l)}\cdot\mathbf{F} = D_{h_l}\cdots D_{h_1} F_{j_1\cdots h_l\cdots h_1}, \end{aligned} \tag{3}$$

where D_h denotes a partial derivative with respect to the variable x_h. By direct computation one verifies that for $m > k = s \geqslant 0$ (Appendix A),

$$\nabla\cdot[\mathbf{C}^{(s)}(\mathbf{F},\mathbf{a})] = \mathbf{C}^{(s+1)}(\nabla\mathbf{F},\mathbf{a}) + \mathbf{C}^{(s)}(\mathbf{F},\nabla\cdot\mathbf{a}). \tag{4}$$

One can generalize the above formula to obtain the $(m-s)$th divergence iteration. In fact, a remarkably useful formula,

$$\nabla^{(m-s)}\cdot[\mathbf{C}^{(s)}(\mathbf{F},\mathbf{a})] = \sum_{t=0}^{t=m-s} \mathbf{C}^{(s+t)}(\nabla^{(t)}\mathbf{F},\nabla^{(m-s-t)}\cdot\mathbf{a}), \tag{5}$$

may be proved by introducing for each t

$$\mathbf{C}^{(s+t)}(\nabla^{(t)}\mathbf{F},\nabla^{(m-s-t)}\cdot\mathbf{a}) = \sum_{H_t} D_{H_t} F_{J_s} D_{H_{m-s-t}} a_{J_s H_t H_{m-s-t}}. \tag{6}$$

The summation in (6) is performed on all t-elements sets H_t that are possible under the condition $H_t \subset CI_s$, where CI_s is understood as a complement of I_s with respect to the set I_m. The detailed calculations necessary to prove Eq. (5) are presented in Appendix A.

For a scalar test function ϕ, the derivative of a generalized scalar function ρ is defined by[9]

$$\langle \rho',\phi\rangle = -\langle \rho,\phi'\rangle. \tag{7}$$

For a generalized vector function $\boldsymbol{\rho}$, the following, formula will be useful

$$\langle\nabla\cdot\boldsymbol{\rho},\phi\rangle = \langle D_i\rho_i,\phi\rangle = \langle \rho_i, -D_i\phi\rangle = \langle \boldsymbol{\rho}, -\nabla\phi\rangle, \tag{8}$$

where $\langle \boldsymbol{\rho},\boldsymbol{\phi}\rangle = \Sigma_i\langle\rho_i,\phi_i\rangle = \langle\rho_i,\phi_i\rangle$ is understood. In order to generalize the above procedure, the generalized sth contraction $\mathbf{C}^{(s)}\langle\mathbf{F},\boldsymbol{\phi}\rangle$ of a generalized tensor function $\mathbf{F}$ will be introduced. Let $\mathbf{F}$ be a tensor generalized function of order k and let $\boldsymbol{\phi}$ be a tensor test function of the order m in the sense that all components of $\boldsymbol{\phi}$ belong to the Schwartz space $\mathscr{D}$. For $0 \leqslant s \leqslant \min(k,m)$ a generalized function $\mathbf{C}_F^{(s)}$ is defined by

$$\mathbf{C}_F^{(s)}(\boldsymbol{\phi}) = \mathbf{C}^{(s)}\langle\mathbf{F},\boldsymbol{\phi}\rangle = \langle F_{J_k\,J_s},\phi_{J_s I_m}\rangle, \tag{9}$$

where a contraction is carried out with respect to the "inner" set of indices, J_s, as explained at the beginning of this section, Eq. (12).

III. ENERGY SPECIFICATIONS OF ORDER l

Let Ω and S_Ω denote, respectively, a simply connected region in the three-dimensional space and the surface bounding the region. A characteristic function $\delta(\Omega)$ is defined, as usual, by

$$\delta(\Omega) = \eta[S_\Omega(x)] = \begin{cases} 1, & S_\Omega(x) \geqslant 0, \\ 0, & S_\Omega(x) < 0, \end{cases} \tag{10}$$

where x is an arbitrary point in the space and $S_\Omega(x) = 0$ defines the surface S. The sources to be considered in the present paper may have an arbitrary support. Furthermore, we make it a rule to employ an integral extended over the entire space. Hence, a source density is given in the whole space even if the distribution in individual cases extends over a finite space only.[10] A generalized function corresponding to a source distribution in this sense may be denoted by $[\mathbf{P}]$ and the physical objects that embody such a source distribution will be called the physical or source bodies. In order to deal with multipole densities of arbitrarily high order a system of notations must be developed which is capable enough to present the symbol's meaning clearly and consistently in any situation of application. Definite forward steps to achieve our goal will be made by introducing two definitions.

Definition 1: An electrostatic multipole source body of order r is a generalized function

$$[\mathbf{P}^{(m)}]_r = \mathbf{C}^{(m-r)}[\nabla^{(m-r)}\delta(\Omega),\mathbf{P}^{(m)}]. \tag{11}$$

The list of the quantities on the right-hand side reads

$\mathbf{P}^{(m)}$	Tensor source density, sufficiently regular and determined in the whole space R^3,
$\delta(\Omega)$	Characteristic function of the source support,
$\mathbf{C}^{(m-r)}$	The $(m-r)$th tensorial contraction defined by (2).

The quantity that appears for $m = r$, $[\mathbf{P}^{(m)}]_m = [\mathbf{P}]_m$, will be called an elementary source body or elementary volume source density. For $m > r$, a source body $[\mathbf{P}^{(m)}]_r$ will be called a physical surface or simply a layer. Occasionally it is convenient to use a traditional notation that is applicable when $m = 0$. For example, $[\mathbf{P}]_0 = [\rho]$ is a source body associated with a volume charge density.

In the present analysis of the interaction energy between an arbitrary source body $[\mathbf{P}]$ and a test field ϕ, the quality of a source body will be categorized by applying the multipole expansion as defined by (11), while the energy integral will be subject to its own multipole expansion. In such an expansion of energy integrals, integrals of the type

$\langle [\mathbf{P}], L\phi \rangle$ will be used. A suitably chosen set of such integrals is the basis of the formalism to be developed. The class of operators L may be generalized, but for the purpose of laying a foundation for the macroscopic electrostatics, we limit it to linear differential operators $L = \nabla^{(m-l)}\nabla^{(l)}$ required for the multipole expansion of energy (Appendix B). With the above remarks in mind, the following definition is introduced.

Definition 2: An electrostatic energy specification of the order l of a multipole source body $[\mathbf{P}]_m$ is a set $S\{W_e\}$ of functionals, where an individual functional W_e is the energy

$$W_e = \mathbf{C}^{(m)}\langle [\mathbf{P}]_m, \nabla^{(m-l)}(\nabla^{(l)}\phi)\rangle, \tag{12}$$

evaluated for a particular test potential of the family $\mathscr{D}$. The factorization $\nabla^{(m-l)}\nabla^{(l)}$ indicates explicitly that $\nabla^{(l)}\phi$, the l th tensorial derivative of test field, is being investigated in the specification.

As is the case with a test function $f(x)$ discussed in the standard reference on the macroscopic averaging process,[11] the test potential here introduced does not need to be specified in detail. All that is required is that it is sufficiently smooth and it vanishes outside bounded regions. Two source bodies $[\mathbf{P}_1]_l$, $[\mathbf{P}_2]_l$ are said to be equivalent if, and only if, for some (l), $(l = 0,1,2,\cdots)$,

$$\langle [\mathbf{P}_1]_l, {}^{(l)}\phi \rangle = \langle [\mathbf{P}_2]_l, {}^{(l)}\phi \rangle. \tag{13}$$

IV. REDUCED MULTIPOLE DENSITIES

In this section it is shown that a distribution of multipole densities of order $m \geqslant l$ contained in a region Ω can be effectively replaced by equivalent volume multipole densities of order l. This result leads naturally to the additive form of the multipole expansion of energy. First consider a single elementary multipole source body $[\mathbf{P}]_m$. In view of Eq. (5) and Def. 2, $[\mathbf{P}]_m$ is equivalent to the sum of the bodies of order l, i.e.,

$$\delta(\Omega)\mathbf{P}^{(m)} \sim (-1)^{m-l}\sum_{t=0}^{m-l}\mathbf{C}^{(t)} \times [\nabla^{(t)}\delta(\Omega), \nabla^{(m-l-t)}\cdot\mathbf{P}^{(m)}]. \tag{14}$$

Indeed, using a definition of the generalized derivative

$$\begin{aligned}
&\mathbf{C}^{(m)}\langle \delta(\Omega)\mathbf{P}^{(m)}, \nabla^{(m-l)}(\nabla^{(l)}\phi)\rangle \\
&\quad = \mathbf{C}^{(l)}\langle (-1)^{m-l} D_{i_{l+1}}\cdots D_{i_m} \\
&\qquad \times [\delta(\Omega)\cdot P_{i_1\cdots i_l\cdots i_m}], D_{h_1}\cdots D_{h_l}\phi\rangle \\
&\quad = \mathbf{C}^{(l)}\langle (-1)^{m-l}\nabla^{(m-l)}(\delta(\Omega)\mathbf{P}^{(m)}), \nabla^{(l)}\phi\rangle,
\end{aligned} \tag{15}$$

and it is sufficient to apply Eq. (5) in order to obtain the decomposition (14). Suppose that a region Ω contains multipole moment densities of orders m for every m from l to $M, l \leqslant m \leqslant M$. From Eq. (5) one obtains

$$\sum_{m=l}^{M}(-1)^{m-l}\nabla^{(m-l)}\cdot[\mathbf{P}^{(m)}] = \sum_{m=l}^{M}\sum_{t=0}^{m-l}(-1)^{m-l}\mathbf{C}^{(t)}(\nabla^{(t)}\delta(\Omega), \nabla^{(m-l-t)}\cdot\mathbf{P}^{(m)}), \tag{16}$$

and after changing the order of summation in (16), the energy specification $S\{W_e\}$ of order l for the source body $\Sigma_l^M[\mathbf{P}]_m = \delta(\Omega)\{\mathbf{P}^{(l)} + \cdots\mathbf{P}^{(n)}\}$ becomes

$$\begin{aligned}
S\{W_e\} &= \sum_{m=l}^{M}\mathbf{C}^{(m)}\langle \delta(\Omega)\mathbf{P}^{(m)}, \nabla^{(m-l)}(\nabla^{(l)}\phi)\rangle \\
&= \sum_{m=l}^{M}(-1)^{m-l}\mathbf{C}^{(l)}\langle \nabla^{(m-l)}\cdot[\mathbf{P}]_m, \nabla^{(l)}\phi\rangle \\
&= \mathbf{C}^{(l)}\left\langle \sum_{m=l}^{M}\sum_{t=0}^{m-l}(-1)^{m-l}\mathbf{C}^{(t)} \times(\nabla^{(t)}\delta(\Omega), \nabla^{(m-l-t)}\cdot\mathbf{P}^{(m)}), \nabla^{(l)}\phi\right\rangle \\
&= \mathbf{C}^{(l)}\left\langle \sum_{t=0}^{M-l}\mathbf{C}^{(t)}(\nabla^{(t)}\delta(\Omega), \right. \\
&\qquad \left. \times \sum_{H_t}\sum_{m=l+t}^{M}(-1)^{m-l}\nabla^{(m-l-t)}\cdot\mathbf{P}^{(m)}\right), \nabla^{(l)}\phi\Big\rangle \\
&= \mathbf{C}^{(l)}\left\langle \sum_{t=0}^{M-l}\mathbf{C}^{(t)}(\nabla^{(t)}\delta(\Omega), \mathbf{P}_{l,t}^{M}), \nabla^{(l)}\phi\right\rangle \\
&= \mathbf{C}^{(l)}\left\langle \sum_{t=0}^{M-l}[\mathbf{P}_t]_l, \nabla^{(l)}\phi\right\rangle.
\end{aligned} \tag{17}$$

Multipole densities

$$\mathbf{P}_{l,t}^{M} = \sum_{H_t}\sum_{m=l+t}^{M}(-1)^{m-l}\nabla^{(m-l-t)}\cdot\mathbf{P}^{(m)} \tag{18}$$

of the source bodies $[\mathbf{P}_t]_l$ will be referred to as the reduced volume source densities of order l. Note here that $t = 0$ corresponds to the interior of a region Ω and $t \geqslant 1$ represent supports of source distributions at the boundary. Assuming $\mathbf{P}_{l,0}^{M} = \mathbf{P}_l^{M}$ and $\mathbf{P}_{l,t}^{M} = \mathfrak{S}_{l,t}^{M}$, $(t \geqslant 1)$, Eq. (17) can be presented in the more explicit form

$$S\{W_e\} = \mathbf{C}^{(l)}\left\langle [\mathbf{P}_l^{M}] + \sum_{t=1}^{M-l}[\mathfrak{S}_{l,t}^{M}], \nabla^{(l)}\phi\right\rangle. \tag{19}$$

Since Eq. (19) remains valid for an arbitrary M, the following property has been proved: An elementary source body $[\Sigma_m\mathbf{P}^{(m)}]$ containing volume multipole moment densities of an arbitrary high order is equivalent, in the sense of the energy specification of order l to source bodies with reduced multipole densities

$$\mathbf{P}_{l,t} = \sum_{H_t}\sum_{m=l+t}^{\infty}(-1)^{m-l}\nabla^{(m-l-t)}\cdot\mathbf{P}^{(m)}. \tag{20}$$

In view of the classical expression for the energy of a continuous charge distribution in an external field

$$W_e = \int_{\Omega}\rho(T)\phi(T)\,d\vartheta_T = \langle [\rho], \phi\rangle, \tag{21}$$

the case $l = 0$ is of particular importance. Indeed, reduced multipole densities become simply the volume charge densities and consequently the multipole expansion of the energy can be written in the "additive" form. Furthermore, the last formula establishes an equivalency relation between a multipole density distribution and an equivalent volume charge distribution.

V. CONCLUDING REMARKS

Evidently the energy specification formalism developed in the present paper deals with multipole distributions of higher order consistently. The basic physical ideas underlying the formalism are commonly accepted ones, and a con-

cept of the generalized derivative has been extensively applied. The multipole expansion of electrostatic energy is a crucial preliminary step that must be taken before dealing with a bulk material based on the macroscopic distribution of molecular species. On realizing this circumstance, it has been shown that the generalized derivative effectively handles an arbitrary source and family of external fields. The interpretation of the intermediate and final results in terms of familiar physical images is naturally accomplished by adopting procedures similar to procedures in conventional electrodynamics. Thus, interaction between a source ρ and an external potential φ, $W_e = \int \rho\varphi d\vartheta$, has been handled along the line discussed above. The subsequent step of development was to transform the terms of the form $\mathbf{P}^{(m)}.\nabla^{(m)}\phi$ into the form $\rho_m\phi$, where ρ_m represents an equivalent charge density. When this is accomplished, the resultant additive character of equivalent charge densities is apparent, as may be schematically shown by

$$\sum \mathbf{P}^{(m)}.\nabla^{(m)}\phi \sim \left(\sum \rho_m\right)\phi, \quad l = 0. \tag{22}$$

It is interesting to note that the basic formula (20) determines an equivalent volume multipole source density of the order l, if one wishes to stop a "reduction" process at the level l. Moreover, Eq. (20) includes the boundary effects that result from forcing all the sources to vanish outside the source body. It can be easily seen that for $l = 0$ the volume moment density becomes simply the volume charge density. Furthermore, a classical Lorentz's "dipole approximation" requires a test field to be in the form $\mathbf{E}_t = -\nabla\phi$ and consequently Eq. (20) yields, inside a region ($t = 0$),

$$P_0 = \rho = \sum (-1)^m \nabla^{(m)}.\mathbf{P}^{(m)}, \tag{23}$$

and at its boundary,

$$-\mathfrak{S}_{0,t} = \mathbf{P}_t = \sum_{H_t}\sum_{m=t} (-1)^m \nabla^{(m-t)}.\mathbf{P}^{(m)}. \tag{24}$$

For $t = 1$ one obtains the more general case of a known formula according to which a volume multipole distribution inside a region is equivalent to the sum of the volume charge density (23) and volume charge density[2]

$$\mathfrak{S}_{0,1} = \sum_{H_t}\sum_{m=1} (-1)^m \nabla^{(m-1)}.\mathbf{P}^{(m)}, \tag{25}$$

which corresponds to a surface charge at the boundary. An additive version of the multipole expansion will prove to be a convenient setting in which the notions of polarization charges and currents have the more rigorous meaning. This is, however, the subject of a separate paper.

ACKNOWLEDGMENTS

The study of this problem was motivated by ideas of Professor Adam Skopec. The author thanks him for stimulating discussions and encouragement in the early stage of this work. He would also like to thank Professor Smio Tani for valuable discussions and a number of useful suggestions concerning the final version of the manuscript. He is grateful to Professor S.V. Jaskolski for his hospitality and to Marquette University for financial support.

APPENDIX A

In this Appendix, Eq. (5) is derived. Let $\mathbf{F}$ and $\mathbf{a}$ denote, respectively, tensorial functions of order s and m ($m > s$). In the following, summation is understood for all repeated indices. Using a definition of the tensorial contraction one obtains

$$\begin{aligned}
\nabla^{(m-s)}.\mathbf{C}^{(s)}(\mathbf{F},\mathbf{a}) &= \nabla^{(m-s)}.(F_{j_1\cdots j_s}a_{j_s\cdots j_1 i_{s+1}\cdots i_m}) \\
&= D_{i_{s+1}}\cdots D_{i_m}(F_{j_1\cdots j_s}a_{j_s\cdots j_1 i_{s+1}\cdots i_m} \\
&= D_{i_{s+1}}\cdots D_{i_{m-1}}(D_{i_m}F_{j_1\cdots j_s} \\
&\quad \times a_{j_s\cdots j_1 i_{s+1}\cdots i_m} + F_{j_1\cdots j_s}D_{i_m}a_{j_s\cdots j_1 i_{s+1}\cdots i_m}) \\
&= D_{i_{s+1}}\cdots D_{i_{m-2}}(D_{i_{m-1}}D_{i_m}F_{j_1\cdots j_s} \\
&\quad \times a_{j_s\cdots j_1 i_{s+1}\cdots i_m} + D_{i_m}F_{j_1\cdots j_s}D_{i_{m-1}}a_{j_s\cdots j_1\cdots i_m} \\
&\quad + D_{i_{m-1}}F_{j_1\cdots j_s}D_{i_m}a_{j_s\cdots j_1\cdots i_m} \\
&\quad + F_{j_1\cdots j_s}D_{i_{m-1}}D_{i_m}a_{j_s\cdots j_1\cdots i_{m-1}i_m})\cdots \\
&= D_{i_{s+1}}\cdots D_{i_m}F_{j_1\cdots j_s}a_{j_s\cdots j_1 i_{s+1}\cdots i_m} \\
&\quad + D_{i_{s+1}}\cdots D_{i_{m-1}}F_{j_1\cdots j_s}D_{i_m}a_{j_s\cdots i_m} \\
&\quad + \cdots D_{i_{s+2}}\cdots D_{i_m}F_{j_1\cdots j_s}D_{i_{s+1}}a_{j_s\cdots j_1 i_{s+1}\cdots i_m} \\
&\quad + \cdots + F_{j_1\cdots j_s}D_{i_{s+1}}\cdots D_{i_m}D_{i_m}a_{j_s\cdots i_m} \\
&= \sum_{t=0}^{t=m-s}\sum_{H_t} D_{H_t}F_{J_s}D_{H_{m-s-t}}a_{J_s H_t H_{m-s-t}} \\
&= \sum_{t=0}^{t=m-s} \mathbf{C}^{(s+t)}(\nabla^{(t)}\mathbf{F},\nabla^{(m-s-t)}\mathbf{a}),
\end{aligned} \tag{A1}$$

where H_t denote all possible t-element ($0 \leqslant t \leqslant m-s$) subsets of the set $\{i_{s+1}\cdots i_m\}$. For $m - s = 1$, (A1) yields

$$\nabla\cdot\mathbf{C}^{(s)}(\mathbf{F},\mathbf{a}) = \mathbf{C}^{(s+1)}(\nabla\mathbf{F},\mathbf{a}) + \mathbf{C}^{(s)}(\mathbf{F},\nabla\cdot\mathbf{a}) \tag{A2}$$

and, for example, for $s = 1$, $m = 2$, one obtained directly

$$\begin{aligned}
\nabla\cdot\mathbf{C}(\mathbf{F},\mathbf{a}) &= \nabla\cdot(F_i a_{ij}) = D_j(F_i a_{ij}) = D_j F_i a_{ij} + F_i D_j a_{ij} \\
&= \mathbf{C}^{(2)}(D_l F_n, a_{ij}) + \mathbf{C}(F_n, D_j a_{ij}) \\
&= \mathbf{C}^{(2)}(\nabla\mathbf{F},\mathbf{a}) + \mathbf{C}(\mathbf{F},\nabla\cdot\mathbf{a}).
\end{aligned} \tag{A3}$$

APPENDIX B

Let $\overline{\Pi_0 P} = \mathbf{R}$ and $\overline{\Pi_0 \Pi} = \xi$ denote the points P and Π, respectively. In addition, let us define $\mathbf{r} = \overline{\pi P} = \mathbf{R} - \xi$. Using the Taylor expansion a potential ϕ can be written[12]

$$\phi(P) \cong \sum_{m=0}^{\infty} (-1)^m (\mathbf{P}^{(m)}.\nabla^{(m)}(1/R) = \int (\rho/r)\, d\vartheta, \tag{B1}$$

where

$$\frac{1}{r} = \sum_{m=0}^{\infty} \frac{1}{m!}(-1)^m (\xi^{(m)}.\nabla^{(m)})(1/R) \tag{B2}$$

and

$$\mathbf{P}^{(m)} = \frac{1}{m!}\int d\vartheta\, \rho\xi^{(m)} \tag{B3}$$

denotes a multipole moment density of order m with respect to Π_0 of a volume source density ρ. A mutual energy between a source distribution and test field then, can be expressed by

$$W_e = \int d\vartheta\, \rho\phi = \langle \rho, \phi \rangle, \tag{B4}$$

where ϕ is an arbitrary test potential. Expanding ϕ in the vicinity of Π_0, one obtains

$$W_e = \int d\vartheta_{II}\, \rho(II)\phi\,(II) = \int d\vartheta_{II}\, \rho(II)\phi\,(\mathbf{\Pi}_0 + \xi)$$
$$= \int d\vartheta_{II}\, \rho(II)\left(\sum_{m=0}^{\infty} \frac{1}{m!}(\xi\cdot\nabla_{II_0})^{(m)}\cdot\phi\,(II_0)\right)$$
$$= \sum_{m=0}^{\infty} (\mathbf{P}_{II_0}^{(m)}\cdot\nabla_{II_0}^{(m)})\,\phi\,(II_0). \tag{B5}$$

Eq. (B5) determines the type of differential operators L to be used in the formalism.

[1]N. Bleinstein and J. K. Cohen, J. Math. Phys. **18**, 194 (1977).
[2]To the knowledge of the author, a notion of the energy specification was used for the first time by A. Skopec in description of dipole distributions in dielectrics (private communication).
[3]D. Mazur and B. R. A. Nijboer, Physica **19**, 971 (1953).
[4]S.R. de Grott, "The Maxwell equations: nonrelativistic and relativistic derivations from electron theory," In *Studies in Statistical Mechanics*, Vol. IV (North-Holland, Amsterdam, 1969).
[5]M. Bouix, C.R. Acad. Sci. **242**, 2858 (1958).
[6]R. Radulet and R. I. Ciric, Rev. Roum. Sci. Tech. Ser. Electrotech. **15**, 565 (1971).
[7]M. Idemen, "The Maxwell equations in the sense of distributions," IEEE Trans. **AP-21**, 736 (1973).
[8]R. J. Gagnon, Amer. J. Phys. **38**, 879 (1970).
[9]I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964).
[10]K. O. Friedrichs, *Mathematical Methods of Electromagnetic Theory* (New York U.P., New York, 1974), p. 49.
[11]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), p. 228.
[12]C. B. Kadafar, Int. J. Eng. Sci. **9**, 831 (1971).
[13]In the abstract of this paper, the term "equivalent reduced volume density," in its full form, has been used. As a result of using the generalized function formalism, any density in the present formalism means a volume density, even though some of them may correspond to a surface or line density in physical interpretation. Hereafter we abbreviate it as the volume density, retaining the italic form to emphasize the point of view of the present formalism.

On the Lorentz dipole approximation in static electrodynamics

Stanislaw J. Dolzycki[a)]
Department of Electrical Engineering, Marquette University, Milwaukee, Wisconsin 53233

(Received 12 June 1979; accepted for publication 29 February 1980)

The energy specification $S\{W_e\} = \langle [\mathbf{P}]_{l=0}, \varphi \rangle$ of electrostatic sources introduced in the preceding paper is further developed. The electrostatic potential Φ due to a source distribution is defined as a generalized function satisfying the energy specification equation $S\{W_e\} = \langle -\epsilon_0 \nabla^{(n)} \cdot \nabla^{(n)} \Phi, \varphi \rangle$, where n denotes the order of a multipole approximation is reviewed. Specifically, the generalized function of an equivalent field corresponding to a classical field intensity $\mathbf{E}$ is introduced in terms of a given energy specification $S\{W_e\}$. Besides the equivalent field, a family of generalized functions $\mathbf{D}_n$ referred to as the characteristic fields is introduced to deal with the displacement vector $\mathbf{D}$. A formal description of the so-called polarization charges comes out of an analysis of an equivalent field. Equivalent fields of magnetostatic problems are discussed on the basis of the magnetostatic energy specification $S\{W_m\}$.

PACS numbers: 41.10.Dq

I. INTRODUCTION

In this paper the electrostatic energy specification will be further developed as a continuation of the previous work.[1] For all practical purposes, the energy specification of order $l = 0$ is of particular interest. In this case, all macroscopic multipole moment densities will be reduced to equivalent *volume* charge densities.[2] For example, the commonly known procedure to handle a volume dipole moment density $\mathbf{P}$ represents an important case of such a situation. Indeed, from a macroscopic point of view, if there are no higher multipole moment densities within a given bounded region Ω the potential Φ at a point $\mathbf{x}$ outside Ω is given by

$$\Phi(\bar{x}) = \int_{\Omega} d^3x \, (\rho(\bar{\mathbf{x}}')/ |\bar{\mathbf{x}} - \bar{\mathbf{x}}'| + \mathbf{P}(\bar{\mathbf{x}}')\cdot(\bar{\mathbf{x}} - \bar{\mathbf{x}}')/|\bar{\mathbf{x}} - \bar{\mathbf{x}}'|^3),$$

where $\mathbf{P}$ is a macroscopic volume dipole density and ρ is a volume charge density located within the region. Using a well-known procedure,[3] the potential can be written

$$\begin{aligned}\Phi &= \int_{\Omega} d^3x' \quad [\rho(\mathbf{x}')/|\mathbf{x} - \mathbf{x}'| + P(\mathbf{x}')\cdot\nabla'(1/|\mathbf{x} - \mathbf{x}'|)] \\ &= \int_{\Omega} d^3x' \, \{\rho(\mathbf{x}')/|\mathbf{x} - \mathbf{x}'| + [\nabla'\cdot(\mathbf{P}1/|\mathbf{x} - \mathbf{x}'|) \\ &\quad - 1/|\mathbf{x} - \mathbf{x}'|\nabla'\cdot\mathbf{P}]\} \\ &= \int_{\Omega} d^3x' \, \rho(\mathbf{x}')/|\mathbf{x} - \mathbf{x}'| + \int_{\Omega} d^3x \, (-\nabla\cdot\mathbf{P}/|\mathbf{x} - \mathbf{x}'|) \\ &\quad + \int_{S_\Omega} dS \, \mathbf{P}\cdot\mathbf{n}/|\mathbf{x} - \mathbf{x}'| \, .\end{aligned}$$

In view of the above formula, one usually considers that the combination of the volume charge density $-\nabla\cdot\mathbf{P}$ and surface charge density $\mathbf{P}\cdot\mathbf{n}$ is equivalent, as far as the resultant potential is concerned, to the volume dipole moment density $\mathbf{P}$. Recently, Bleinstein *et al.*[4] have mentioned that this potential equivalency can be used to demonstrate the existence as well as nonuniqueness of nonradiating sources in stationary fields. To quote Ref. 4, "...one can use Green's theorem to replace a source distribution in a domain by a monopole-dipole distribution over any surface bounding that domain such that each yields the same field outside the bounding surface. The difference of these source distributions then yields zero field outside the bounding surface."

In this paper, a more general and rigorous analysis will be worked out on the same problem. Using the formalism developed in the previous paper, the energy W_e of a given source distribution will be systematically replaced by the standardized form of the energy specification $S\{W_e\} = \langle [\rho], \phi \rangle$, where $[\rho] = [\mathbf{P}]_{l=0}$ is a reduced volume charge density and ϕ is an arbitrary test potential of the external field. Furthermore, by virtue of the requirement that the energy specification $S\{W_e\}$ be kept invariant, the generalized function of an electrostatic potential Φ will be defined uniquely once the order of multipole approximation is specified. The classical Lorentz dipole approximation appears to be the lowest order case of the so-defined multipole expansion formalism. It will be shown how to work out a formal description of polarization sources up to an arbitrary order of multipole expansion. One advantage of the present formalism is that, due to the mathematical streamlining, so to speak, a more adequate interpretation and definite meaning is assigned to the so-called true and bound charge-currents that otherwise must be presented in terms of figurative comparisons and arguments of a highly speculative nature ("polarization charge," "cutting of dipoles" etc.). It becomes evident, also, the such expressions as " the same field outside a domain" and "difference between sources" acquire quite definite and precise meaning when considered in the here proposed scheme of energy specification. The formalism developed in the previous paper provides all formulas needed in this systematic reduction of energy specification to the standardized form.

II. EQUIVALENT AND CHARACTERISTIC FIELDS

In this section the nature of the polarization charge is discussed by means of energy specification of order $l = 0$. It is shown that a formal description of polarization sources

[a)]On leave of absence from the Technical University of Wroclaw, Poland.

 0022-2488/81/020325-06$1.00

can be given in terms of characteristic and equivalent sources by transforming the so-called energy specification equation. To begin with, a simple case of energy specification $S\{W_e\}$ will be discussed. Let the point charges $q_1,...,q_n$ be placed at the points $T_1,...,T_n$. Let a test potential ϕ be a sufficiently smooth scalar function vanishing outside the bounded region. The energy of the system consisting of the point charges $\{q_1,...,q_n\}$ and an external field represented by the test potential ϕ is given by

$$W_e = \sum_{k=1}^{n} q_k \phi(T_k). \tag{1}$$

By definition, W_e is always finite and, in view of discussion given at the beginning of the previous paper, it is assumed to be the only meaningful energy-characterization of electromagnetic sources. In other words, the interaction energy between a given charge distribution and an external source is considered a crucial entity to be produced by a macroscopic experiment. Its fundamental significance arises from the fact that it remains invariant throughout the application of the present formalism. The set of numbers obtained from Eq. (1) for all elements ϕ of a given space $\mathscr{D}$ of test potentials represents a linear functional defined on $\mathscr{D}$. This functional will be referred to as the energy specification $S\{W_e\}$ of the point charge system. Similarly, for a volume charge density ρ localized in a region Ω and arbitrary test potential ϕ, one can define an integral

$$W_e = \int_{\Omega} \rho(T)\phi(T)\, d\vartheta \tag{2}$$

representing the electric field energy of the system.

The energy specification $S\{W_e\}$ represents a set of energies associated with a given set up of sources and family of external test fields. Note that such energies are functionals of the source distribution and are invariant under coordinate transformations. A specific analytical form of W_e can be deduced from the multipole expansion of energy. Indeed, this expansion defines the characteristic way in which various source multipoles interact with an external field. Electric charges interact with an external potential, dipoles with an electric field, quadrupoles with a field gradient, and so forth. For example, for a given test potential, a distribution of dipoles localized at the points $T_1,...,T_n$ has the energy given by

$$W_e = \sum_{k=1}^{n} \mathbf{p}_k \cdot (\nabla\phi)_{T_k}, \tag{3}$$

where $\mathbf{p}_k$ is a dipole moment at the point T_k. Similarly, for a volume dipole moment density $\mathbf{P}$ one obtains

$$W_e = \int_{\Omega} \mathbf{P}\cdot\nabla\phi\, d\vartheta, \tag{4}$$

and for a volume quadrupole moment density $P_{\alpha\beta}$

$$W_e = \int_{\Omega} \left(\sum_{\alpha,\beta} P_{\alpha\beta} \frac{\partial^2\phi}{\partial\alpha\partial\beta} \right) d\vartheta. \tag{5}$$

It is the goal of the present scheme to exploit the wealth of forms that are obtained by transforming the functional sets (3)–(5) as exhaustively and properly as possible to cover various aspects of electromagnetism. Such transformations are dictated by the basic principle that an energy is a scalar. Probably, it is not improper to explain the goal of the present scheme more colloquially by comparison with the advantage of variational formalism.[5] There is a stationarity requirement imposed on a functional derived from a scalar Lagrangian function serves in compact form as a representative for the wealth of field equations. In other words, one starts with a compact functional form and obtains a variety of equations and their consequences upon developing the formalism. One may describe that situation by stating that the variational formula is the "pre-generating functional form" of the field equations. The purpose of the present paper is to work on the simplest version of the energy specification $S\{W_e\}$, a functional set written in shorthand, and to assert that it is possible to transform $S\{W_e\}$ from its "pre-generating functional form" to a standardized form $S\{\langle[\rho],\phi\rangle;\phi\in\mathscr{D}\}$ useful in electrostatics. In other words, one can always, at least formally, relate an appropriate energy specification with a type of potential that represents the laboratory sitiuation in which actual measurements are to be performed. Therefore it is essential that in the standardized form the energy is always expressed in terms of a test potential itself rather than in terms of its derivatives as may happen for "pre-generating functional forms."

It is proposed that one work according to the conceptual deduction scheme

$$W_e \Rightarrow \langle[\rho],\phi\rangle,$$

where the deduction is based on the invariance of energy

$$S\{W_e\} = S\{\langle[\rho],\phi\rangle,\phi\in\mathscr{D}\}; \tag{6}$$

$[\rho]$ represents a volume source density of the equivalent field. For simplicity of notation, the energy specification $S\{\langle[\rho],\phi\rangle;\phi\in\mathscr{D}\}$ will be designated as $\langle[\rho],\phi\rangle$, where ϕ should be considered an arbitrary element of a given space $\mathscr{D}$ of test potentials. An observation of fundamental importance is that the version of the multipole expansion developed in Ref. 1 (Eq. 22) makes it possible to express, for $l=0$, an arbitrary source body as the combination of energy-equivalent *volume* charge densities.[2]

For example, employing the definition of generalized derivative the expression (4) can be transformed as

$$\begin{aligned} W_e &= \int_{\Omega} \mathbf{P}\cdot\nabla\phi\, d\vartheta \Rightarrow \int \delta(\Omega)\mathbf{P}\cdot\nabla\phi d\vartheta \\ &= \langle\delta(\Omega)\mathbf{P},\nabla\phi\rangle = \langle -\nabla\cdot(\delta(\Omega)\mathbf{P}),\phi\rangle \\ &= \langle -\delta(\Omega)\nabla\cdot\mathbf{P} - \nabla\delta(\Omega)\cdot\mathbf{P},\phi\rangle = \langle[\rho],\phi\rangle \\ &= S\{W_e\}, \end{aligned} \tag{7}$$

where a volume charge density $[\rho]$ of the equivalent field is

$$[\rho] = \delta(\Omega)(-\nabla\cdot\mathbf{P}) - \nabla\delta(\Omega)\cdot\mathbf{P}. \tag{8}$$

The last formula corresponds to a well-known property that a volume dipole moment density distribution is equivalent to the volume charge density $\rho = -\nabla\cdot\mathbf{P}$ within the region Ω and the surface charge density on the boundary. Although in classical formulation one usually restricts oneself to presentation of the above property for volume-type distributions, it is quite legitimite to apply, formally, the same procedure to

other source distributions. Thus, for a dipole curve with a linear density $P_\alpha(T_\alpha)$, one obtains

$$W_e = \int_\alpha \mathbf{P}_\alpha(T_\alpha)(\nabla\phi)_{T_\alpha}\, dt$$
$$\Rightarrow \int \delta_\alpha(T-T_\alpha)\mathbf{P}_\alpha(T_\alpha)\cdot(\nabla\phi)\, d\vartheta$$
$$= \langle \delta_\alpha(T-T_\alpha)\mathbf{P}_\alpha, \nabla\phi\rangle = \langle -\nabla[\delta_\alpha(T-T_\alpha)\mathbf{P}_\alpha], \phi\rangle$$
$$= \langle [\rho], \phi\rangle = S\{W_e\}, \tag{9}$$

where

$$[\rho] = \delta_\alpha(T-T_\alpha)(-\nabla\cdot\mathbf{P}_\alpha) - \nabla\delta_\alpha\cdot\mathbf{P}_\alpha . \tag{10}$$

Similarly, for formula (3) one obtains

$$W_e = \sum_{k=1}^{n} \mathbf{p}_k\cdot(\nabla\phi)_{T_k} \Rightarrow \int \sum_{k=1}^{n} \delta(T-T_k)\mathbf{p}_k\cdot(\nabla\phi)\, d\vartheta$$
$$= \sum_{k=1}^{n} \langle \delta(T-T_k)\mathbf{p}_k, \nabla\phi\rangle$$
$$= \sum_{k=1}^{n} \langle -\nabla\cdot(\delta(T-T_k)\mathbf{p}_k), \phi\rangle$$
$$= \langle [\rho], \phi\rangle = S\{W_e\}, \tag{11}$$

where

$$[\rho] = \sum_{k=1}^{n} \delta(T-T_k)(-\nabla\cdot\mathbf{p}_k) - \sum_{k=1}^{n} \nabla\delta(T-T_k)\cdot\mathbf{p}_k , \tag{12}$$

and, using (5),

$$W_e = \int_\Omega \left(\sum_{\alpha,\beta} P_{\alpha,\beta} \frac{\partial^2\phi}{\partial\alpha\partial\beta} \right) d\vartheta \Rightarrow \mathbf{C}^{(2)}\langle \delta(\Omega)\mathbf{P}^{(2)}, \nabla^{(2)}\phi\rangle$$
$$= \mathbf{C}\langle -\nabla\cdot(\delta(\Omega)\mathbf{P}^{(2)}, \nabla\phi\rangle$$
$$= \mathbf{C}\langle -\delta(\Omega)\nabla\cdot\mathbf{P}^{(2)} - \nabla\delta(\Omega)\cdot\mathbf{P}^{(2)}, \nabla\phi\rangle$$
$$= \langle \nabla\cdot[-\delta(\Omega)\nabla\cdot\mathbf{P}^{(2)} - \nabla\delta(\Omega)\cdot\mathbf{P}^{(2)}], \phi\rangle$$
$$= \delta(\Omega)(\nabla^{(2)}\cdot\mathbf{P}^{(2)}) + 2\nabla\delta(\Omega)\cdot(\nabla\cdot\mathbf{P}^{(2)})$$
$$+ \nabla^{(2)}\delta(\Omega)^{(2)}\cdot\mathbf{P}^{(2)}, \phi\rangle$$
$$= \langle [\rho], \phi\rangle = S\{W_e\}; \tag{13}$$

$$[\rho] = \delta(\Omega)\nabla^{(2)}\cdot\mathbf{P}^{(2)} + \nabla\delta(\Omega)\cdot[2\nabla\cdot\mathbf{P}^{(2)}] + \nabla^{(2)}\delta(\Omega)^{(2)}\cdot\mathbf{P}^{(2)}. \tag{14}$$

Formulas (8), (10), (12), and (14) indicate a general character of the formalism and reveal existing regularities. The first terms in these formulas represent the interior of supports and associated volume charge densities, whereas the following terms represent volume charge densities that are equivalent to surface distributions at the boundary. Equation (14) appears to be particularly relevant to the philosophy used here. In that case the second term represents an equivalent dipole density $2\nabla\cdot P^{(2)}$ localized on the boundary $\nabla\delta(\Omega)$. the product $\nabla\delta(\Omega)\cdot 2\nabla\cdot\mathbf{P}^{(2)}$, however, should be interpreted as a volume charge density resulting from a nonuniform dipole distribution $2\nabla\cdot\mathbf{P}^{(2)}$. The third term represents an equivalent volume charge density due to a quadrupole distribution associated with a "dipole" surface $\nabla^{(2)}\cdot\delta(\Omega)$. It is interesting to note that all terms in the above equations are meaningful since source densities $\mathbf{P}$ are functions defined in the whole space. This remark is methodically important since usually the first term in Eq. (12), for example, is assumed to be zero. The energy specification formalism directly relates boundary effects with analytical properties of test potentials. Indeed, the nature of boundary effects depends upon what differential aspects of external potentials are taken into account (slope, curvature, and so on). In other words, one may maintain that the boundary effects included in $[\rho]$ and analytical properties of external feilds are, in a sense, "dual" phenomena.

The merits of the formalism may become more evident when a multipole of higher order is handled. Taking into account the structure of Eqs. (8)–(14), one easily finds that a general form of an equivalent volume charge density is now given by

$$[\rho] = \sum_k \delta(T-T_k)q_k + \delta_\alpha(T-T_\alpha)\tau(T)$$
$$+ \delta_s(T-T_s)\sigma(T) + \delta(\Omega)\rho_\Omega$$
$$- \nabla\cdot\left\{ \sum_k \delta(T-T_k)\mathbf{p}_k + \delta_\alpha\mathbf{P}_k + \delta_s\mathbf{P}_s + \delta(\Omega)\mathbf{P} \right.$$
$$\left. - \nabla\cdot\left[\left(\sum_k \delta(T-T_k)P_{\alpha,\beta;k} + \cdots + \delta(\Omega)P_{\alpha,\beta}\right) + \cdots\right]\right\}. \tag{15}$$

Assigning the symbols $\mathscr{P}^{(l)}$ for elementary source bodies of order l, Eq. (15) can be written

$$[\rho] = \mathscr{P}^{(0)} - \nabla\cdot\{\mathscr{P}^{(1)} - \nabla\cdot(\mathscr{P}^{(2)} - \cdots\}$$
$$= \mathscr{P}^{(0)} - \nabla\cdot\mathscr{P}^{(1)} + \nabla^{(2)}\cdot\mathscr{P}^{(2)} - \cdots$$
$$= \sum_{l=0}^{\infty} [\rho^{(l)}], \tag{16}$$

where $[\rho^{(l)}]$ represents an equivalent source body of order 0 corresponding to a given multipole source body of order l. Now, using the concept of energy specification of order 0, one can define the electrostatic potential Φ and, in turn, families of so-called characteristic and equivalent fields satisfying the multipole version of Maxwell's equation $\nabla\cdot\mathbf{D} = \mathscr{P}^{(0)}$.

Suppose that the electric potential Φ is a generalized function satisfying the energy specification equation

$$\langle -\epsilon_0\nabla^{(n)}\cdot\nabla^{(n)}\Phi, \phi\rangle = S\{W_e\} = \langle [\rho], \phi\rangle . \tag{17}$$

Equation (17) has the same form as Poisson's equation except for the fact that additional iterations of the operator div(grad) are considered. Defining the generalized vector function $\mathbf{E}$ by $\mathbf{E} = -\nabla\Phi$, Eq. (17) can be written

$$\epsilon_0\nabla^{(n)}\cdot\nabla^{(n-1)}\mathbf{E} = [\rho], \tag{18}$$

and, using (18) and (16), one obtains

$$\langle \epsilon_0\nabla^{(n)}\cdot\nabla^{(n-1)}\mathbf{E}, \phi\rangle$$
$$= \langle \mathscr{P}^{(0)} - \nabla\mathscr{P}^{(1)} + \cdots + (-1)^n\nabla^n$$
$$\cdot[\mathscr{P}^{(n)} - \nabla\cdot\mathscr{P}^{(n+1)} + \cdots], \phi\rangle$$
$$= \langle \nabla^{(n)}\cdot\{\epsilon_0\nabla^{(n-1)}\mathbf{E} + (-1)^{n+1}\mathscr{P}^{(n)}$$
$$+ (-1)^{n+2}\nabla\cdot\mathscr{P}^{(n+1)} + \cdots\}, \phi\rangle$$
$$= \langle \mathscr{P}^{(0)} - \nabla\mathscr{P}^{(1)} + \cdots + (-1)^{n-1}\nabla^{(n-1)}\cdot\mathscr{P}^{(n-1)}, \phi\rangle,$$

or

$$\nabla^{(n)}\cdot[\epsilon_0\nabla^{(n-1)}\mathbf{E} + (-1)^{n+1}\mathscr{P}^{(n)} + \cdots]$$
$$= \mathscr{P}^{(0)} - \nabla\cdot\mathscr{P}^{(1)} + \cdots + (-1)^{n-1}\nabla^{(n-1)}\cdot\mathscr{P}^{(n-1)}. \tag{19}$$

Equation (19) clearly displays the nature of a "multipole

approximation" of order n and demonstrates how to apply the energy specification equation (17) to each specific electrostatic problem. Hence, a formal separation between the so-called true and bounded sources is a direct consequence of the energy specification equation (17). Indeed, assuming $n = 1$, one obtains the classical Lorentz dipole approximation. Equation (17) becomes the Poisson equation

$$\langle -\epsilon_0 \Delta\Phi,\phi\rangle = S\{W_e\} = \langle [\rho],\phi\rangle . \tag{20}$$

Equation (20) can also be written as

$$\begin{aligned} S\{W_e\} &= \langle [\rho],\phi\rangle = \langle \mathscr{P}^{(0)} - \nabla\mathscr{P}^{(1)} + \cdots,\phi\rangle \\ &= \langle -\epsilon_0\Delta\Phi,\phi\rangle = \langle \epsilon_0\nabla\Phi,\nabla\phi\rangle \\ &= \langle -\epsilon_0\mathbf{E},\nabla\phi\rangle = \langle \epsilon_0\nabla\cdot\mathbf{E},\phi\rangle \end{aligned} \tag{21}$$

or

$$\nabla\cdot(\epsilon_0\mathbf{E}) = \mathscr{P}^{(0)} - \nabla\cdot\mathscr{P}^{(1)} + \cdots . \tag{22}$$

On the other hand, for $n = 1$, Eq. (19) becomes

$$\nabla\cdot(\epsilon_0\mathbf{E} + \mathscr{P}^{(1)} - \nabla\cdot\mathscr{P}^{(2)} + \cdots) = \mathscr{P}^{(0)} \tag{23}$$

or

$$\nabla\cdot\mathbf{D} = \mathscr{P}^{(0)}, \tag{24}$$

where $\mathbf{D}$ is a generalized displacement vector. To describe approximations of higher orders, it is convenient to introduce the family D_n satisfying

$$\nabla^{(n)}\cdot\mathbf{D}_n = \mathscr{P}^{(0)} - \nabla\cdot\mathscr{P}^{(1)} + \cdots + (-1)^{n-1}\cdot\mathscr{P}^{(n-1)}. \tag{25}$$

It is seen that in the "quadrupole approximation" the displacement vector D_2 is defined by charge distributions $\mathscr{P}^{(0)}$ and equivalent dipole distributions $\mathscr{P}^{(1)}$. Similarly, for $n = 3$, Eq. (25) becomes

$$\nabla^{(3)}\cdot D_3 = \mathscr{P}^{(0)} - \nabla\cdot\mathscr{P}^{(1)} + \nabla^{(2)}\cdot\mathscr{P}^{(2)} \tag{26}$$

and (18) becomes

$$\epsilon_0\nabla^{(3)}\cdot\nabla^{(2)}\mathbf{E} = [\rho]. \tag{27}$$

It is interesting to note that for every n all sources are necessary to determine the vector$\mathbf{E}$. Since $[\rho]$ is a sum of all equivalent source bodies of the system, the field $\mathbf{E}$ can be interpreted as an equivalent field that is generated by all sources for any multipole approximation. At this point a certain clarification seems to be proper. The field $\mathbf{E}$ in Eq. (22) is a generalized function and thereby it depends upon a given family of external fields. This dependence on the auxiliary fields is crucial to recognize the structure of the formalism. Using an analogy with quantum theory one may assert that each individual distribution of external sources determines the experimental arrangements that can be set up for the system, just as the wavefunction in quantum mechanics determines the state of a system. A measurement made on the system yields a real number which can be interpreted as the interaction energy between a given source distribution $[\mathbf{P}]$ and an auxiliary testing source. The results of independently performed experiments obtained for all possible auxiliary fields within a given family define the energy specification $S\{W_e\}$. Subsequently, the equivalent field $\mathbf{E}$ due to $[\mathbf{P}]$ is determined in terms of the energy specification $S\{W_e\}$ rather than directly by means of sources as is the case in traditional field theory. Similar remarks apply to vectors $\mathbf{D}_n$.

A different structure shown in Eq. (25) indicates, however, that vectors $\mathbf{D}_n$ depend explicitly on the order n of the multipole approximations. Indeed, it can be seen that they are determined by the energy specification of all multipole densities of orders less than the order of a given approximation. In this sense, the vectors $\mathbf{D}_n$ represent the characteristic fields of the multipole approximation. These results offer a simple and consistent interpretation of so-called bound or polarization charges. In view of Eqs. (18) and (25), polarization sources can be defined as a difference between equivalent and characteristic sources. As is shown above, the character of polarization sources depends upon the order of the multipole approximation. Specifically, in the classical dipole approximation free charges are the sources of the characteristic field $\mathbf{D}$ and polarization charges are the sources that should be added to free charges in order to obtain sources of the equivalent field $\mathbf{E}$.

III. ARBITRARY CURRENT DISTRIBUTION AS A VOLUME CURRENT DENSITY

Infinite values of energies for idealized sources in a magnetostatic field can be avoided by the introduction of appropriate magnetic energy specifications $S\{W_m\}$.[6] Consider a current loop localized in an external magnetic flux Ψ_m. The classical expression for the magnetic energy is given by

$$W_m = i\Psi_m = i\oint_\alpha \mathbf{dl}\cdot\mathscr{A} , \tag{28}$$

where $\mathscr{A}$ is a corresponding vecotr test potential. In the analogous way, for an arbitrary volume current density $\mathbf{j}(T)$ one can assign a real number

$$W_m = \int_\Omega \mathbf{j}(T)\cdot\mathscr{A}(T)\, d\vartheta, \tag{29}$$

which corresponds to the magnetic energy of a current loop in an external field. The set of all numbers obtained through (28) or (29) for a given family of vector test potentials defines the linear functional $S\{W_m\}$. It will be referred to, in the following, as the magnetic energy specification. Two magnetic sources are considered to be equivalent if their magnetic energy specifications are identical. It can be shown that the concept of generalized derivative makes it possible to transform the functional set $S\{W_m\}$ of a system to the standardized from $\langle[\mathbf{j}],\mathscr{A}\rangle$. In fact, the scheme of conceptual deduction $W_m \Rightarrow \langle[\mathbf{j}],\mathscr{A}\rangle$ based on the invariancy condition

$$S\{W_m\} = S\{\langle[\mathbf{j}],\mathscr{A}\rangle;\ \mathscr{A}\in\mathscr{D}\rangle \tag{30}$$

may be applied where, as previously, the symbol $\langle[\mathbf{j}],\mathscr{A}\rangle$ will be understood as $S\{\langle[\mathbf{j}],\mathscr{A}\rangle,\mathscr{A}\in\mathscr{D}\}$. For example, Eqs. (28) and (29) become

$$\begin{aligned} i\oint_\alpha \mathbf{dl}\cdot\mathscr{A} &= i\int \boldsymbol{\delta}_i\cdot\mathscr{A}\, d\vartheta \Rightarrow \langle\boldsymbol{\delta}_\alpha i,\mathscr{A}\rangle \\ &= \langle[\mathbf{j}],\mathscr{A}\rangle = S\{W_m\}, \end{aligned} \tag{31}$$

$$\int_\Omega \mathbf{j}\cdot\mathscr{A}\, d\vartheta = \int \delta(\Omega)\mathbf{j}\cdot\mathscr{A}\, d\vartheta = \langle[\mathbf{j}],\mathscr{A}\rangle = S\{W_m\}. \tag{32}$$

The energy specification of a magnetic dipole can be obtained directly from Eq. (28). Indeed, using Stoke's theorem

one obtains

$$\langle \delta_\alpha i, \mathscr{A} \rangle = \langle \nabla \times \delta_s i, \mathscr{A} \rangle = \langle \delta_s i, \nabla \times \mathscr{A} \rangle. \tag{33}$$

Let $S \to 0$ and $i \to \infty$ in Eq. (33) so that the product iS remains constant. In such a case the loop becomes a magnetic dipole having the moment $\mathbf{m} = \mathbf{e}_m$ $(\lim iS)$, where $\mathbf{e}_m$ is the unit vector normal to S and $\delta_s i \to \delta_m$. Hence the energy specification $S\{W_m\}$ of magnetic dipoles localized at the points $T_1, \ldots, T_n$ and having magnetic moment $\mathbf{M} = \Sigma \delta(T - T_k)\mathbf{m}_k$ is given by

$$W_m \Rightarrow \left\langle \sum_{k=1}^{n} \delta(T - T_k \mathbf{m}_k, \nabla \times \mathscr{A} \right\rangle$$
$$= \left\langle \nabla \times \left(\sum_{k=1}^{n} \delta(T - T_k)\mathbf{m}_k \right), \mathscr{A} \right\rangle = \langle [\mathbf{j}], \mathscr{A} \rangle, \tag{34}$$

where the volume current density of the equivalent field is

$$[\mathbf{J}] = \sum_{k=1}^{n} \nabla \times (\delta(T - T_k)\mathbf{m}_k)$$
$$= \sum_{k=1}^{n} \delta(T - T_k)(\nabla \times \mathbf{m}_k) + \sum_{k=1}^{n} \nabla \delta \times \mathbf{m}_k \,. \tag{35}$$

Similarly, for a volume magnetic moment density distributed within a bounded region Ω the energy specification takes the form

$$W_m \Rightarrow = \langle \delta(\Omega)\mathbf{M}, \nabla \times \mathscr{A} \rangle = \langle \nabla \times (\delta(\Omega)\mathbf{M}), \mathscr{A} \rangle$$
$$= \langle [\mathbf{j}], \mathscr{A} \rangle = S\{W_m\},$$

where

$$[\mathbf{j}] = \delta(\Omega)\nabla \times \mathbf{M} + \nabla \delta(\Omega) \times \mathbf{M}. \tag{36}$$

In a similar fashion one finds, for a surface magnetic dipole moment density

$$W_m \Rightarrow \langle \nabla \times (\delta_s (T - T_s)\mathbf{M}_s), \mathscr{A} \rangle = \langle [\mathbf{j}], \mathscr{A} \rangle = S\{W_m\},$$

where

$$[\mathbf{j}] = \delta_s \nabla \times \mathbf{M}_s + \nabla \delta_s (T - T_s) \times \mathbf{M}_s \,. \tag{37}$$

The first term in Eq. (37) represents a surface current density and the second term corresponds to the linear current density on the boundary α of the surface S. Suppose that the generalized function of a vector potential $\mathbf{A}$ and a generalized magnetic field vector $\mathbf{B}$ are defined by the energy specification equation

$$1/\mu_0 \langle \nabla \times (\nabla \times \mathbf{A}), \mathscr{A} \rangle = S\{W_m\} = \langle [\mathbf{j}], \mathscr{A} \rangle \tag{38}$$

and assume that

$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{39}$$

Equations (34)–(37) indicate that an equivalent current density $[\mathbf{j}]$ can be written in the form

$$[\mathbf{j}] = [\mathbf{j}^{(0)}] + [\mathbf{j}^{(1)}] = [\mathbf{j}^{(0)}] + \nabla \times \mathscr{M}$$
$$= [j^{(0)}] + \nabla \times \left[\sum_k \delta(T - T_k)\mathbf{m}_k + \delta_\alpha (T - T_\alpha)\mathbf{M}_\alpha \right.$$
$$\left. + \delta_s \mathbf{M}_s + \delta(\Omega)\mathbf{M}_\Omega \right]. \tag{40}$$

Substituting Eqs. (39) and (40) into (38) yields

$$1/\mu_0 \langle \nabla \times \mathbf{B}, \mathscr{A} \rangle = \langle [\mathbf{j}^{(0)}] + \nabla \times \mathscr{M}, \mathscr{A} \rangle = \langle [\mathbf{j}], \mathscr{A} \rangle \tag{41}$$

or

$$1/\mu_0 \langle \nabla \times (\mathbf{B} - \mu_0 \mathscr{M}), \mathscr{A} \rangle = \langle [\mathbf{j}^{(0)}], \mathscr{A} \rangle, \tag{42}$$

where $\mathbf{B} - \mu_0 \mathscr{M} = \mathscr{H}$ is referred to as the magnetic field intensity.

This demonstrates that the vectors $\mathbf{B}$ and $\mathscr{H}$ represent, respectively, an equivalent and characteristic field with respect to the energy specification Eq. (30).

IV. CONCLUSIONS REMARKS

The formalism developed in this paper makes it possible to (a) express an arbitrary source in the form of an equivalent volume source density, (b) achieve a significant simplification and uniformity of traditional formulas, and (c) provide opportunities for new interesting physical interpretations. For example, in a time-variant case the electric test field in Lorentz's approximation assumes the form $\mathbf{E}_t = -\nabla\phi - \partial \mathscr{A}/(\partial t)$. A total polarization charge-current can be found directly from the energy specification $S\{W_e\} + S\{W_m\}$. Indeed, for a volume dipole moment density $\mathbf{P}$ one obtains

$$\langle \delta(\Omega)\mathbf{P}, \mathbf{E}_t \rangle = \langle \delta(\Omega)\mathbf{P}, -\nabla\phi - \partial \mathscr{A}/\partial t \rangle$$
$$= \langle \delta(\Omega)\nabla \cdot \mathbf{P} + \nabla \delta \cdot \mathbf{P}, \phi \rangle + \langle \delta(\Omega)\partial \mathbf{P}/\partial t, \mathscr{A} \rangle$$
$$= \langle [\rho], \phi \rangle + \langle [\mathbf{j}], \mathscr{A} \rangle = S\{W_e\} + S\{W_m\},$$

where $[\rho]$ and $[\mathbf{j}]$ denote, respectively, equivalent charges and currents. The result is consistent with classical theory, as it should be. It is hardly surprising that classical integral formulas are phrased succintly in terms of the energy specification $S\{W_e\}$ and $S\{W_m\}$. As a matter of fact, the distributional approach simplifies an integration by parts and this is the property that makes the generalized derivative a very convenient tool with which to deal. In particular, the rigorous description of the so-called polarization charge-currents, in view of the "additive" version of the multipole expansion, seems to be expected. After all, polarization sources are a part of the physical description of a problem. However, if one examines the mathematical structure of the energy specification formalism, a desired linkage between between the physical and mathematical aspects is now available in rigorous analytical form.

It should be emphasized that the formulas representing energy integrals can be viewed as a bilinear form-type formalism in which both topological properties of sources and analytical characterization of probing agents are considered simultaneously. It is believed that such an approach within classical electrodynamics is conceptually closer to formalisms used in quantum theory and abstract field theory.

ACKNOWLEDGMENTS

The author is indebted to Professor Smio Tani for many valuable discussions concerning the presentation of the work and suggested improvements in notation. He also thanks Professor Jan K. Sedivy for discussions and encouragement and the referee of the journal for suggestions.

[1] S. J. Dolzycki, J. Math. Phys. **22**, 320 (1981).
[2] See footnote 13 in Ref. 1.

[3]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), p. 145.
[4]N. Bleinstein and J. K. Cohen, J. Math. Phys. **18**, 194 (1977).
[5]This point has been suggested by S. Tani. For classical mechanics, the same kind of point of view has been stated before, see, e.g., C. Lanczos, *The Variational Principles of Mechanics*, 4th ed. (University of Toronto Press, Toronto, 1970), Chaps. 1 and 2.
[6]A. Skopec, Technical University of Wroclaw, Poland, preprint (May 1974) (unpublished).

An example of an $\mathscr{H}$-space [a)]

G. A. J. Sparling
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260
K. P. Tod
Mathematical Institute, St. Giles, Oxford, England

(Received 8 May 1978; accepted for publication 14 Dcember 1979)

The good cut equation for a specific asymptotic shear is solved and the metric of the associated $\mathscr{H}$-space is obtained. The $\mathscr{H}$-space is found to be type N, asymptotically flat and positive frequency.

PACS numbers: 04.20.Cv

Since $\mathscr{H}$-space was introduced by Newman,[1] the subject has been developed in a series of papers.[2,3,4] However one stumbling block up to now has been the absence of any specific nontrivial examples of $\mathscr{H}$-spaces. It is our purpose here to describe such an example which leads to a left-flat space of some interest in its own right, and which provides an arena for the testing of future developments in the subject.

An $\mathscr{H}$-space is defined as the space of asymptotically shear-free cuts (good cuts) of the complexified future null infinity $\mathbb{C}\mathscr{I}^+$ of an asymptotically flat space–time $\mathscr{M}$.[1] Equivalently, the $\mathscr{H}$-space of $\mathscr{M}$ is the space of solutions of the good cut equation,[1]

$$\eth^2 Z(\zeta,\tilde{\zeta}) = \sigma^0(Z,\zeta,\tilde{\zeta}), \tag{1.1}$$

which are regular on the entire Riemann sphere of ζ when $\tilde{\zeta} = \bar{\zeta}$. Here $\sigma^0(u,\zeta,\tilde{\zeta})$ is the asymptotic shear of an outgoing Bondi family of null hypersurfaces in $\mathscr{M}$ (see Refs. 1, 2 for definitions).

This solution space is known to be four-dimensional[3] so that $\mathscr{H}$-space is a four-dimensional complex manifold. In terms of the solution of (1.1) there is a natural definition of a quadratic metric on the $\mathscr{H}$-space which automatically gives rise to a self-dual curvature tensor.

No solutions of (1.1) for nonzero σ^0 have hitherto been found, except for the essentially trivial case when σ^0 is linear in u. In this paper we solve (1.1) and obtain the metric of $\mathscr{H}$-space in the particular case

$$\sigma^0(u,\zeta,\tilde{\zeta}) = \lambda/[u^3(1+\zeta\tilde{\zeta})^2]. \tag{1.2}$$

Here λ is a parameter for bookkeeping purposes and will turn out to measure the "strength" of the curvature.

While σ^0 given by (1.2) is singular at zero u (so that a space–time whose asymptotic shear was this would have a singularity in its radiation field) this singularity is easily removed by a complex translation in u. We shall return to this point, but for ease of calculation we work with (1.2).

The good cut equation becomes

$$\eth^2 Z \equiv ((1+\zeta\tilde{\zeta})^2 Z_{,\zeta})_{,\tilde{\zeta}} = \lambda/[Z^3(1+\zeta\tilde{\zeta})^2]. \tag{1.3}$$

This may be directly integrated, the constants of integration being fixed by the regularity requirement, but it is easier to make the ansatz (with hindsight)

$$(Z(z^a,\zeta,\tilde{\zeta}))^2 = z^2 + \lambda s^2,$$
$$z = z^a l_a(\zeta,\tilde{\zeta}),$$
$$s = s^a l_a(\zeta,\tilde{\zeta}), \tag{1.4}$$
$$z^a = (u,X,Y,v),$$
$$l_a = \frac{1}{1+\zeta\tilde{\zeta}}(1,\zeta,\tilde{\zeta},\zeta\tilde{\zeta}).$$

The z^a are the four parameters on which the solution depends (and thus are coordinates on $\mathscr{H}$-space) and the good cut equation (1.3) serves to fix the four s^a in terms of the z^a.

Substituting (1.4) into (1.3) leads to

$$z\eth s - s\eth z = 1/(1+\zeta\tilde{\zeta}), \tag{1.5}$$

which gives three conditions on the s^a. To fix s^a uniquely, we recall that

$$\eta^{ab} l_a l_b = 0,$$

where η^{ab} is the Minkowski metric in null coordinates:

$$\eta^{ab} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} = \eta_{ab}.$$

Thus in (1.4) the trace

$$\eta_{ab}(z^a z^b + \lambda s^a s^b),$$

is undetermined and we may choose s^a such that

$$\eta_{ab} s^a s^b = 0. \tag{1.6}$$

Then (1.6) and (1.5) lead to

$$s^a = \frac{1}{uv - XY}(Y,v,0,0), \tag{1.7}$$

and (1.7) with (1.4) provide the full solution of the good cut equation.

The general methods of Refs. 2 or 3 may be used to obtain the metric from $Z(z^a,\zeta,\tilde{\zeta})$ but it is quicker to use the methods of the Penrose Twistor Theory.[5] To define the conformal metric first, suppose two infinitesimally separated points of $\mathscr{H}$-space to have coordinates z^a and $z^a + dz^a$ with corresponding good cuts

$$u = Z(z^a,\zeta,\tilde{\zeta}),$$
$$u + du = Z(z^a,\zeta,\tilde{\zeta}) + dZ(z^a,\zeta,\tilde{\zeta}),$$

where

$$dZ = dz^a Z_{,a},$$

then the condition that dz^a be a null displacement[3,5] is that du vanish along a curve of constant $\tilde{\zeta}$, i.e., that one can find a value $\tilde{\zeta}_0$ of $\tilde{\zeta}$ such that

[a)]Work supported in part by NSF grant.

 0022-2488/81/020331-02$1.00

$$du \equiv dz^a Z_{,a}(z^b, \zeta, \tilde{\zeta}_0) = 0, \tag{1.8}$$

for all ζ. From (1.4) with an obvious notation this means

$$z\,dz + \lambda s\,ds = 0. \tag{1.9}$$

The left-hand side of (1.9) is quadratic in ζ, giving three conditions involving z^a, dz^a, and $\tilde{\zeta}_0$. The object now is to eliminate $\tilde{\zeta}_0$ and obtain a quadratic relation among the dz^a. This calculation is greatly facilitated by the Penrose "blob" notation[6] with the result

$$2du\,dv - 2dX\,dY - \frac{2\lambda}{(uv - XY)^3}(Y\,dv - v\,dY)^2 = 0, \tag{1.10}$$

as the condition for dz^a to be null. We now observe that the metric

$$ds^2 = 2du\,dv - 2dX\,dY - \frac{2\lambda}{(uv - XY)^3}(Y\,dv - v\,dY)^2, \tag{1.11}$$

has Plebański's second form for the general left-flat metric[7] if one makes the identifications

$$p = v, \quad q = Y, \quad x = u, \quad y = X,$$

$$\Theta = \frac{1}{2(uv - XY)} = \frac{1}{2(px - qy)}. \tag{1.12}$$

Further, Θ satisfies Plebański's second "heavenly" equation

$$\Theta_{,px} - \Theta_{,qy} - (\Theta_{,xx}\Theta_{,yy} - (\Theta_{,xy})^2) = 0.$$

Thus (1.11) is already a left-flat metric and the $\mathscr{H}$-space metric which we are seeking can differ from (1.11) by at most a constant conformal factor. Twistor methods may again be used to find that this constant is unity, so that (1.11) is the $\mathscr{H}$-space metric arising from the shear (1.2).

Since (1.11) is in Kerr-Schild form, the curvature must be algebraically special. The simplest way to calculate the curvature is to use the results of Plebański[7] and one then finds that the $\mathscr{H}$-space is type N. Further the curvature is nonsingular everywhere except on the surface

$$uv - XY = 0. \tag{1.13}$$

Again, since (1.11) is in Kerr–Schild form, there is a canonical Minkowski background in which to discuss properties of the metric. Regarded as an "already linearized" solution on this background, the solution is one of Penrose's elementary states[8] and is singular on the light-cone of the origin (1.13). It is possible to deal with this singularity and the singularity in σ^0 simultaneously. If in (1.2) we make a translation

$$u \mapsto u - it^a l_a(\zeta, \tilde{\zeta}), \tag{1.14}$$

where l_a is as in (1.4) and t^a is timelike and future-pointing with respect to η^{ab}, then σ^0 becomes holomorphic in the lower half of the complex u-plane, i.e., positive frequency. With

$$\sigma^0 = \frac{\lambda}{(u - it^a l_a)^3(1 + \zeta\tilde{\zeta})^2},$$

we transform the Z of (1.5) by

$$Z \mapsto Z + it^a l_a(\zeta, \tilde{\zeta}), \tag{1.15}$$

to obtain a solution of the new good cut equation. With (1.15), the limit Z_0 of Z for zero λ is

$$Z_0 = (z^a + it^a) l_a(\zeta, \tilde{\zeta}),$$

so it is natural to transform to new coordinates on the $\mathscr{H}$-space given by

$$z'^a = z^a + it^a,$$

so that

$$Z_0 = z'^a l_a(\zeta, \tilde{\zeta}).$$

This is just a translation of the background Minkowski space, with the result that the curvature is nonsingular away from the (background) light-cone of the point

$$z'^a = it^a.$$

In particular, this means that the curvature is nonsingular in the (background) past-tube in primed coordinates and we may take this as a definition of positive frequency for the Weyl tensor in this case. There is as yet no definition of positive frequency in a general curved space but any definition will presumably reduce to this one in this case. This leads to the conjecture that a positive frequency σ^0 gives rise to a positive frequency $\mathscr{H}$-space.

With the identification of (1.11) as an elementary state, we see that it is asymptotically flat. (At least on the real section in the primed coordinates, where the usual definition of asymptotic simplicity is applicable. The question of whether this solution is asymptotically flat according to the general definition of Ref. 4 is currently under investigation.) We are thus led to conjecture in general that asymptotic flatness of the $\mathscr{H}$-space arises from appropriate large u behavior of σ^0, specifically that

$$\sigma^0 \sim u^{-3}, \quad \text{large } |u|,$$

for asymptotic flatness.

We conclude by remarking that a number of generalizations of the solution presented here are possible, giving solutions of the good cut equation for other shears and $\mathscr{H}$-spaces of other algebraic types.

[1] E.T. Newman, in *General Relativity and Gravitation*, edited by G. Shaviv and J. Rosen (Wiley, New York, 1975).
[2] M. Ko, E.T. Newman and K.P. Tod, in *Asymptotic Structure of Space-Time*, edited by F.P. Esposito and L. Witten (Plenum, New York, 1976).
[3] R.O. Hansen, E.T. Newman, R. Penrose, and K.P. Tod, Proc. R. Soc. (Lond) A **363**, 445 (1978).
[4] M. Ko, M. Ludvigsen, E.T. Newman, and K.P. Tod, to appear in Phys. Rep.
[5] R. Penrose, G.R.G. 7, 31 (1976).
[6] Z. Perjes, Phys. Rev. D **11**, 2031 (1975).
[7] J. Plebański, J. Math. Phys. **16**, 2395 (1975).
[8] R. Penrose and M.A.H. MacCallum, Phys. Rep. C **6**, 242 (1973).

Space-time algebra approach to curvature

G. Sobczyk
Institute for Theoretical Physics, University of Wroclaw, Cybulskiego 36, Wroclaw, Poland

(Received 6 December 1978; accepted for publication 17 July 1979)

Introducing the notions of vector and bivector differentiation into the Dirac algebra, considered as a Clifford algebra, makes possible an extremely concise and geometrically transparent treatment of the curvature tensor and its properties, and of related topics such as Lorentz invariants, characteristic equations, Petrov types, and principal null directions by explicit construction.

PACS numbers: 04.20.Cv, 02.40,Ky, 02.40.Re

INTRODUCTION

The Riemann curvature tensor is the crucial geometric ingredient in the study of general relativity. It is for this reason that the curvature tensor and its properties and invariants have attracted wide attention. The original classification of the curvature tensor for empty space was carried out by Petrov[1] using matrix methods. Subsequently, a number of different methods and refinements have been introduced. Noteworthy of mention is the spinor approach used by Witten,[2] and later refined by Penrose[3] in his systematic study of the coincidence patterns of the four principal null directions. But, as anyone who is familiar with calculations with spinors knows, these methods are only adapted to certain kinds of problems. Classical tensor methods have also been used with some success, for example,[4] but the computational aspects of this approach are formidable. Thorpe[5] notes that computations are considerably simplified by using the Hodge star operator to make the space of bivectors into a complex Euclidean space, but he ignores the possibility of utilizing the underlyng Lie algebra of bivectors. Stehney[6] modifies Thorpe's approach to the requirements of matrix methods and produces a classification scheme based on the minimal polynomial of a complex 3×3 matrix, but her methods lack conceptual clarity, and her algorithm works only for repeated principal null directions.

The purpose of the present work is to cover much the same ground as the above authors, but in a coordinate-free formalism whose power, simplicity, and geometric transparency have yet to be recognized; a formalism which has all the advantages of each of the above mentioned approaches, and the defects of none.

In Sec. 1, following Hestenes,[7,8] we introduce the 16-dimensional Clifford algebra called the *Dirac algebra* of space-time in agreement with the name given its matrix representation. (Clifford algebra of 2^n-dimensions has been extensively developed in the book, *Clifford Algebras and Geometric Calculus: A Unified Language for Mathematics and Physics*,[9] using an abstract approach,[10] rather than a matrix representation such as is used by Cartan,[11] and others). The even subalgebra, consisting of scalars, bivectors, and pseudoscalars of the Dirac algebra, make up the *Pauli algebra* of space. The Pauli algebra can be fruitfully compared to the popular Gibbs-Heaviside vector algebra, because many identities of the former are the "complexified" versions of the latter. A discussion of bivectors and null bivectors is given, and a multiplication table of basis elements is included.

In Sec. 2 we complement the algebraic machinery introduced in Sec. 1 by introducing the operations of vector and bivector differentiation. These operations simplify and generalize the operation of contraction in tensor algerbra. They were originally developed as a coordinate-free tool for the study of linear transformations in Ref. 12.

In Sec. 3 we study general properties of linear operators on the space of bivectors by decomposing it into the sum of dual and antidual operator parts. A dual operator is equivalent to a general linear operator on a complex three-dimensional Euclidean space. Using the new method of bivector derivatives, the determinant, characteristic polynomial and Cayley–Hamilton theorem are derived for dual operators. In our approach it is unnecessary to introduce the Hodge-star operator, because in the Dirac algebra duality is simply expressed by multiplication by the unit pseudoscalar element. Finally, we show that an antidual bivector operator can be expressed entirely in terms of two symmetric trace-free vector operators. In another paper[13] we show how the problem of the classification of these symmetric vector operators is directly correlated to the Petrov classification.

In Sec. 4 we give a complete classification of dual operators based on explicit construction of their principal null bivectors. The classification of a dual skew-symmetric operator is equivalent to the classification of an electromagnetic field by its principal null directions. A dual symmetric operator with vanishing trace is equivalent to the conformal curvature tensor. The Petrov–Penrose classification of dual symmetric operators is carried out by construction of its four principal null bivectors, based on a new canonical form involving a complex scalar, a bivector, and a null eigenbivector. This new canonical form provides a simple geometric criterion for the various coincidence patterns of the four principal null directions. In addition, it makes it trivial to give simple examples of conformal curvature tensors of any desired type.

In Sec. 5, curvature invariants, which are complex scalars, are defined in terms of the bivector derivative, and it is shown that a curvature operator has nine complex scalars, three of which are real. When the extra Bianchi identity is imposed, these 15 real scalars reduce down to the well

[a]Research supported by NSF grant ϕ GF41959 through an exchange program with State University of New York at Stony Brook.

 0022-2488/81/020333-10$1.00

known 14 real invariants of the Riemann curvature tensor. Various well known properties and identities of Riemann curvature are then derived in the spacetime algebra (STA) formalism. In each case the simplicity and geometric transparency of our methods are apparent. A table is included comparing the appearance of well known formulas in the tensor, STA, and spinor formalism. We believe a close examination of this table and the methods of this paper will show the judgment of Misner, Thorne, Wheeler[14] (p. 1165) that "the spinor formalism is a more powerful method than any other for deriving the Petrov–Pirani algebraic classification of the conformal curvature tensor, and for proving theorems about algebraic properties of curvature tensors," needs reexamination. See also Ref. 13.

1. SPACE-TIME ALGEBRA

Let x be a generic point in spacetime. Following Hestenes,[7] we select a set of orthonormal vectors e_0, e_1, e_2, e_3 tangent to the point x, and subject them to the rules:

$$e_0^2 = 1, e_1^2 = e_2^2 = e_3^2 = -1, \tag{1.1}$$

$$e_u e_v = -e_v e_u \quad \text{for} \quad u,v = 0,1,2,3 \quad \text{and} \quad u \neq v. \tag{1.2}$$

The orthonormal vectors $\{e_u\}$, under the rules for geometric multiplication (1.1) and (1.2) generate a real Clifford Algebra of $2^4 = 16$ dimensions called the *Dirac Algebra* $\mathscr{D}$ in agreement with the name given its matrix representation. Symbollically we write $\mathscr{D} = \mathscr{D}_0 + \mathscr{D}_1 + \mathscr{D}_2 + \mathscr{D}_3 + \mathscr{D}_4$, to express the Dirac algebra $\mathscr{D}$ as the sum of linear subspaces of scalars, vectors, bivectors, trivectors, and pseudoscalars, respectively.

For purposes of orientation and fixing the notation that will be used here, let us review some of the basic operations and identities in $\mathscr{D}$. Let a, b be vectors in $\mathscr{D}_1$,

$$a = \alpha^u e_u \equiv \sum_{u=0}^{3} \alpha^u e_u, \quad b = \beta^u e_u,$$

then

$$a\cdot b \equiv \tfrac{1}{2}(ab + ba) = \alpha^0\beta^0 - \alpha^1\beta^1 - \alpha^2\beta^2 - \alpha^3\beta^3 \equiv g(a,b) \tag{1.3}$$

and

$$\begin{aligned} a \wedge b \equiv{}& \tfrac{1}{2}(ab - ba) \\ ={}& \begin{vmatrix} \alpha^1 & \alpha^0 \\ \beta^1 & \beta^0 \end{vmatrix} e_1 \wedge e_0 + \begin{vmatrix} \alpha^2 & \alpha^0 \\ \beta^2 & \beta^0 \end{vmatrix} e_2 \wedge e_0 \\ &+ \begin{vmatrix} \alpha^3 & \alpha^0 \\ \beta^3 & \beta^0 \end{vmatrix} e_3 \wedge e_0 + \begin{vmatrix} \alpha^3 & \alpha^2 \\ \beta^3 & \beta^2 \end{vmatrix} e_3 \wedge e_2 \\ &+ \begin{vmatrix} \alpha^1 & \alpha^3 \\ \beta^1 & \beta^3 \end{vmatrix} e_1 \wedge e_3 + \begin{vmatrix} \alpha^2 & \alpha^1 \\ \beta^2 & \beta^1 \end{vmatrix} e_2 \wedge e_1. \end{aligned} \tag{1.4}$$

From the definitions (1.3) and (1.4), it is clear that

$$ab = \tfrac{1}{2}(ab + ba) + \tfrac{1}{2}(ab - ba) = a\cdot b + a \wedge b, \tag{1.5}$$

i.e., the geometric product of two vectors can be decomposed into the sum of an inner product or (real) scalar part, and an outer product or bivector part. The metric tensor $g(a, b)$ of spacetime is determined by the inner product and is, of course, invariant under local Lorentz transformations.

Define the bivectors

$$E_1 = e_1 \wedge e_0 = e_1 e_0, \quad E_2 = e_2 \wedge e_0, \quad E_3 = e_3 \wedge e_0, \tag{1.6}$$

$$E_4 = e_3 \wedge e_2 = e_3 e_2, \quad E_5 = e_1 \wedge e_3, \quad E_6 = e_2 \wedge e_1.$$

The unit pseudoscalar I, defined by

$$I = e_0 \wedge e_1 \wedge e_2 \wedge e_3 = e_0 e_1 e_2 e_3 = E_1 E_2 E_3, \tag{1.7}$$

has the property $I^2 = -1$, and assigns a unique orientation to the Dirac algebra $\mathscr{D}$. The duality of the bivectors E_1, E_2, E_3 and E_4, E_5, E_6 has the simple algebraic expression

$$E_4 = IE_1 = E_1 I, \quad E_5 = IE_2, \quad E_6 = IE_3. \tag{1.8}$$

Note that the bivectors E_1, E_2, E_3 satisfy the following rules of multiplication:

$$E_1^2 = E_2^2 = E_3^2 = 1 \quad (E_4^2 = E_5^2 = E_6^2 = -1), \tag{1.9}$$

$$E_i E_j = -E_j E_i, \quad \text{for } i = j = 1,2,3 \text{ and } i \neq j, \tag{1.10}$$

and generate a $2^3 = 8$ dimensional Clifford algebra called the *Pauli albegra* $\mathscr{P}$, which is the even subalgebra of $\mathscr{D}$ consisting of the scalars, bivectors, and pseudoscalars.

Operations similar to (1.3) and (1.4) can be defined in the Pauli algebra $\mathscr{P}$. Thus, let A, B be bivectors in $\mathscr{D}_2$, then

$$A = \alpha^i E_i \equiv \sum_{i=1}^{3} \alpha^i E_i, \quad B = \beta^i E_i,$$

where α^i and β^i are "complex" scalars of the form

$$\alpha^i = \alpha'^i + \alpha''^i I \quad \text{and} \quad \beta^i = \beta'^i + \beta''^i I$$

and I is the unit pseudoscalar defined in (1.7). Now define:

$$\begin{aligned} A \circ B \equiv{}& \tfrac{1}{2}(AB + BA) = \alpha^1\beta^1 + \alpha^2\beta^2 + \alpha^3\beta^3 \\ \equiv{}& G(A,B) \end{aligned} \tag{1.11}$$

and

$$\begin{aligned} A \times B \equiv{}& \tfrac{1}{2}(AB - BA) = \begin{vmatrix} \alpha^2 & \alpha^3 \\ \beta^2 & \beta^3 \end{vmatrix} IE_1 + \begin{vmatrix} \alpha^3 & \alpha^1 \\ \beta^3 & \beta^1 \end{vmatrix} IE_2 \\ &+ \begin{vmatrix} \alpha^1 & \alpha^2 \\ \beta^1 & \beta^2 \end{vmatrix} IE_3. \end{aligned} \tag{1.12}$$

From (1.11) and (1.12) it follows that

$$AB \equiv \tfrac{1}{2}(AB + BA) + \tfrac{1}{2}(AB - BA) = A \circ B + A \times B, \tag{1.13}$$

i.e., the geometric product of bivectors can be decomposed into the sum of a symmetric product, or complex scalar part, and a Lie product, or bivector part. The metric tensor $G(A,B)$ defined by the symmetric product (1.11) turns the space of bivectors $\mathscr{D}_2$ into a complex Euclidean space, as is noted by Thorpe,[5] and like $g(a, b)$ is Lorentz invariant.

The operations $A \circ B$ and $A \times B$ in the Pauli algebra can be expressed entirely in terms of the operations (1.3) and (1.4) in the Dirac algebra. Thus, let $A = a \wedge b$ and $B = c \wedge d$, then

$$A \circ B = A\cdot B + A \wedge B = \text{scalar} + \text{pseudoscalar} \tag{1.14}$$

where

$$A\cdot B = (a \wedge b)\cdot(c \wedge d) = (a\cdot d)(b\cdot c) - (a\cdot c)(b\cdot d),$$

$$A \wedge B = a \wedge b \wedge c \wedge d = \begin{vmatrix} \alpha^0 & \alpha^1 & \alpha^2 & \alpha^3 \\ \beta^0 & \beta^1 & \beta^2 & \beta^3 \\ \gamma^0 & \gamma^1 & \gamma^2 & \gamma^3 \\ \delta^0 & \delta^1 & \delta^2 & \delta^3 \end{vmatrix} I,$$

and

$$A \times B = (a \wedge b) \times (c \wedge d)$$
$$= a \wedge (b \cdot (c \wedge d)) + (a \cdot (c \wedge d)) \wedge b \qquad (1.15)$$
$$= (b \cdot c) a \wedge d - (b \cdot d) a \wedge c$$
$$+ (a \cdot c) d \wedge b - (a \cdot d) c \wedge b.$$

Note also the duality relations

$$I(A \cdot B) = (IA) \wedge B \quad \text{and} \quad I(A \wedge B) = (IA) \cdot B \qquad (1.16)$$

between $A \cdot B$ and $A \wedge B$.

There are two triple products in the Pauli algebra $\mathscr{P}$ built up from the symmetric and Lie products. They are given by

$$(A \times B) \circ C = \begin{vmatrix} \alpha^1 & \alpha^2 & \alpha^3 \\ \beta^1 & \beta^2 & \beta^3 \\ \gamma^1 & \gamma^2 & \gamma^3 \end{vmatrix} I = A \circ (B \times C) \qquad (1.17)$$

and

$$A \times (B \times C) = (A \circ B) C - (A \circ C) B. \qquad (1.18)$$

From (1.17), it follows that three bivectors A, B, C are linearly independent over the complex scalars iff their triple product (1.17) is nonvanishing. The identities (1.11), (1.12), (1.17), and (1.18) of the Pauli algebra obviously parallel their Gibbs–Heaviside vector algebra counterparts, and this suggests that the former are in some sense the "complexified" version of the latter.

We conclude this section with a discussion and classification of bivectors.[11] A bivector B is said to be *simple* if

$$B^2 = B \cdot B + B \wedge B = B \cdot B, \qquad (1.19)$$

i.e., B^2 is a (real) scalar. The bivector B is said to be *null* if

$$B^2 = 0 \text{ and } B \neq 0. \qquad (1.20)$$

A simple bivector can always be factored into the product of two anticommuting (orthogonal) Dirac vectors, i.e., $B = ab = -ba$. A non-null bivector $C \neq 0$ can always be uniquely expressed in the form

$$C = \rho e^{I\theta} A, \text{ for } \rho \geqslant 0,\ 0 \leqslant \theta < 2\pi, \text{ and } A^2 = 1, \qquad (1.21)$$

and a null bivector N can always be uniquely expressed in the form

$$N = \rho(1 + A_1) A_2, \quad \rho > 0, \text{ and}$$
$$A_1^2 = A_2^2 = 1, \text{ and } A_1 A_2 = -A_2 A_1. \qquad (1.22)$$

To prove (1.21), note that we can define $\rho^2 e^{2I\theta} \equiv C^2 \neq 0$, and $A = \rho^{-1} e^{-I\theta} C$, from which the required properties easily follow. For the case of the null bivector N, there exists an orthonormal basis a_u related to the orthonormal basis e_u of (1.1) by a proper Lorentz transformation, which satisfies:

$$N = \rho a_2 n = \rho a_2 (a_0 + a_1) = \rho a_2 a_0 a_0 (a_0 + a_1) \qquad (1.23)$$
$$= \rho A_2 (1 - A_1) = \rho (1 + A_1) A_2,$$

where $n = a_0 + a_1$ is a null vector, $A_1 = a_1 \wedge a_0 = a_1 a_0$, and $A_2 = a_2 \wedge a_0 = a_2 a_0$.

The following is a multiplication table for $A_1, A_2, A_{12} \equiv A_1 A_2$, and a null bivector $N = (1 + A_1) A_2$:

	A_1	A_2	A_{12}	N
A_1	1	A_{12}	A_2	N
A_2	$-A_{12}$	1	$-A_1$	$1 - A_1$
A_{12}	$-A_2$	A_1	-1	$-1 + A_1$
N	$-N$	$1 + A_1$	$-1 - A_1$	0

(1.24)

2. VECTOR AND BIVECTOR DIFFERENTIATION

Two notions of differentiation are fundamental to the methods of this work, the vector derivative ∂_v, defined for differentiable $\mathscr{D}$-valued functions of a vector variable $f: \mathscr{D}_1 \to \mathscr{D}$, and the bivector derivative ∂_B, defined for differentiable functions of a bivector variable $F: \mathscr{D}_2 \to \mathscr{D}$. The vector derivative is characterized by two properties:

$$\partial_B \text{ behaves algebraically like a vector in } \mathscr{D}_1. \qquad (2.3)$$
$$a \cdot \partial f \equiv a \cdot \partial_v f(v) \equiv (d/dt) f(v + ta)|_{t=0}. \qquad (2.2)$$

The bivector derivative is characterized by two similar properties:

$$\partial_B \text{ behaves algebraically like a vector in } \mathscr{D}_2. \qquad (2.3)$$
$$A \cdot \partial F \equiv A \cdot \partial_B F(B) \equiv (d/dt) F(B + tA)|_{t=0}. \qquad (2.4)$$

We shall not be precise in specifying conditions for vector and bivector differentiability, because we shall be concerned here only with derivatives of linear functions, which always exist.

Because of property (2.1), ∂_v can be expressed in terms of the orthonormal basis $\{e_u\}$ by

$$\partial \equiv \partial_v = e_0 e_0 \cdot \partial_v - e_1 e_1 \cdot \partial_v - e_2 e_2 \cdot \partial_v - e_3 e_3 \cdot \partial_v. \qquad (2.5)$$

Simple but important formulas for the vector derivative are

$$a \cdot \partial v = a = \partial v \cdot a, \qquad (2.6)$$
$$\partial v = 4 \Leftrightarrow \partial \cdot v = 4 \quad \text{and} \quad \partial \wedge v = 0, \qquad (2.7)$$
$$a \wedge \partial v = 3a = \partial v \wedge a, \qquad (2.8)$$
$$\partial_v \wedge \partial_u u \wedge v = 12 = \partial_v \partial_u u \wedge v, \qquad (2.9)$$

which can be easily derived from (2.1), (2.2), (2.5) and algebraic identities from Sec. 1. For example, to prove (2.6), note that

$$a \cdot \partial_v v \equiv (d/dt)(v + ta)|_{t=0} = a.$$

Identity (2.7) follows by using (2.5) and (2.6) to get

$$\partial_v v = e_0 e_0 \cdot \partial_v v - e_1 e_1 \cdot \partial_v v - e_2 e_2 \cdot \partial_v v - e_3 e_3 \cdot \partial_v v$$
$$= e_0^2 - e_1^2 - e_2^2 - e_3^2 = 4.$$

Identity (2.8) then follows by using (1.5), (2.1), (2.7), and (2.6) to get

$$a \wedge \partial v = a \partial v - a \cdot \partial v = 4a - a = 3a.$$

Because of its property (2.3), the bivector derivative ∂_B can be expressed in terms of the orthonormal timelike bivector basis $\{E_i\}$ by

$$\partial \equiv \partial_B = E_1 E_1 \circ \partial_B + E_2 E_2 \circ \partial_B + E_3 E_3 \circ \partial_B. \qquad (2.10)$$

Simple, but important, formulas for the bivector derivative

are

$$A\cdot\partial B = A = \partial B\cdot A, \tag{2.11}$$

$$A\wedge\partial B = -I(IA)\cdot\partial B = A = \partial B\wedge A, \tag{2.12}$$

$$A\circ\partial B = A\cdot\partial B + A\wedge\partial B = 2A = \partial B\circ A \tag{2.13}$$

$$\partial B = 6 \Leftrightarrow \partial\circ B = 6 \text{ and } \partial\times B = 0, \tag{2.14}$$

$$A\times\partial B = 4A = \partial B\times A, \tag{2.15}$$

$$\partial_B\times\partial_A A\times B = 24 = \partial_B\partial_A A\times B, \tag{2.16}$$

$$\partial_C\circ\partial_B\times\partial_A A\times B\circ C = 48 = \partial_C\partial_B\partial_A A\times B\circ C, \tag{2.17}$$

and these formulas can be derived from (2.3), (2.4), (2.10) and the algebraic identities in Sec. 1. For example, to prove the left-hand side of (2.11), use definition (2.4) to get

$$A\cdot\partial B = (d/dt)(B+tA)|_{t=0} = A|_{t=0} = A.$$

The left-hand side of (2.12) is a consequence of (2.11) and (1.16). To prove the right-hand side of (2.11), we use (2.10), (1.14), and what we have just proved, to get

$$\begin{aligned}\partial B\cdot A &= E_1E_1\circ\partial B\cdot A + E_2E_2\circ\partial B\cdot A + E_3E_3\circ\partial B\cdot A\\ &= E_1(E_1\cdot\partial - I(IE_1)\cdot\partial)B\cdot A + \cdots\\ &= E_1(E_1\cdot A - I(IE_1)\cdot A) + \cdots\\ &= E_1E_1\circ A + E_2E_2\circ A + E_3E_3\circ A = A.\end{aligned}$$

The right-hand side of (2.12) now easily follows from the right-hand sides of (2.11) and (1.16). Finally, to see that (2.16) is a consequence of (2.14) and (2.13), first use (2.3) and (1.14) and write

$$A\times\partial B = A\partial B - A\circ\partial B = 6A - 2A = 4A.$$

There is a close relationship between the vector and bivector derivatives of a linear function $F(B)$. It is given by

$$\partial_B F(B) = \partial_v\wedge\partial_u F(B) = \tfrac{1}{2}\partial_v\wedge\partial_u F(u\wedge v), \tag{2.18}$$

where $B = \frac{1}{2}u\wedge v$. This relationship is checked for the identity $F(B) = B$ by comparing (2.9) and (2.14). The vector and bivector derivatives, and their natural generalization to 2^n-dimensional Clifford algebra were originally developed as coordinate-free tools for use in linear algebra and differential geometry in Ref. 12, and since have been extensively used in Ref. 9.

3. BIVECTOR OPERATORS

By a bivector operator $F(B)$ we mean a linear bivector-valued function of the bivector variable B. If in addition F satisfies

$$F(IB) = IF(B), \tag{3.1}$$

we say that F is *dual*. If instead F satisfies

$$F(IB) = -IF(B), \tag{3.2}$$

we say that F is *antidual*. A bivector operator can always be split into the sum of dual and antidual parts, as is evident in

$$F(B) = S(B) + T(B), \tag{3.3}$$

where

$$S(B) = \tfrac{1}{2}[F(B) - IF(IB)],$$

and

$$T(B) = \tfrac{1}{2}[F(B) + IF(IB)].$$

Using formulas (2.11), (2.12), and (2.13), we calculate derivatives of $F(B)$, finding

$$A\circ\partial F = A\circ\partial S + A\circ\partial T = 2S(A), \tag{3.4}$$

since

$$A\circ\partial S = S(A\circ\partial B) = 2S(A) \tag{3.5}$$

and

$$\begin{aligned}A\circ\partial T &= T(A\cdot\partial B) - T(A\wedge\partial B)\\ &= T(A) - T(A) = 0.\end{aligned} \tag{3.6}$$

From (3.4) and (2.13) it follows that

$$\partial F = \tfrac{1}{2}\partial_A A\circ\partial F = \partial S = \partial\circ S + \partial\times S, \tag{3.7}$$

which shows that the bivector derivative of F is completely determined by the bivector derivative of its dual part. As a consequence of this, it follows that

$$\partial T = 0, \tag{3.8}$$

i.e., the derivative of an antidual operator vanishes.

An operator $F(B)$ is said to be *symmetric* (with respect to the metric g) if

$$F(A)\cdot B = A\cdot F(B) \Leftrightarrow F(A) = F^\dagger(A) \equiv \partial_B F(B)\cdot A \tag{3.9}$$

and *skew-symmetric* (w.r.t.g) if

$$F(A)\cdot B = -A\cdot F(B) \Leftrightarrow F(A) = -F^\dagger(A). \tag{3.10}$$

Differentiating the first expressions in (3.9) and (3.10) by $\partial_A\partial_B$ gives, with the help of (2.11),

$$\partial\times F = \tfrac{1}{2}(\partial F - F\overleftarrow{\partial}) = 0 \tag{3.11}$$

and

$$\partial\circ F = \tfrac{1}{2}(\partial F + F\overleftarrow{\partial}) = 0, \tag{3.12}$$

respectively, where $\overleftarrow{\partial}$ differentiates to the left. Thus, symmetric operators have vanishing *curl*, whereas skew-symmetric operators have vanishing *trace*. Symmetric bivector operators are known in the literature as curvature operators, and will be studied in Sec. 5.

An operator is said to be *dual symmetric* if it is both dual and symmetric, and *dual skew-symmetric* if it is both dual and skew-symmetric. An operator is dual symmetric iff

$$F(A)\circ B = A\circ F(B), \tag{3.13}$$

i.e., F is symmetric w.r.t. the metric G, or equivalently, iff

$$A\circ\partial F = 2F(A) = \partial F\circ A. \tag{3.14}$$

To establish (3.13), note by using (1.16) that

$$\begin{aligned}F(A)\wedge B &= -I(IF(A))\cdot B = -IF(IA)\cdot B\\ &= -I(IA)\cdot F(B) = A\wedge F(B)\end{aligned}$$

and combine this result with (3.9). Property (3.14) follows directly from (3.13) and (2.13). An operator is dual skew-symmetric iff

$$F(A)\circ B = -A\circ F(B), \tag{3.15}$$

i.e., F is skew-symmetric w.r.t. the metric G, or, equivalently, iff

$$F(B) = \tfrac{1}{4}B\times(\partial\times F). \tag{3.16}$$

The proof of (3.15) is similar to that of (3.13). The proof of (3.16) follows by using (1.18), (3.15), and (2.13) to get

$$B\times(\partial\times F) = B\circ\partial F - \partial F\circ B = 4F(B).$$

There is an important identity satisfied by dual operators $F(B)$. It is given by

$$F(A\times C)+F(A)\times C+A\times F(C)=\tfrac{1}{2}\partial\circ FA\times C+\tfrac{1}{2}(A\times C)\times(\partial\times F). \tag{3.17}$$

In the special case that F is also symmetric, (3.17) reduces to

$$F(A\times C)+F(A)\times C+A\times F(C)=\tfrac{1}{2}\partial\circ FA\times C. \tag{3.18}$$

In the special case that F is dual skew-symmetric, (3.17) reduces to (3.16). Identity (3.18) follows by equating the right sides of the identitites

$$A\times(C\times\partial)F-C\times(A\times\partial)F=\partial\circ FA\times C-2F(A\times C)$$

and

$$\begin{aligned}A\times(C\times\partial)F-C\times(A\times\partial)F&=\partial\circ FA\circ C-2CF(A)-\partial\circ FCA+2AF(C)\\&=-2C\times F(A)-2F(C)\times A,\end{aligned}$$

and the general identity (3.17) follows by combining (3.18) and (3.16).

We will now find the determinant, the characteristic equation, and the Cayley–Hamilton theorem for a dual operator $F(B)$. Define

$$\det(F)=(1/48)\partial_C\circ\partial_B\times\partial_A\,F(A)\times F(B)\circ F(C). \tag{3.19}$$

In terms of the orthonormal basis $\{E_i\}$, with the help of (1.17) and (2.10), it is not difficult to check that

$$\det(F)=-IF(E_1)\times F(E_2)\circ F(E_3)=|F(E_i)\circ E_j|. \tag{3.20}$$

For $F(B)=B$, from (2.17) or (1.7) it can be seen that $\det(F)=1$, as would be expected. Carrying out the indicated differentiation in (3.19) gives

$$\det(F)=1/48[8\partial\circ F^3-6\partial\circ F\;\partial\circ F^2+(\partial\circ F)^3], \tag{3.21}$$

which expresses the $\det(F)$ in terms of the complex scalars $\partial\circ F$, $\partial\circ F^2$, and $\partial\circ F^3$. Note the these three complex scalars correspond to six real scalars, and are Lorentz invariant; more about them later. In the case that F is dual skew-symmetric, $\det(F)=0$, since $\partial\circ F=0=\partial\circ F^3$.

To obtain the characteristic polynomial for F, define

$$F'\equiv F-\lambda\equiv F(B)-\lambda B. \tag{3.22}$$

Then $\psi(\lambda)$ is given by

$$\psi(\lambda)=\det(F')=\det(F-\lambda). \tag{3.23}$$

Using (3.23), and (3.19) or (3.21), we compute

$$\psi(\lambda)=\lambda^3-\tfrac{1}{2}\partial\circ F\lambda^2-\tfrac{1}{4}[\partial\circ F^2-\tfrac{1}{2}(\partial\circ F)^2]\lambda-1/48[8\partial\circ F^3-6\partial\circ F\partial\circ F^2+(\partial\circ F)^3]. \tag{3.24}$$

In the case that F is dual skew-symmetric, $\psi(\lambda)$ simplifies to

$$\psi(\lambda)=\lambda[\lambda+\tfrac{1}{2}(\partial\circ F^2)^{1/2}][\lambda-\tfrac{1}{2}(\partial\circ F^2)^{1/2}].$$

The Cayley–Hamilton for F says simply that

$$\psi(F)\equiv 0, \tag{3.26}$$

i.e., F satisfies its characteristic equation. The method of proof of (3.26) is to decompose $\det(F)A$, which is the last term in $\psi(F)$, into the sum of the other terms. This is accomplished in the following steps:

$$\begin{aligned}48A\det(F)&=A\partial_3\circ\partial_2\times\partial_1F_1\times F_2\circ F_3\\&=6\partial_2\times\partial_1F_1\times F_2\circ F(A)\\&=6\partial_2\times\partial_1F_1\times F_2F(A)-6\partial_2\times\partial_1(F_1\times F_2)\times F(A)\\&=[6(\partial\circ F)^2-12\partial\circ F^2]F(A)-24\partial\circ FF^2(A)+48F^3(A).\end{aligned}$$

This formulation and proof of the Cayley–Hamilton theorem was first found for linear transformations in Ref. 12.

We will now show that an antidual operator $T(B)$ can be expressed entirely in terms of two symmetric trace-free vector operators. First consider the identity

$$\begin{aligned}T(B)&=\tfrac{1}{2}(B\cdot\partial_v)\cdot\partial_uT(u\wedge v)\\&=\tfrac{1}{2}B\cdot\partial_v\partial_uT(u\wedge v)-B\times\partial_AT(A)\\&=\tfrac{1}{2}B\cdot\partial_v\partial_uT(u\wedge v).\end{aligned} \tag{3.27}$$

The last equality is a consequence of (3.6) and (3.8), since

$$B\times\partial_AT(A)=B\partial_AT(A)-B\circ\partial_AT(A)=0.$$

Now define the vector operators

$$t(v)\equiv\partial_u\cdot T(u\wedge v)\quad\text{and}\quad\bar{t}(v)\equiv\partial_u\cdot T(u\wedge vI). \tag{3.28}$$

An easy consequence of (3.8) is that $t(v)$ and $\bar{t}(v)$ satisfy

$$\partial_vt(v)=0=\partial_v\bar{t}(v),$$

which means $t(v)$ and $\bar{t}(v)$ are symmetric trace-free operators. We can now express (3.27) in the form

$$\begin{aligned}T(B)&=\tfrac{1}{2}B\cdot\partial_v[\partial_u\cdot T(u\wedge v)+\partial_u\cdot T(u\wedge vI)I]\\&=E(B)+D(B),\end{aligned} \tag{3.29}$$

where

$$E(B)\equiv\tfrac{1}{2}B\cdot\partial_vt(v)=E^\dagger(B) \tag{3.30}$$

is an antidual symmetric bivector operator, and

$$D(B)\equiv\tfrac{1}{2}B\cdot\partial_v\bar{t}(v)I=-D^\dagger(B) \tag{3.31}$$

is an antidual skew-symmetric bivector operator. The symmetry of $E(B)$ follows from the steps

$$\begin{aligned}E^\dagger(B)&\equiv\partial_AE(A)\cdot B=\tfrac{1}{2}\partial_A[A\cdot\partial_vt(v)]\cdot B=\tfrac{1}{2}\partial_AA\cdot[\partial_vt(v)\cdot B]\\&=\tfrac{1}{2}[SB{:}\partial{:}v]t(v)\cdot B=E(B),\end{aligned}$$

and the skew-symmetry of $D(B)$ can be similarly established.

We have the following important properties of $E(B)$ and $D(B)$:

$$\partial_u\wedge E^k(u\wedge v)=0,\quad\text{for}\quad k=1,2,\cdots \tag{3.32}$$

and

$$\partial_u\wedge D^{2k}(u\wedge v)=0=\partial_u\cdot D^{2k-1}(u\wedge v),\quad\text{for}\quad k=1,2,\cdots, \tag{3.33}$$

which can be proved by using induction on k and the symmetry of $t(v)$ and $\bar{t}(v)$.

Combining the results of (3.3), (3.13), (3.16), and (3.29), we find that a general bivector operator can always be decomposed into

$$F(B)=[H(B)+E(B)]+[J(B)+D(B)], \tag{3.34}$$

where $H(B)$ is dual symmetric, $E(B)$ is antidual symmetric, $J(B)$ is dual skew-symmetric, and $D(B)$ is antidual skew-symmetric. The classification of trace-free symmetric vector

operator is carried out in Ref. 13 by reducing the problem to the Petrov classification of a correlated Weyl tensor.

4. CLASSIFICATION OF DUAL OPERATORS

Let F be a dual operator, i.e., one satisfying (3.1). The F has the characteristic polynomial $\psi(\lambda)$ given by (3.24), and setting

$$\psi(\lambda) = 0 \tag{4.1}$$

gives the characteristic equation for F. The solutions $\lambda_1, \lambda_2, \lambda_3$ are the eigenvalues of F. Writing

$$\psi(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3), \tag{4.2}$$

we find, on expanding the right-hand side of (4.2) and equating the coefficients of λ with those in (3.24), that

$$\tfrac{1}{2}\partial \circ F^k = \lambda_1^k + \lambda_2^k + \lambda_3^k \quad \text{for} \quad k = 1,2,3. \tag{4.3}$$

The characteristic roots of (4.1) have multiplicity 1, 2, or 3 according to whether

$$\lambda_1 \neq \lambda_2 \neq \lambda_3 \quad \text{for multiplicity 1,} \tag{4.4}$$

$$\lambda_1 \neq \lambda_2 = \lambda_3 \quad \text{for multiplicity 2,} \tag{4.5}$$

$$\lambda_1 = \lambda_2 = \lambda_3 \quad \text{for multiplicity 3,} \tag{4.6}$$

Conditions for (4.4), (4.5), (4.6) can be given in terms of $\partial \circ F$, $\partial \circ F^2$, $\partial \circ F^3$,[13,15]

We see from (3.23) and (3.20) that, for each eigenvalue λ_k,

$$[F(E_1) - \lambda_k E_1] \times [F(E_2) - \lambda_k E_2] \circ [F(E_3) - \lambda_k E_3] = 0,$$

which implies, because of (1.17), that there exist eigenbivectors satisfying

$$F(C_k) = \lambda_k C_k \quad \text{for} \quad k = 1,2,3. \tag{4.7}$$

We will consider the classification of dual symmetric and dual skew-symmetric operators separately. This is justified by the fact that we can always decompose F into

$$F(B) = H(B) + J(B), \tag{4.8}$$

where

$$H(B) \equiv \tfrac{1}{2}[F(B) + F^\dagger(B)] = \tfrac{1}{2}\partial_B F(B) \cdot B$$

is dual symmetric, and

$$J(B) = \tfrac{1}{2}[F(B) - F^\dagger(B)] = \tfrac{1}{4}B \times (\partial \times F)$$

is dual skew-symmetric.

Let $J(B)$ be a dual skew-symmetric operator. Then by (3.16), $J(B)$ can be writtern in the form

$$J(B) = B \times Q, \tag{4.9}$$

where $Q = \frac{1}{4}\partial \times J$. From (4.9) we calculate

$$J^2(B) = (B \times Q) \times Q = B\,Q \circ Q - B \circ Q\,Q \tag{4.10}$$

and

$$J^3(B) = B \times Q\,Q^2, \tag{4.11}$$

from which is follows that $\partial J^k = 4Q^k$, which implies

$$\partial \circ J = 0, \quad \partial \circ J^2 = 4Q^2, \quad \partial \circ J^3 = 0. \tag{4.12}$$

The characteristic polynomial (3.25) of $J(B)$ can be written in terms of Q^2, getting

$$\psi(\lambda) = \lambda[\lambda + (Q^2)^{1/2}][\lambda - (Q^2)^{1/2}]. \tag{4.13}$$

From (4.13) it is clear that the key to the classification of $F(B)$ is the bivector Q. The canonical forms (1.21) and (1.22) for a bivector tell us that

$$Q = 0 \quad \text{or} \quad Q = \rho e^{I\theta} A_1 \quad \text{or} \quad Q = \rho(1 + A_1)A_2.$$

The case $Q = 0$ is trivial.

For the case $Q = \rho e^{I\theta} A_1$, we construct the null bivectors $N = (1 + A_1)A_2$, and $M = (1 - A_1)A_2$, and note, with the help of table (1.23), that

$$A_1 = \tfrac{1}{2}N \times M, \quad N \circ M = 2, \quad A_1 \circ N = 0 = A_1 \circ M. \tag{4.14}$$

It then follows, using (1.18), that

$$\begin{aligned} J(B) &= B \times Q = \tfrac{1}{2}\rho e^{I\theta} B \times (N \times M) \\ &= \tfrac{1}{2}\rho e^{I\theta}(B \circ N\,M - B \circ M\,N). \end{aligned} \tag{4.15}$$

From the canonical form (4.15) of $J(B)$, with the help of (4.14), we can read off the eigenbivectors and eigenvalues of J. Thus,

$$J(A_1) = 0A_1, \quad J(N) = -\rho e^{I\theta} N, \quad J(M) = \rho e^{I\theta} M. \tag{4.16}$$

For the case $Q = \rho N$, where $N = (1 + A_1)A_2$, $Q^2 = 0$,

$$J(B) = \rho B \times N = \rho(B \circ A_1 N - B \circ N A_1) \tag{4.17}$$

is the desired canonical form. We calculate

$$\begin{aligned} &J(N) = 0, \quad J(A_{12}) = \rho A_1, \quad J(A_1) = \rho N, \\ &J(A_2) = -\rho A_1, \end{aligned} \tag{4.18}$$

from which it follows that N is the only eigenbivector of $J(B)$. The above cases can be summarized in the following table enumerating the number of null eigenbivectors of $J(B)$:

$$\begin{array}{cc} 11 & Q^2 \neq 0 \\ 1(Q \neq 0) \quad -(Q = 0) & Q^2 = 0. \end{array} \tag{4.19}$$

Of course it closely parallels that given by Penrose,[3] in his spinor classification of an electromagnetic field. The bivector Q represents an electromagnetic field at a point in spacetime.

We will now carry out the classification of a dual symmetric operator $H(B)$ into the so-called Petrov types. Because of (4.7), H has eigenbivectors and values satisfying

$$H(C_k) = \lambda_k C_k, \quad \text{for} \quad k = 1,2,3. \tag{4.20}$$

That orthogonal bivectors correspond to distinct eigenvalues follows from the standard argument:

$$(\lambda_i - \lambda_j) C_i \circ C_j = H(C_i) \circ C_j - C_i \circ H(C_j) = 0. \tag{4.21}$$

Furthermore, because of the bivector classifications (1.21) and (1.22), and the fact that H is dual, each eigenbivector C of H can be replaced by a time-like eigenbivector A, with $A^2 = 1$, or by a null bivector $N = (1 + A_1)A_2$, having the same eigenvalue as C. We will always assume that the eigenbivectors C_k of H have been so normalized. The operator $H(B)$ is said to be of Petrov

Type I: if $\{C_k\}$ spans a three-dim. space,
Type II: if $\{C_k\}$ spans a two-dim. space,
Type III: if $\{C_k\}$ spans one-dim. space.

Suppose H is Type I. If the eigenvalues λ_k are distinct, then by (4.21) the C_k's are orthogonal. This excludes the possibility that one or more of the eigenbivectors are null, because inspection of table (1.23) shows that if an eigenbivector C is orthogonal to a null eigenbivector $N = (1 + A_1)A_2$, then C

must be of the form $C = A_i + \alpha N$, so that the eigenbivectors C_k could span at most a two-dim. space. If the eigenvalues are not distinct, simple orthonormal space-like bivectors can still be choosen with the same multiplicity as the repeated roots.

Suppose H is type II. Then the eigenvalues of H cannot all be distinct, for othewise, because of (4.21), H would be Type I. Also, H cannot have two orthonormal time-like eigenbivectors A_1, A_2, for in this case, letting $A_3 = IA_1 \times A_2$, we find by using (3.18) that

$$H(A_1 \times A_2) + H(A_1) \times A_2 + A_1 \times H(A_2) = \tfrac{1}{2}\partial \circ H A_1 \times A_2,$$

or

$$H(A_3) = (\tfrac{1}{2}\partial \circ H - \lambda_1 - \lambda_2)A_3,$$

so that A_3 would be an eigenbivector also, contradicting the assumption that H is Type II. Thus, H must have a null eigenbivector $N = (1 + A_1)A_2$ satisfying

$$H(N) = \lambda_N N, \tag{4.22}$$

and a time-like bivector of the form $C_1 = A_1 + \alpha N$, satisfying

$$H(C_1) = \lambda_1 C_1. \tag{4.23}$$

Equations (4.22), (4.23) imply

$$H(A_1) = \lambda_1 A_1 + \beta_1 N, \tag{4.24}$$

where

$$\beta_1 \equiv A_1 \circ H(A_2) = A_2 \circ H(A_1) = \alpha(\lambda_1 - \lambda_N).$$

In the degenerate case when $\lambda_1 = \lambda_N$, (4.24) reduces to

$$H(A_1) = \lambda_1 A_1, \quad \text{and} \quad \beta_1 = 0. \tag{4.25}$$

Finally, note that $A_1 \times N = N$, and using this in identity (3.18), together with (4.24) and (4.22), shows that

$$\tfrac{1}{2}\partial \circ H = \lambda_1 + 2\lambda_N, \tag{4.26}$$

for Type II.

Suppose H is Type III. Then H has one eigenbivector, a null bivector N, satisfying

$$H(N) = \lambda_N N, \tag{4.27}$$

and

$$\tfrac{1}{2}\partial \circ H = 3\lambda_N. \tag{4.28}$$

The above classification scheme can be refined by introducing the notion of principal null directions of H. These are null bivectors M which satisfy

$$H(M) \circ M = 0 \quad \text{and} \quad M^2 = 0, \tag{4.29}$$

and were used by Penrose[3] in his refinement of the Petrov classification of the conformal curvature tensor using spinors. The condition (4.29) was first noted in a remark by Thorpe.[5] The principal null bevectors of H are explicitly calculated below, and their coincidence patterns are specified by new and simple conditions.

For the case that the λ_k's are distinct, $H(B)$ has a basis of orthonormal time-like eigenbivectors A_1, A_2, A_3. In terms of this basis we can write

$$H(B) = B \circ A_1 \lambda_1 A_1 + B \circ A_2 \lambda_2 A_2 + B \circ A_3 \lambda_3 A_3. \tag{4.30}$$

Imposing the condition (4.29) leads to the equations

$$M \circ M = \alpha_1^2 + \alpha_2^2 + \alpha_3^2 = 0, \quad \text{or} \quad \alpha_3^2 = -(\alpha_1^2 + \alpha_2^2),$$

and

$$H(M) \circ M = \lambda_1 \alpha_1^2 + \lambda_2 \alpha_2^2 - \lambda_3(\alpha_1^2 + \alpha_2^2) = 0,$$

where $M = \alpha_1 A_1 + \alpha_2 A_2 + \alpha_3 A_3$, which has solutions

$$\alpha_1 = \pm\sqrt{\lambda_2 - \lambda_3}, \quad \alpha_2 = \pm\sqrt{\lambda_3 - \lambda_1},$$
$$\alpha_3 = \sqrt{\lambda_1 - \lambda_2} \tag{4.31}$$

which correspond to four distinct principal null directions.

For all other cases there will be a null eigenbivector $N = (1 + A_1)A_2$ for which $H(N) = \lambda_N N$. In these cases, we expand $H(B)$ in the basis $A_1, A_2, A_{12} = A_1 A_2$, finding, with the help of (1.23) and (3.18),

$$H(B) = (B \circ A_1 \lambda_1 + B N \beta_1)A_1 + B \circ H_2 A_2 + (B \circ H_2 - \lambda_N B \circ N)A_{12}, \tag{4.32}$$

where $\beta_1 \equiv A_1 \circ H(A_2)$, $H_2 \equiv H(A_2)$, and $\lambda_1 \equiv \tfrac{1}{2}\partial \circ H - 2\lambda_N$, $\lambda_N \equiv N \circ H_2$.

Note that $H(B)$ is defined entirely in terms of the independent quantities

$$H_2, \quad N, \quad \tfrac{1}{2}\partial \circ H, \tag{4.33}$$

where H_2 is an arbitrary bivector (six parameters), N is an arbitrary null bivector (four parameters), and $\tfrac{1}{2}\partial \circ H$ is an arbitrary complex scalar (two parameters), making up 12 independent parameters in all.

We are now ready to solve for principal null bivectors by imposing (4.29) on the expansion (4.32). This is done in the steps below:

$$M = \alpha_1 A_1 + \alpha_2 A_2 + \alpha_3 A_{12}, \quad M \circ M = \alpha_1^2 + \alpha_2^2 - \alpha_3^2 = 0,$$
$$H(M) = [\lambda_1 \alpha_1 + (\alpha_2 - \alpha_3)\beta_1]A_1 + M \circ H_2 A_2 + [M \circ H_2 - \lambda_N(\alpha_2 - \alpha_3)]A_{12}$$
$$= [\lambda_1 \alpha_1 - y\beta_1]A_1 + [\beta_1 \alpha_1 + \lambda_N \alpha_2 - \beta_2 y]A_2 + [\beta_1 \alpha_1 + (\lambda_N - \beta_2)y + \lambda_N \alpha_2]A_{12}$$

and

$$H(M) \circ M = \beta_2 y^2 + (\delta x - 2\alpha_1 \beta_1)y \tag{4.34}$$

in terms of the new variables:

$$x = \alpha_3 + \alpha_2, \quad y = \alpha_3 - \alpha_2, \quad \delta = \lambda_1 - \lambda_N,$$
$$\beta_1 = A_1 \circ H_2, \quad \beta_2 = A_{21} \circ H_2.$$

Thus the equation

$$H(M) \circ M = 0 \Leftrightarrow \delta x y + \beta_2 y^2 = 2\beta_1 \alpha_1 y, \tag{4.35}$$

and for $y \neq 0 \neq \beta_1$, we find

$$\alpha_1 = [1/2\beta_1](\delta x + \beta_2 y].$$

Squaring the equation in (4.35) leads to the equation

$$y^2[(\delta x + \beta_2 y)^2 - 4\beta_1^2 x y] = 0. \tag{4.36}$$

Analysis of equation (4.36) together with (4.35) leads to the following classification scheme of the principal null directions of $H(B)$:

		1111($\neq\lambda_k$'s)		
	211($\beta_1^2 \neq \delta\beta_2$)	22 ($\beta_1^2 = \delta\beta_2$)		$\delta \neq 0$
31($\beta_1 \neq 0$)	4($\beta_1 = 0 \neq \beta_2$)	—($\beta_1 = 0 = \beta_2$)		$\delta = 0$
III	II	I		(4.37)

As an example of the kind of analysis involved in the above classification, we will carry it to completion in the more involved case $\beta_1 \neq 0 \neq \beta_2$ and $\delta \neq 0$. In this case, set $x = 1$ in (4.36), and factor the resulting quadratic equation in y, getting

$$y^2\left\{\left[y+\left(\frac{\delta}{\beta_2}-\frac{2\beta_1^2}{\beta_2^2}\right)+\epsilon\right]\left[y+\left(\frac{\delta}{\beta_2}-\frac{2\beta_1^2}{\beta_2^2}\right)-\epsilon\right]\right\} = 0, \tag{4.38}$$

where

$$\epsilon = 2\frac{\beta_1}{\beta_2}\sqrt{\frac{\delta\beta_2-\beta_1^2}{\beta_2^2}}.$$

This equation reduces to

$$y^2\left(y-\frac{\beta_1^2}{\beta_2^2}\right)^2 = 0, \quad \text{when} \quad \delta\beta_2 = \beta_1^2. \tag{4.39}$$

We see that for $\epsilon \neq 0$ (or $\delta\beta_2 \neq \beta_1^2$), Eq. (4.38) has double solution for $y = 0$, and two single solutions corresponding to the zeros of the other factors. The corresponding principal null bivectors can be exhibited explicitly by going back to the original variables. Similarly, (4.39) gives two double principal null bivectors for each of the roots of its repeated factors, when $\delta\beta_2 = \beta_1^2$).

In the case that $\frac{1}{2}\partial\circ H = \lambda_1 + \lambda_2 + \lambda_3 = 0$, the three Petrov types can be efficiently characterized by the canonical forms

$$H(B) = (2\lambda_1+\lambda_2)B\circ A_1A_1 + (2\lambda_2+\lambda_1)B\circ A_2A_2 - (\lambda_1+\lambda_1)B, \tag{4.40}$$

for type I,

$$H(B) = \lambda_N B - 3\lambda_N B\circ C_1C_1 + \mu B\circ NN,$$

where

$$\left\{\begin{array}{l} C_1 = A_1 + \alpha N \\ \mu = \beta_2 + 3\lambda_N\alpha^2 \end{array}\right\}, \tag{4.41}$$

for type II, and

$$H(B) = B\circ NC^1 = B\circ C_1N,$$

where

$$C_1 = \beta_1A_1 + \tfrac{1}{2}\beta_2N, \tag{4.42}$$

for type III. The canonical form (4.40) can be derived immediately from (4.30) and the fact that $B \equiv \sum_{k=1}^{3} B\circ A_kA_k$. To derive (4.41), we use the properties (4.22)–(4.26), together with (1.18) and (1.23) and the fact that $C_1\times N = N$, to calculate

$$H(B)\times N = H(B)\times(C_1\times N) = H(B)\circ C_1N - H(B)\circ NC_1 = \lambda_N(B\times N - 3B\circ C_1N),$$

which implies that

$$[H(B) - \lambda_N B + 3\lambda_N B\circ C_1C_1]\times N \equiv 0 \qquad \text{for all } B.$$

Applying A_2x to this last identity, and again utilizing (1.18) and (1.23), yields

$$H(B) - \lambda_N B + 3\lambda_N B\circ C_1C_1 = A_2\circ[H(B) - \lambda_N B + 3\lambda_N B\circ C_1C_1]N = B\circ[H(A_2) - \lambda_N A_2 + 3\lambda_N\alpha C_1]N. \tag{4.43}$$

Applying (4.43) with $B = A_2$ gives

$$H(A_2) - \lambda_N A_2 + 3\lambda_N + C_1 = [A_2\circ H_2 - \lambda_N + 3\lambda_N\alpha^2]N = \mu N, \tag{4.44}$$

where $\mu \equiv \beta_2 + 3\lambda_N\alpha^2$. Then (4.41) now follows trivially from (4.43) and (4.44). Finally, to derive (4.42), we note from (4.32) that for type III, $H(B)$ reduces to the form

$$H(B) = B\circ N\beta_1A_1 + B\circ H(A_2)N, \tag{4.45}$$

from which it follows that

$$H(A_2) = \beta_1A_1 + \beta_2N. \tag{4.46}$$

Together, (4.45) and (4.46) imply (4.42), where $C_1 \equiv \beta_1A_1 + \frac{1}{2}\beta_2N$.

5. RIEMANN CURVATURE: INVARIANTS AND PROPERTIES

Recall that a curvature operator $R(B)$ is a bivector operator satisfying (3.9). From (3.34) it follows tht $R(B)$ can be written in the form

$$R(B) = H(B) + E(B) = R^\dagger(B), \tag{5.1}$$

where

$$H(B) \equiv \tfrac{1}{2}[R - I(RI)](B) = H^\dagger(B)$$

is a dual symmetric bivector operator, and

$$E(B) \equiv \tfrac{1}{2}[R + I(RI)](B) = E^\dagger(B)$$

is an antidual symmetric bivector operator. We shall now study the Lorentz invariants of R in terms of complex scalars of R. By *complex scalars* of R we mean all possible rational linear combinations of complex scalar derivatives of R^k and its dual $(RI)^k$, for $k = 1, 2, \cdots$. Thus,

$$\partial\circ R + I\partial\circ R^2 + 3\partial\circ(RI)^2 - 2\partial\circ(RI)^4 \tag{5.2}$$

is a complex scalar of R. Note that (5.2) is also a Lorentz invariant of R; we will show that all Lorentz invariants of R can be so expressed.

Squaring both sides of (5.1), considered as an operator equation, leads to

$$R^2(B) = [H^2 + E^2](B) + [HE + EH](B), \tag{5.3}$$

where

$$H^2(B) = \tfrac{1}{4}[R^2 - IR^2I](B) - \tfrac{1}{4}[(RI)^2 - I(RI)^2I](B)$$

and

$$E^2(B) = \tfrac{1}{4}[R^2 - IR^2I](B) + \tfrac{1}{4}[(RI)^2 - I(RI)^2](B)$$

are dual symmetric operators, and

$$[HE + EH](B) = \tfrac{1}{2}[R^2 + IR^2I](B)$$

is antidual. Since H^2 and E^2 are symmetric, it follows by (3.11) that

$$\partial\times H^2 = 0 = \partial\times E^2. \tag{5.4}$$

Because of (3.8), derivatives of $R^2(B)$ can be entirely expressed in terms of H^2 and E^2, getting

$$\partial R^2 = \partial H^2 + \partial E^2 = \partial\circ H^2 + \partial\circ E^2. \tag{5.5}$$

Now defining $K \equiv E^2$, note that

$$\partial \circ HK = KH(\partial_B) \circ B = \partial \circ KH, \tag{5.6}$$

$$\partial \circ EHE = \partial_B \circ [E(B) \circ \partial_A EH(A)] = E(\partial_A) \circ EH(A)$$
$$= \partial \cdot E^2 H - \partial \wedge E^2 H = \partial \cdot KH - \partial \wedge KH,$$

and more generally,

$$\partial \circ E^i H E^j$$
$$\text{fs1.05m} = \begin{cases} 0 & \text{if } i+j \text{ is odd} \\ (-1)^{\frac{1}{2}(i+j)} H + \partial \cdot K^{\frac{1}{2}(i+j)} H & \text{if } i+j \text{ is even.} \end{cases} \tag{5.7}$$

From the above remarks it follows that the complex scalars of R can be expressed entirely in terms of complex scalars of the form $\partial \circ H^i K^j$. But the characteristic equations of H, K and HK are all of the third order; with the help of the Cayley–Hamilton theorem (3.26), and (3.32), it follows that all complex scalars of R can be expressed in terms of rational polynomials in

$$\partial \circ H,\ \partial \circ H^2,\ \partial \circ H^3;\quad \partial \cdot K,\ \partial \cdot K^2,\ \partial \cdot K^3;$$
$$\partial \circ HK,\ \partial \circ (HK)^2,\ \partial \circ (HK)^3 \tag{5.8}$$

and their complex conjugates. Thus, R has a total of $3 \times 6 - 3 = 15$ independent invariants, and, as we shall shortly see, the added symmetry of the Riemann curvature tensor reduces this number to the well known 14. (If the same analysis of invariants is carried out for a general bivector operator given by (3.34), in addition to the 15 invariants found in (5.8), there are 15 more given by $\partial \circ J^2$; $\partial \cdot L$, $\partial \cdot L^2$, $\partial \cdot L^3$; $\partial \circ (HL)$, $\partial \circ (HL)^2$, $\partial \circ (HL)^3$; $\partial \circ HJ^2$, $\partial \circ H^2 J$, where $L \equiv D^2$, making up a total of 30 independent scalar invariants.)

For the remainder of this section, let $R(B)$ be a bivector operator with the property

$$\partial_a \wedge R(a \wedge b) = 0. \tag{5.9}$$

An operator with the property (5.9) is called Riemann curvature, because it is equivalent to the usual Riemann curvature tensor R_{ijkl} by way of the identification

$$R_{ijkl} \equiv R(e_i \wedge e_j) \cdot (e_k \wedge e_l), \tag{5.10}$$

the same as is made by Thorpe in Ref. 5. The identities

$$(a \wedge b) \cdot [\partial_d \wedge \partial_c \wedge R(c \wedge d)]$$
$$\equiv [(a \wedge b) \cdot \partial_d] \cdot [\partial_c \wedge R(c \wedge d)]$$
$$- R(a \wedge b) + R^\dagger(a \wedge b)$$

and

$$(a \wedge b \wedge c) \cdot [\partial_v \wedge R(v \wedge d)] \equiv R(c \wedge d) \cdot (a \wedge b) + R(a \wedge d) \cdot (b \wedge c) + R(b \wedge d) \cdot (c \wedge a),$$

together with (5.9), show that

$$R(a \wedge b) \cdot (c \wedge d) = (a \wedge b) \cdot R(c \wedge d) \tag{5.11}$$

and

$$R(a \wedge b) \cdot (c \wedge d) + R(b \wedge c) \cdot (a \wedge d) + R(c \wedge a) \cdot (b \wedge d) = 0. \tag{5.12}$$

Identity (5.11) say that $R(B)$ is a symmetric operator, and (5.12) is the famous Bianchi identity. (The other Bianchi identity in this formalism has the form $\nabla \wedge R = 0$, and can be found in (9); this paper is exclusively concerned with local properties of operators at a point x in curved spacetime.) Thus, (5.9) is equivalent to the two well known conditions (5.11) and (5.12).

Now let $S(B) \equiv R(B) \cdot B$ be the *sectional curvature* defined by $R(B)$. The sectional curvature satisfies the important identity

$$\partial_v \wedge \partial_u S(u \wedge v)|_{u=a, v=b} = 2\partial_v \wedge [v \cdot R(a \wedge v)]|_{v=b}$$
$$= 6R(a \wedge b). \tag{5.13}$$

A well-known consequence of this identity is that $S(u \wedge v) \equiv 0$ iff $R(u \wedge v) \equiv 0$, for all $u, v \in \mathscr{D}_1$.

From the curvature operator $R(B)$ we construct the Ricci operator by contraction:

$$R(b) \equiv \partial_a \cdot R(a \wedge b). \tag{5.14}$$

The Ricci tensor is identified by

$$R_{ij} \equiv R(e_i) \cdot e_j, \tag{5.15}$$

and the property that the Ricci tensor is symmetric is equivalent to

$$\partial_B \times R(B) = \tfrac{1}{2}\partial_a \wedge [\partial_a \cdot R(a \wedge b)] = \tfrac{1}{2}\partial_b \wedge R(b) = 0. \tag{5.16}$$

Scalar curvature is constructed by contracting (5.14), getting

$$R \equiv \partial_b \cdot R(b) = R^i{}_i = 2\partial_B \cdot R(B). \tag{5.17}$$

Notice that we use only the domain to distinguish between Riemann, Ricci, and scalar curvature.

We now decompose $R(B)$, as is done in Refs. 15 and 4, by writing

$$R(B) = C(B) + E(B) + G(B), \tag{5.18}$$

where

$$C(B) = R(B) - \tfrac{1}{2}B \cdot \partial_v [R(v) - (1/6)vR],$$
$$E(B) = \tfrac{1}{2}B \cdot \partial_v [R(v) - (1/4)vR],$$

and

$$G(B) = (1/12)BR.$$

The conformal curvature operator $C(B)$ has the properties

$$\partial_a C(a \wedge b) = 0 = \partial_B C(B) \quad \text{and} \quad C(IB) = IC(B). \tag{5.19}$$

The Einstein operator $E(B)$ has the properties

$$E(B) = \tfrac{1}{2}B \cdot \partial_v E(v) = \tfrac{1}{2}[E(a) \wedge b + a \wedge E(b)] \text{ and}$$
$$E(IB) = -IE(B), \tag{5.20}$$

where $E(v) \equiv \partial_a \cdot E(a \wedge b) = R(v) - \frac{1}{4}vR$ and $\partial_b E(b) = 0$. An important consequence of the fact that $E(B)$ is completely determined by the symmetric vector operator $E(v)$, as given in (5.20), is that

$$\partial_B E^k(B) = \partial_B \cdot E^k(B), \text{ for } k = 1,2,3,\cdots. \tag{5.21}$$

[Recall (3.32)]. The operator $G(B)$ satisfies

$$\partial_a G(a \wedge b) = \tfrac{1}{4}bR,\ \partial_B G(B) = \tfrac{1}{2}R,\ G(IB) = IG(B). \tag{5.22}$$

A comparison of the decompositions (5.1) and (5.18), together with the properties of C, E, and G given above, shows that

$$H(B) = C(B) + G(B) \tag{5.23}$$

and that $E(B)$ has been correctly identified. Because of properties (5.19), (5.22), and (5.23), $\partial_B H(B) = \partial_B \cdot H(B)$ and therefore the 15 real invariants determined by (5.8) reduce to the well known $15 - 1 = 14$ for Riemann curvature.

If spacetime is not empty but is filled with sourceless electromagnetic fields, the Ricci operator (5.14) satisfies

$$R(v) = -QvQ, \tag{5.24}$$

where Q is the electromagnetic bivector defining $J(B)$ in (4.9). It is easy to check that in this case the scalar curvature $R \equiv \partial_v \cdot R(v) = 0$, as is well known. From this it follows that $E(v) \equiv R(v)$, and from (5.20) we calculate

$$E(B) = B \cdot Q\, Q - B \wedge Q\, Q = (B \cdot Q - B \wedge Q)Q. \tag{5.25}$$

Equation (5.25) shows that the Einstein operator determines Q uniquely up to a phase $e^{I\theta}$. Further discussion of these problems in the language of spinors can be found in (2), (3), and (15). There is a discussion of Maxwell's equation and properties of electromagnetic fields in the STA formalism in Ref. 7.

To demonstrate the geometric transparency of the spacetime algebra (STA) formalism, we give a new geometric argument for the well known numbers of independent parameters (IP) of the Riemann, Conformal Weyl, and Einstein tensors. Let $F(B)$ be a general bivector operator. Then $F(B)$ has $6 \times 6 = 36$ IP, since both the domain and range of F are the six-dim. bivector space $\mathscr{D}_2$. Taking the contraction and curl of $F(B)$ defines the operators

$$f(b) \equiv \partial_a \cdot F(a \wedge b) \tag{5.26}$$

and

$$T(b) \equiv \partial_a \wedge F(a \wedge b). \tag{5.27}$$

The operator $T(b)$ determines $4 \times 4 = 16$ IP of $F(B)$, since the domain and range of T are the four-dim. spaces $\mathscr{D}_1$ and $\mathscr{D}_3$. A similar argument shows that $f(b)$ also determines $4 \times 4 = 16$ IP of $F(B)$; but these degrees of freedom are not completely independent of those determined by $T(b)$, since it is easy to show that

$$\partial_b \cdot T(b) = \partial_b \wedge f(b). \tag{5.28}$$

The relation (5.28) shows that $f(b)$ and $T(b)$ have six parameters in common, i.e., they determine a common bivector. The proof that (5.28) is an integrability condition which guarantees the existence of an operator $F(B)$ satisfying (5.26) and (5.27) will be given elsewhere.

From the above considerations we can read off the numbers of IP for the various operators and their corresponding tensors. Thus, for Riemann curvature, (5.27) vanishes leaving $36 - 16 = 20$ IP. For conformal curvature, both (5.27) and (5.26) vanish, and taking into consideration (5.28), this leaves $36 - 16 - 16 + 6 = 10$ IP. For Einstein curvature, since it is completely determined by (5.26), and (5.28) vanishes, taking into account that $\partial_b \cdot E(b) = 0$, gives $16 - 6 - 1 = 9$ IP.

To bring out the advantages of the STA formalism over the tensor and spinor formalisms, we present the following table of how basic quantities and relationships find expression in each.

TENSOR	STA	SPINOR
$F_{uv} = -F_{vu} = (e_u \wedge e_v)\cdot Q$	Q	$\phi^{\alpha\beta}$
$F^*_{uv} = \frac{1}{2}\epsilon_{uvrs}F^{rs}$	IQ	$-i\phi^{\alpha\beta}$
$R_{uvrs} = (e_u \wedge e_v)\cdot R(e_r \wedge e_s)$	$R(B) = C(B) + E(B) + G(B)$	$\psi_{\alpha\beta\gamma\delta}, \phi_{\alpha\beta\dot{\mu}\dot{\nu}}, \Lambda$
$\left\{\begin{matrix} R_{uvrs} + R_{ruvs} + R_{vrus} = 0 \\ R_{uvrs} = R_{rsuv} \end{matrix}\right\}$	$\partial_a \wedge R(a \wedge b) = 0$	$\left\{\begin{matrix} \psi^{\dot{\alpha}}{}_{\beta}{}^{\beta}{}_{\dot{\alpha}} = \psi^{\alpha}{}_{\beta}{}^{\beta}{}_{\alpha} \\ \psi^{\alpha\beta\mu\nu} = \psi^{\mu\nu\alpha\beta}, \phi^{\dot{\alpha}\dot{\beta}\mu\nu} = \phi^{\mu\nu\dot{\alpha}\dot{\beta}} \end{matrix}\right\}$
$\left\{\begin{matrix} E_{uvrs}E^{rs}{}_{pq}C^{pqtw}E_{tw}{}^{mn}E_{mn}{}^{uv} \\ + E^*_{uvrs}E^{rs}{}_{pq}C^{pqtw}E_{tw}{}^{mn}E_{mn}{}^{uv} \end{matrix}\right\}$	$\partial_B \circ E^4 C(B)$	$\phi^{rs\dot{\alpha}\dot{\beta}}\phi_{\rho\sigma\dot{\alpha}\dot{\beta}}\psi^{\rho\sigma\kappa\lambda}\phi_{\kappa\lambda\dot{\mu}\dot{\nu}}\phi_{\gamma\delta}{}^{\dot{\mu}\dot{\nu}}$

(5.29)

ACKNOWLEDGMENTS

I am very grateful to Z. Oziewicz for critically reading the early version of this work and making many valuable suggestions, to R. Ablamowicz for also reading the early version and for calling my attention to important references, to B. Jancewicz for inviting me to the Univeristy of Wroclaw, to J. Ray for a helpful correspondence regarding aspects of this work, and to the Institute for Theoretical Physics for much needed hospitality and support.

[1] A.Z. Petrov, Dokl. Akad. Nauk. SSSR **105**, 905 (1955).
[2] L. Witten, Phys. Rev. **113**, 357 (1959).
[3] R. Penrose, Ann. Phys. (N. Y.) **10**, 171 (1960).
[4] P.J. Greenberg, Stud. Appl. Math. **3**, 277 (1972).
[5] J.A. Thorpe, J. Math. Phys. **10**, 1 (1969).
[6] A. Stehney, J. Math. Phys. **17**, 1793 (1976).
[7] D. Hestenes, *Space–Time Algebra* (Gordon and Breach, New York, 1966).
[8] D. Hestenes, Am. J. Phys. **47**, 399 (1979). (This is the most recent of many articles by Hestenes exploring the physical content of the *Spacetime algebra* approach to physics.)
[9] D. Hestenes and G. Sobczyk, *Clifford Algebras and Geometric Calculus: A Unified Language for Mathematics and Physics* Mathematics Lecture Note Series (Addison-Wesley, Reading Mass., 1980).
[10] M. Riesz, *Clifford Numbers and Spinors*, Lecture series No. 38, The Institute for Fluid Dynamics and Applied Math, University of Maryland (1958).
[11] E. Cartan, *The Theory of Spinors* (M.I.T. P., Cambridge, Mass., 1966).
[12] G. Sobczyk, thesis, Ariz. State Univ. (1971).
[13] G. Sobczyk, "Plebański Classification of the Tensor of Matter" Acta Phys. Pol. B, (August 1980).
[14] C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[15] J. Plebański, Acta. Phys. Pol. **26**, 963 (Nov. 1964).

Linearization stability of Einstein equations coupled with self-gravitating scalar fields

R. V. Saraykar and N. E. Joshi
Department of Mathematics, Nagpur University Campus, Nagpur 440010, India

(Received 7 August 1979; accepted for publication 7 December 1979)

In this paper, we extend the work of Fischer–Marsden and Moncrief on the linearization stability of vacuum spacetimes to the case of vacuum Einstein equations coupled with self-gravitating scalar fields. We prove that the coupled system is linearization stable under some suitable conditions. We also prove the relation between linearization stability and the condition that spacetime admits Killing fields analogous to the work of Moncrief.

PACS numbers: 04.40. + c

In this paper, we extend the work of Fischer–Marsden[1] and Moncrief[2] on linearization stability of vacuum spacetimes to the case of gravity coupled with self-gravitating scalar fields. Such a system is considered by Francaviglia[3] where he proved the "existence" results following Fisher–Marsden.[4] We refer to Francaviglia[3] for details regarding Lagrangian and Hamiltonian formulations of the system. The sign conventions we follow are those of Ref. 1.

The configuration space for the system is $\mathcal{M}\times\mathcal{F}$ where $\mathcal{M}$ is the space of Riemannian metrics on M, and $\mathcal{F}$ is the space of C^∞ functions on M, M being a three-dimensional compact orientable Riemannian manifold without boundary.

The Einstein equation is

$$G_{\mu\nu}=\chi T_{\mu\nu},\quad \text{where } G_{\mu\nu}=R_{\mu\nu}-\tfrac{1}{2}R\,g_{\mu\nu},$$

is the classical Einstein tensor and T is the stress-energy tensor. In our problem, the scalar field is described by

$$T_{\mu\nu}=-\beta\,(2\phi_{,\mu}\phi_{,\nu}-g_{\mu\nu}(\phi_{,\rho}\phi^{,\rho}+m^2\phi^2)),$$

m^2,β being two positive constants related respectively to the mass of the field and to the choice of units. As remarked in Ref. 3, this case applies to π^0 mesons ($m\neq 0$) and to the Brans–Dicke field. If β is allowed to take negative values, the validity of the results can be extended to the C field proposed by Hoyle–Narlikar, which is related to the so-called "steady-state universe."

In our system, the evolution equations can be derived from the variational principle

$$I=\int\left[\pi^{ij}(\partial g_{ij}/\partial t)+\sigma(\partial\phi/\partial t)-N\cdot\mathscr{H}_T-X\cdot\mathscr{J}_T\right]dt,$$

where π is the momentum density conjugate to g, $\sigma=-4\beta\gamma\mu_g$ being scalar density conjugate to ϕ. π,σ are to be obtained by using the so-called Christodoulou–DeWitt metric defined in Ref. 3. We refer to it for details. μ_g denotes the volume element corresponding to g. $\mathscr{H}_T$ is the total Hamiltonian defined by

$$\mathscr{H}_T=\mathscr{H}_G+\mathscr{H}_F,$$

where

$$\mathscr{H}_G=\pi'\cdot\pi'-\tfrac{1}{2}(\mathrm{tr}\pi')^2-R\,(g),\tag{1}$$

and

$$\mathscr{H}_F=-2\beta\,(\gamma^2+A\,(\phi))\mu_g\,.$$

The expression for $\mathscr{H}_F$ follows from the stress-energy tensor (cf., Ref. 3, p. 512). Here $\pi=\pi'\otimes\mu_g$ and $A\,(\phi)=\phi_{,i}\phi^{,i}+m^2\phi^2$. Lastly $\mathscr{J}_T$ is the moments constraint,

$$\mathscr{J}_T=\mathscr{J}_G+\mathscr{J}_F,$$

where

$$\mathscr{J}_G=-2(\delta_g\pi)=2\pi_i^{\,j}|_j\,,\tag{2}$$

and

$$\mathscr{J}_F=-\sigma\phi_{,i}\,.$$

The signs here follow the sign convention of the shift vector field in Ref. 1.

The constraint equations are

$$\mathscr{H}_T=0\quad\text{and}\quad\mathscr{J}_T=0.$$

The evolution equations derived from the variational principle can be written in a compact form as

$$(\partial/\partial t)\begin{pmatrix}g\\ \phi\\ \pi\\ \sigma\end{pmatrix}=J\circ D\Phi^*(N,X),\tag{3}$$

where $\Phi=(\mathscr{H}_F,\mathscr{J}_T)$ and J being the antisymmetric matrix

$$\begin{pmatrix}0 & I\\ -I & 0\end{pmatrix};$$

N is the lapse function and X is the shift vector field. * denotes the adjoint operator.

$$D\Phi^*(N,X)=D\mathscr{H}_T^*\cdot N+D\mathscr{J}_T^*\cdot X\,.$$

We obtain

$$\begin{aligned}&D\mathscr{H}_T^*\cdot N\\ &=[-NS_g(\pi,\pi)+(N\ \text{Eing–Hess}\ N-g\Delta N)^{\#}\mu_g\\ &\quad+\beta N\,(2\tilde{\phi}-gA\,(\phi))\mu_g-\beta N\gamma^2 g\mu_g,\\ &\quad-4\beta N\,(\Delta\phi+m^2\phi)\,\mu_g+4\beta\,(\nabla N\cdot\nabla\phi^{\#})\mu_g,\\ &\quad 2N\,(\pi'-\tfrac{1}{2}(\mathrm{tr}\pi')g),(-\sigma' N/4\beta))].\end{aligned}\tag{4}$$

See Ref. 1 for notations. In addition,

$$\tilde{\phi}\in S_2(M),\quad \tilde{\phi}_{ij}=\phi_{,i}\cdot\phi_{,j},\quad i,j=1,2,3;\quad \sigma=\sigma'\otimes\mu_g\,.$$

To find $D\mathscr{J}_T^*\cdot X$, we proceed as in Ref. 1 for the term $\sigma\nabla\phi$, and get

 0022-2488/81/020343-05$1.00

$$D\mathcal{J}_T^*\cdot X = (L_X\pi, -L_X g, +L_X\sigma, -L_X\phi).$$

Here $\mathcal{J}_T = \mathcal{J}_T(g,\phi,\pi,\sigma)$. We write

$$D\mathcal{J}_T^*\cdot X = (L_X\pi, L_X\sigma, -L_X g, -L_X\phi), \tag{5}$$

changing the order for consistency of notation. Thus evolution equations are obtained as

$$\partial g/\partial t = 2N(\pi' - \tfrac{1}{2}(\mathrm{tr}\pi')g) - L_X g, \tag{6}$$

$$\partial\pi/\partial t = NS_g(\pi,\pi) - [N \text{ Eing–Hess } N - g\Delta N]^{\#}\mu_g + \beta N\gamma^2 g\mu_g - \beta N(2\tilde{\phi} - gA(\phi))\mu_g - L_X\pi, \tag{7}$$

$$\partial\phi/\partial t = -\sigma' N/4\beta - L_X\phi, \tag{8}$$

$$\partial\sigma/\partial t = 4\beta N(\Delta\phi + m^2\phi)\mu_g - 4\beta(\nabla N\cdot\nabla\phi^{\#})\mu_g - L_X\sigma. \tag{9}$$

The negative signs of Lie derivatives in Eqs. (6)–(9) are due to the sign of the shift vector field in Ref. 1. We now derive conservation laws analogous to Refs. 1 and 3.

We have, for diffeomorphism η on M,

$$\mathcal{H}(\eta^*g, \eta^*\pi, \eta^*\phi, \eta^*\sigma) = \eta^*\mathcal{H}(g,\pi,\phi,\sigma),$$

and

$$\mathcal{J}(\eta^*g, \eta^*\pi, \eta^*\phi, \eta^*\sigma) = \eta^*\mathcal{J}(g,\pi,\phi,\sigma).$$

Thus, if η_t is a curve in $\mathcal{D}(M)$ with $\eta_0 = id$ and

$$(d/dt)\eta_t\big|_{t=0} = X,$$

then

$$D\mathcal{H}(g,\pi,\phi,\sigma)\cdot(L_X g, L_X\pi, L_X\phi, L_X\sigma) = L_X\mathcal{H}(g,\pi,\phi,\sigma),$$

$$D\mathcal{J}(g,\pi,\phi,\sigma)\cdot(L_X g, L_X\pi, L_X\phi, L_X\sigma) = L_X\mathcal{J}(g,\pi,\phi,\sigma).$$

Hence

$$D\Phi(g,\pi,\phi,\sigma)\cdot(L_X g, L_X\pi, L_X\phi, L_X\sigma) = L_X\Phi(g,\pi,\phi,\sigma).$$

We now consider

$$\frac{\partial\mathcal{H}(g,\phi,\pi,\sigma)}{\partial t}$$

$$= D\mathcal{H}(g,\phi,\pi,\sigma)\cdot[(\partial g/\partial t),(\partial\phi/\partial t),(\partial\pi/\partial t),(\partial\sigma/\partial t)]$$

$$= D\mathcal{H}\cdot J\circ\left[D\Phi^*\cdot\begin{pmatrix}N\\X\end{pmatrix}\right]$$

(Change of order of variables is for convenience)

$$= D\mathcal{H}[J\circ D\mathcal{H}(g,\phi,\pi,\sigma)^*\cdot N + D\mathcal{J}(\gamma,\phi,\pi,\sigma)^*\cdot X]$$

$$= D\mathcal{H}\cdot J\circ D\mathcal{H}^*\cdot N + D\mathcal{H}\cdot(-L_X g, -L_X\phi, -L_X\pi, -L_X\sigma)$$

$$= D\mathcal{H}\cdot J\circ D\mathcal{H}^*\cdot N + D\mathcal{H}\cdot(-L_X g, -L_X\phi, -L_X\pi, -L_X\sigma).$$

Then

$$\frac{\partial\mathcal{H}}{\partial t} = D\mathcal{H}\cdot J\circ D\mathcal{H}^*\cdot N - L_X\mathcal{H}(g,\phi,\pi,\sigma).$$

We now compute $D\mathcal{H}\cdot J\circ D\mathcal{H}^*\cdot N$, as in Ref. 1:

$$D\mathcal{H}\cdot J\circ D\mathcal{H}^*\cdot N$$

$$= D\mathcal{H}(g,\pi,\phi,\sigma)\cdot(D_\pi\mathcal{H}^*\cdot N, -D_g\mathcal{H}^*\cdot N, D_\sigma\mathcal{H}^*\cdot N, -D_\phi\mathcal{H}^*\cdot N)$$

$$= D_g\mathcal{H}\cdot(D_\pi\mathcal{H}^*\cdot N) - D_\pi\mathcal{H}\cdot(D_g\mathcal{H}^*\cdot N) + D_\phi\mathcal{H}\cdot(D_\sigma\mathcal{H}^*\cdot N) - D_\sigma\mathcal{H}\cdot(D_\phi\mathcal{H}^*\cdot N),$$

where $\mathcal{H} = \mathcal{H}_G + \mathcal{H}_F$.

Out of these,

$$D_\phi\mathcal{H}_G^*\cdot N = 0 = D_\sigma\mathcal{H}_G^*\cdot N,$$

$$D_\phi\mathcal{H}_G = 0, \quad D_\sigma\mathcal{H}_G = 0,$$

$$D_\pi\mathcal{H}_F^*\cdot N = 0, \quad D_\pi\mathcal{H}_F = 0, \quad D_\sigma\mathcal{H}_F \neq 0.$$

Thus

$$D\mathcal{H}\cdot J\circ D\mathcal{H}^*\cdot N$$

$$= D_g\mathcal{H}_G\cdot(D_\pi\mathcal{H}_G^*\cdot N) + D_g\mathcal{H}_F\cdot(D_\pi\mathcal{H}_G^*\cdot N) - [D_\pi\mathcal{H}_G\cdot D_g\mathcal{H}_G^*\cdot N + D_\pi\mathcal{H}_G\cdot D_g\mathcal{H}_F^*\cdot N] + D_\phi\mathcal{H}_F\cdot(D_\sigma\mathcal{H}_F^*\cdot N) - D_\sigma\mathcal{H}_F\cdot(D_\phi\mathcal{H}_F^*\cdot N),$$

Now

$$D_g\mathcal{H}_F\cdot h = -\beta\gamma^2 g\cdot h + \beta(2\tilde{\phi} - gA(\phi))\cdot h,$$

therefore

$$D_g\mathcal{H}_F\cdot(D_\pi\mathcal{H}_G^*\cdot N)$$

$$= D_g\mathcal{H}_F\cdot(N\partial_\pi\mathcal{H}_G) \quad \text{(cf. Ref. 1)}$$

$$= D_g\mathcal{H}_F\cdot(-2k\cdot N) = -2ND_g\mathcal{H}_F\cdot k \tag{10}$$

$$= -2N[-\beta\gamma^2 g\cdot k + \beta(2\tilde{\phi} - gA(\phi))\cdot k]$$

$$= 2N\beta[\gamma^2 g\cdot k - 2\tilde{\phi}\cdot k + A(\phi)g\cdot k].$$

$$D_\pi\mathcal{H}_G\cdot\omega = \partial_\pi\mathcal{H}_G\cdot\omega = -2k\cdot\omega,$$

therefore

$$D_\pi\mathcal{H}_G\cdot(D_g\mathcal{H}_G^*\cdot N)$$

$$= -2k\cdot[-\beta\gamma^2 Ng\mu_g + \beta N(2\tilde{\phi} - gA(\phi))\mu_g] \tag{11}$$

$$= 2N\beta(\gamma^2 g\cdot k - 2\tilde{\phi}\cdot k + A(\phi)g\cdot k)\mu_g.$$

So that

$$D_g\mathcal{H}_F\cdot D_\pi\mathcal{H}_G^*\cdot N - D_\pi\mathcal{H}_G\cdot D_g\mathcal{H}_F^*\cdot N = 0, \tag{12}$$

$$D_\phi\mathcal{H}_F\cdot\psi = -2\beta(2m^2\phi\cdot\psi + 2\nabla\phi^{\#}\cdot\nabla\psi) = -4\beta(m^2\phi\cdot\psi + \nabla\phi^{\#}\cdot\nabla\psi).$$

$$D_\sigma\mathcal{H}_F^*\cdot N = \frac{D_\gamma\mathcal{H}_F^*\cdot N}{-4\beta} = -\frac{N\sigma'}{4\beta},$$

$$(D_\gamma\mathcal{H}_F\cdot\epsilon = -4\beta\gamma\cdot\epsilon\mu_g = \sigma\cdot\epsilon).$$

Therefore

$$D_\phi\mathcal{H}_F\cdot D_\sigma\mathcal{H}_F^*\cdot N = (m^2\phi N\sigma' + \nabla\phi^{\#}\cdot\nabla(N\sigma')), \tag{13}$$

$$D_\phi\mathcal{H}_F^*\cdot N = -4\beta N(\Delta\phi + m^2\phi) + 4\beta(\nabla N\cdot\nabla\phi^{\#}).$$

Therefore

$$-D_\sigma\mathcal{H}_F\cdot(D_\phi\mathcal{H}_F^*\cdot N) = -N\sigma'(\Delta\phi + m^2\phi) + \sigma(\nabla N\cdot\nabla\phi^{\#}). \tag{14}$$

Hence

$$D_\phi\mathcal{H}_F\cdot(D_\sigma\mathcal{H}_F^*\cdot N) - D_\sigma\mathcal{H}_F\cdot(D_\phi\mathcal{H}_F^*\cdot N)$$

$$= \nabla\phi^{\#}\cdot(\nabla N)\sigma + \nabla\phi^{\#}\cdot N(\nabla\sigma) - N\sigma\Delta\phi + \sigma\nabla N\cdot\nabla\phi^{\#}$$

$$= -N\sigma\cdot\Delta\phi + N\nabla\phi^{\#}\cdot\nabla\sigma + 2\sigma\cdot\nabla N\cdot\nabla\phi^{\#}. \tag{15}$$

If we calculate $\delta(N^2\mathcal{J}_F)$, we get

$$\delta(N^2\sigma\nabla\phi^{\#}) = -(N^2\sigma\phi^{,i})|_i .$$

Hence

$$\begin{aligned}\delta(N^2\mathscr{J}_F) &= +2N\nabla N\cdot\sigma\phi^{,i} + N^2(\sigma\phi^{,i})_{,i}\\ &= 2N\nabla N\cdot\sigma\nabla\phi^{\#} + N^2[-\sigma\Delta\phi + \nabla\phi^{\#}\cdot\nabla\sigma]\\ &= N[2\nabla N\cdot\sigma\nabla\phi^{\#} - N\sigma\Delta\phi + N\nabla\phi^{\#}\cdot\nabla\sigma].\end{aligned}$$

Therefore

$$\begin{aligned}&D_\phi\mathscr{H}_F\cdot(D_\sigma\mathscr{H}_F^*\cdot N) - D_\sigma\mathscr{H}_F\cdot(D_\phi\mathscr{H}_F^*\cdot N)\\ &\quad = (1/N)\delta(N^2\mathscr{J}_F)\\ &\quad = -(1/N)\mathrm{div}(N^2\mathscr{J}_F). \qquad (16)\end{aligned}$$

Thus combining Eqs. (12) and (16) and using results of Ref. 1 for the remaining terms in $D\mathscr{H}\cdot J\circ D\mathscr{H}^*\cdot N$, we get

$$\begin{aligned}&D\mathscr{H}\cdot J\circ D\mathscr{H}^*\cdot N\\ &\quad = -(1/N)\mathrm{div}(N^2\mathscr{J}_F) - (1/N)\mathrm{div}(N^2\mathscr{J}_G),\end{aligned}$$

and hence

$$\begin{aligned}&(\partial\mathscr{H}/\partial t)\\ &= -(1/N)\mathrm{div}(N^2\mathscr{J}_F) - (1/N)\mathrm{div}(N^2\mathscr{J}_G) - L_X\mathscr{H},\end{aligned}$$

or

$$(\partial\mathscr{H}/\partial t) + (1/N)\mathrm{div}(N^2\mathscr{J}) + L_X\mathscr{H} = 0. \qquad \text{(CI)}$$

For the evolution equation for $\mathscr{J}$, we write, for vector field Y, independent of t,

$$\begin{aligned}&\frac{d}{dt}\int\langle Y,\mathscr{J}(g,\phi,\pi,\sigma)\rangle\\ &\quad = \int\langle Y,(d/dt)\mathscr{J}(g,\phi,\pi,\sigma)\rangle\\ &\quad = \int\langle Y,D\mathscr{J}\cdot[(\partial g/\partial t),(\partial\phi/\partial t),(\partial\pi/\partial t),(\partial\sigma/\partial t)]\rangle\\ &\quad = \int\left\langle Y,D\mathscr{J}\cdot J\circ D\Phi^*\cdot\begin{pmatrix}N\\ X\end{pmatrix}\right\rangle\\ &\quad = \int\left\langle D\Phi\cdot J^*\circ D\mathscr{J}^*\cdot Y,\begin{pmatrix}N\\ X\end{pmatrix}\right\rangle\\ &\quad = -\int\left\langle D\Phi\cdot J\circ D\mathscr{J}^*\cdot T,\begin{pmatrix}Y\\ X\end{pmatrix}\right\rangle \quad (J^* = -J)\\ &\quad = \int\langle D\Phi\cdot(L_Yg, L_Y\phi, L_Y\pi, L_Y\sigma),(N,X)\rangle\\ &\quad = -\int Y(dN)\mathscr{H} - \int\langle Y, L_X\mathscr{J}\rangle \quad \text{(cf. Ref. 1).}\end{aligned}$$

Thus we get

$$\frac{d\mathscr{J}}{dt} + L_X\mathscr{J} + (dN)\mathscr{H} = 0. \qquad \text{(CII)}$$

This is the required evolution equation for $\mathscr{J}$. From (CI) and (CII) we get analogue of Theorem 3.1, Ref. 1.

We now prove the linearization stability of the equations:

Theorem: A solution $({}^{(4)}g,\bar{\phi})$ of the coupled system is linearization stable if the following conditions are satisfied:

(i) $\mathrm{tr}\pi$ is a constant multiple of the volume element on M;

(ii) One of g, ϕ, π, σ is nontrivial or g is not flat;

(iii) There are no simultaneuous symmetries of g, ϕ, π, σ on M, i.e., if $L_Xg = 0, L_X\phi = 0, L_X\pi = 0$ and $L_X\sigma = 0$, then $X = 0$.

Proof: As argued in Ref. 1, it is enough to show that $D\Phi^*$ is injective and has injective symbol. Symbol of $D\Phi^*$:

$$\begin{aligned}\Sigma_\xi(s,V) = &((-\xi\otimes\xi + \|\xi\|^2 g)\mu_g\cdot s + (\xi_1\pi^{1i}\delta^j_k + \xi_1\pi^{1j}\delta^i_k\\ &- \xi_k\pi^{ij})V^k{}_{,}\xi\otimes\nabla\phi)s\ \mu_g,(g_{ik}\xi_j + g_{jk}\xi_i)V^k,0).\end{aligned}$$

Thus if $\Sigma_\xi(s,V) = 0$, then third component zero gives $V = 0$. Then from the first component, we get $s = 0$. Thus the symbol is injective.

We now show that $D\Phi^*$ is injective. Let $D\Phi^*(N,X) = 0$. To show $N = 0$ and $X = 0$, $D\Phi^*(N,X) = 0$ is equivalent to the following four equations:

$$\begin{aligned}&-NS_g(\pi,\pi) + (N\ \text{Eing–Hess}\ N - g\Delta N)^{\#}\mu_g\\ &\quad + \beta N(2\tilde{\phi} - gA(\phi) - \gamma^2 g)\mu_g + L_X\pi = 0, \qquad (17)\\ &4\beta N(\Delta\phi + m^2\phi) - 4\beta\nabla N\cdot\nabla\phi^{\#} + L_X\sigma = 0, \qquad (18)\\ &2N(\pi' - \tfrac{1}{2}(\mathrm{tr}\pi')g) - L_Xg = 0, \qquad (19)\\ &(\sigma' N/4\beta) + L_X\phi = 0. \qquad (20)\end{aligned}$$

Taking tr of Eq. (17) we get,

$$\begin{aligned}&(N/2)\mathscr{H}_G + [-2\Delta N + \beta N(2\phi_{,i}\phi^{,i} - 3A(\phi))\\ &\quad - 3\beta N\gamma^2]\mu_g + \mathrm{tr}L_X\pi = 0.\end{aligned}$$

But

$$\begin{aligned}&\mathscr{H}_T = \mathscr{H}_G + \mathscr{H}_F\\ &\quad = 0 \Rightarrow \mathscr{H}_G = -\mathscr{H}_F = 2\beta(\gamma^2 + A(\phi))\mu_g.\end{aligned}$$

Therefore

$$(N/2)\mathscr{H}_G = N\beta(\gamma^2 + A(\phi))\mu_g.$$

So the above equation becomes,

$$\begin{aligned}&[N\beta\gamma^2 + N\beta A(\phi) - 2\Delta N + 2\beta N\ \phi_{,i}\phi^{,i} - 3N\beta A(\phi)\\ &\quad - 3N\beta\gamma^2]\mu_g + \mathrm{tr}L_X\pi = 0.\end{aligned}$$

Using $A(\phi) = \phi_{,i}\phi^{,i} + m^2\phi^2$, we get

$$-2\beta N\gamma^2 - 2N\beta m^2\phi^2 - 2(\Delta N)\mu_g + \mathrm{tr}L_X\pi = 0,$$

i.e.,

$$(\Delta N)\mu_g + \beta N(m^2\phi^2 + \gamma^2) - \tfrac{1}{2}\mathrm{tr}L_X\pi = 0. \qquad (21)$$

Taking tr of Eq. (19) gives

$$-N\,\mathrm{tr}\pi' + 2\delta_g X = 0,$$

or

$$\mathrm{div}X = -\tfrac{1}{2}N\,\mathrm{tr}\pi', \qquad (22)$$

$$\begin{aligned}\mathrm{tr}(L_X\pi) &= X\cdot d\,\mathrm{tr}\pi - \pi\cdot L_Xg + (\mathrm{div}X)(\mathrm{tr}\pi)\\ &= -\pi\cdot L_Xg - \tfrac{1}{2}N(\mathrm{tr}\pi')^2,\end{aligned}$$

from (22) and using $\mathrm{tr}\pi = \text{constant}$. Hence (21) becomes

$$\Delta N + N\beta(\gamma^2 + m^2\phi^2) + \tfrac{1}{2}(L_Xg)\cdot\pi' + \tfrac{1}{4}N(\mathrm{tr}\pi')^2 = 0.$$

From (19),

$$\begin{aligned}\tfrac{1}{2}(L_Xg)\pi' &= \tfrac{1}{2}\pi'\cdot 2N(\pi' - \tfrac{1}{2}(\mathrm{tr}\pi')g)\\ &= N(\pi'\cdot\pi' - \tfrac{1}{2}(\mathrm{tr}\pi')^2).\end{aligned}$$

Hence

$$\begin{aligned}\tfrac{1}{2}(L_Xg)\pi' + \tfrac{1}{4}N(\mathrm{tr}\pi')^2 &= N(\pi'\cdot\pi' - \tfrac{1}{4}(\mathrm{tr}\pi')^2)\\ &= N(\pi' - \tfrac{1}{2}(\mathrm{tr}\pi')g)\cdot(\pi' - \tfrac{1}{2}(\mathrm{tr}\pi')g).\end{aligned}$$

Thus we get finally

$$\Delta N[(\pi' - \tfrac{1}{2}(\mathrm{tr}\pi')g)\cdot(\pi' - \tfrac{1}{2}(\mathrm{tr}\pi')g) + \beta(\gamma^2 + m^2\phi^2)]N = 0.$$

Now, since β is positive, the coefficient of N is positive definite and we get N is constant. Then we get $N = 0$ if one of π, σ, ϕ is nonzero.

If π, σ, ϕ are all identically zero, then $\mathscr{H} = 0$ gives $R = 0$ and (17) gives (since N is constant)

$$NR^{ij} - \tfrac{1}{2}Ng^{ij}\cdot R = 0 \Rightarrow NR^{ij} = 0.$$

Then (g is not flat $\Longleftrightarrow R^{ij} \neq 0$) $\Rightarrow N = 0$.

Thus condition (ii) forces N to be zero. Now, $N = 0$ gives using Eqs. (17)–(20),

$$L_X\pi = 0, \quad L_X\sigma = 0, \quad L_X g = 0, \quad L_X\phi = 0.$$

Then condition (iii) and above equations imply that $X = 0$. $D\Phi^*$ is thus injective. Linearization stability is thus proved.

Moncrief's Condition: We now wish to prove the following theorem:

Theorem: (Analogue of Th. 5.5, Ref. 1): Let $({}^{(4)}g, \bar{\phi})$ be a solution to the field equations $\mathrm{Ein}({}^{(4)}g) = T$. Let $\Sigma_0 = i_0(M)$ be a compact Cauchy hypersurface with induced metric g_0, scalar function ϕ_0 and canonical conjugate quantities π_0, σ_0. $\phi_0 = i_0^*\bar{\phi}$). Then $\mathrm{Ker}D\Phi\,(g_0, \phi_0, \pi_0, \sigma_0)^*$ is isomorphic to the space of simultaneous Killing vector fields of ${}^{(4)}g$ and $\bar{\phi}$.

In fact $(Y_\perp, -Y_\parallel) \in \mathrm{Ker}D\Phi\,(g_0, \phi_0, \pi_0, \sigma_0)^*$ iff there exists a simultaneous Killing vector field ${}^{(4)}Y$ of $({}^{(4)}g, \bar{\phi})$ whose normal and tangential components to Σ_0 are $Y_\perp$ and $Y_\parallel$.

$\bar{\phi}$ is a function on V_4 appearing in $T_{\mu\nu}$.

Proof (analogous to that in Ref. 1): Necessary condition: Let F_t be the flow of ${}^{(4)}Y$. For t in a neighbourhood of 0, $i_t = F_t \circ i_0$ is a well-defined one-parameter family of spacelike embeddings with generator ${}^{(4)}Y_t = {}^{(4)}Y \circ i_t$. Let $(Y_\perp(t), Y_\parallel(t))$ be the normal and tangential components of ${}^{(4)}Y_t$. Let $(g(t), \pi(t), \phi(t), \sigma(t))$ be the usual quantities with their conjugates induced on Σ_t by $({}^{(4)}g, \bar{\phi})$. For a family of embeddings given by $i_t = F_t \circ i_0$, this will be the same as metrics, scalar functions and their conjugates induced on Σ_0 by $F_t^{*(4)}g$ and $F_t^*\bar{\phi}$.

Since ${}^{(4)}Y$ is a Killing vector field of ${}^{(4)}g$ and $\bar{\phi}$, $F_t^{*(4)}g = {}^{(4)}g$ and $F_t^*\bar{\phi} = \bar{\phi}$, so $g(t) = g_0$, $\pi(t) = \pi_0$, $\phi(t) = \phi_0$ and $\sigma(t) = \sigma_0$, $\forall t$. Hence by the adjoint form of the evolution equations,

$$0 = \begin{pmatrix} \partial g/\partial t \\ \partial\pi/\partial t \\ \partial\phi/\partial t \\ \partial\sigma/\partial t \end{pmatrix} = J \circ D\Phi\,((g(t), \pi(t), \phi(t), \sigma(t))^* \begin{pmatrix} Y_\perp(t) \\ -Y_\parallel(t) \end{pmatrix}.$$

Evaluating at $t = 0$, we get $(Y_\perp, -Y_\parallel) \in \mathrm{Ker}D\Phi^*$. Sufficient condition: For sufficiency, we require the following analogue of Proposition 4.7, Ref. 1, whose proof can be easily extended in our case:

Proposition: Let ${}^{(4)}Y$ be a vector field on V_4 with flow F_t and let $i_t = F_t \circ i_0$. Let ${}^{(4)}h = L_{(4)_\nu}{}^{(4)}g_0$ and $\bar{\psi} = L_{(4)_\nu}\bar{\phi}$. Let $g(t)$, $\pi(t)$, $\phi(t)$, $\sigma(t)$ be the usual quantities on $\Sigma_t = i_t(M)$ and let $(h(t), \omega(t), \psi(t), \tau(t))$ be the infinitesimal deformations of (g, π, ϕ, σ) induced on Σ_t by ${}^{(4)}h$ and $\bar{\psi}$.

Then

$$\begin{pmatrix} h(t) \\ \omega(t) \\ \psi(t) \\ \tau(t) \end{pmatrix} = J \circ D\Phi\,(g(t), \pi(t), \phi(t), \sigma(t))^* \begin{pmatrix} Y(t) \\ Y_{\mathrm{shift}}(t) \end{pmatrix}.$$

Proof of sufficiency: Let $(Y_\perp, -Y_\parallel) \in \mathrm{Ker}D\Phi^*$. We wish to extend $(Y_\perp, -Y_\parallel)$ to a simultaneous Killing field ${}^{(4)}Y$. Choose a slicing i_t and let N_t, X_t be its lapse and shift. To define $Y_\perp$ and $Y_\parallel$, take the perp–perp and perp–parallel projections of Killing equations $L_{(4)_\nu}{}^{(4)}g = 0$ and let $Y_\perp$ and $Y_\parallel$ be subjected to the condition

$$L_{Y_\perp}{}^{(4)}{}_Z\bar{\phi} = L_{T_t \cdot Y_{\mathrm{shift}}}\bar{\phi},$$

i.e.,

$$L_{(4)_\nu}\bar{\phi} = 0.$$

This gives

$$(\partial Y_\perp/\partial_t) + L_X Y_\perp + L_{Y_\parallel}N = 0,$$

$$-(\partial Y_\parallel/\partial t) - L_X Y_\parallel + N\,\mathrm{grad}\,Y_\perp - Y_\perp\,\mathrm{grad}N = 0,$$

subjected to $L_{(4)_\nu}\bar{\phi} = 0$. For given $N(t,x)$, $X(t,x)$ and initial conditions $(Y_\perp, Y_\parallel)$ together with the additional constraint, these equations define a unique $Y_\perp$, $Y_\parallel$ on V_4 with the given initial conditions. Thus we get a vector field ${}^{(4)}Y$ on V_4 with these normal and tangential components on each hypersurface and satisfying $L_{(4)_\nu}\bar{\phi} = 0 = \bar{\psi}$ (say). Let $(h(t), \omega(t), \psi(t), \tau(t), U(t), V(t))$ be the induced deformations of $(g, \pi, \phi, \sigma, N, X)$. By construction, $\{4\}_{h_{\perp\perp}} = 0$, $\{4\}_{h_{\perp\parallel}} = 0$.

Hence

$$U(t) = \tfrac{1}{2}N(t)\,{}^{(4)}h_{\perp\perp}(t) = 0,$$

and

$$V(t) = N(t)\,{}^{(4)}h_{\perp\parallel}(t) = 0.$$

Thus $(h(t), \omega(t), \psi(t), \tau(t))$ satisfies the linear system

$$(\partial/\partial t)\begin{pmatrix} h \\ \omega \\ \psi \\ \tau \end{pmatrix} = J \circ D\left[D\Phi^*\begin{pmatrix} N \\ X \end{pmatrix}\right]\begin{pmatrix} h \\ \omega \\ \psi \\ \tau \end{pmatrix}.$$

Hence, by the above proposition, on Σ_0,

$$\begin{pmatrix} h(0) \\ \omega(0) \\ \psi(0) \\ \tau(0) \end{pmatrix} = J \circ D\Phi^*\begin{pmatrix} Y_\perp(0) \\ Y_\parallel(0) \end{pmatrix} = 0.$$

Thus $(h(t), \omega(t), \psi(t), \tau(t)) = (0,0,0,0)\forall t$. Hence, since $h(t) = 0, h_{\perp\perp}(t) = 0$, and $h_{\perp\parallel}(t) = 0$, ${}^{(4)}h = 0$, and $\bar{\psi} = 0$, by construction. Thus ${}^{(4)}Y$ is a simultaneous Killing field as required.

It also follows that the dimension of $\mathrm{Ker}D\Phi^*$ is equal to the number of linearly independent nontrivial simultaneous Killing fields of ${}^{(4)}g$ and $\bar{\phi}$.

Remarks: (1) Since our evolution equations are in the adjoint form, we can prove Moncrief's splitting[5] as given in Ref. 1, Theorem 6.1, in our case also.

(2) We hope that the results regarding linearization stability can be proved in the noncompact case. The procedure to be followed may be similar to that in Ref. 6 due to some technical difficulties as explained there.

[1]A.E. Fischer and J.E. Marsden, Nuovo Cimento (1978).
[2]V. Moncrief, J. Math. Phys. **16**, 493 (1975).
[3]M. Francaviglia, "On the symplectic formulation of the Einstein system of evolution in presence of a self-gravitating scalar field," Springer Lecture notes in Mathematics, No. 570, p. 498.
[4]A.E. Fischer and J.E. Marsden, J. Math. Phys. **13**, 546 (1972).
[5]V. Moncrief, J. Math. Phys. **16**, 1556 (1975).
[6]Y. Choquet-Bruhat, A.E. Fischer, and J.E. Marsden, Nuovo Cimento (1978).

Statistical mechanics and the gravothermal catastrophe

P. Cally and J. J. Monaghan
Mathematics Department, Monash University, Clayton 3168, Australia

(Received 17 July 1979; accepted for publication 24 December 1979)

Jensen's inequality is applied to the canonical partition function of a self-gravitating system to determine the best independent particle potential. The inequality allows the stability to be analyzed very easily. We recover the results of Lynden-Bell and Wood for the onset of an instability in an isothermal sphere in a heat bath. Our eigenvalue analysis leads to results very similar to those of Horwitz and Katz, but we differ in the description of the $l = 1$ perturbation.

PACS numbers: 05.20. – y, 04.20. – q

I. INTRODUCTION

In a series of papers Horwitz and Katz[1-3] have analyzed the statistical mechanics of a self-gravitating system by approximately evaluating the various partition functions. Their procedure involves replacing the Boltzmann factor by its representation as a functional integral. They evaluate the resulting multiple integral by a saddle point method. The necessary condition for the stationary point is that the particles move independently in the mean field. The requirement that the stationary point gives a maximum can be made the basis of a stability analysis. In this way they recovered and extended the results of Lynden-Bell and Wood.[4]

We show, in this paper, that the results found by Horwitz and Katz for the canonical partition function can be obtained very easily by applying Jensen's inequality[5] to the configurational integral.

II. A VARIATIONAL METHOD

We work with the canonical partition function Z and assume that the potential is of such a form that Z exists. A discussion of these assumptions is given by Ipser[6] and by Horwitz and Katz[2]. Z is defined for N equal particles by

$$Z = \frac{1}{N!}\int e^{-\beta E}\, d\Omega\,, \quad \beta = \frac{1}{kT}\,, \tag{1}$$

where quantum factors have been omitted because they are unimportant for the systems we consider. Performing the integration over momenta, and defining $d\Omega_s$ to be the spatial part of the phase space volume element, we find

$$Z = (2\pi mkT)^{3N/2} Q\,, \tag{2}$$

where the mass of each particle is m and Q, the configurational integral, is defined by

$$Q = \frac{1}{N!}\int e^{-\beta V}\, d\Omega_s\,, \tag{3}$$

and V is the potential energy. In practice, V is usually too complicated to allow Q to be evaluated. To estimate Q we suppose that we have a reference system of N particles with the same mass as before, but with a potential energy U which allows the configurational integral to be evaluated. Now write Eq. (3) as

$$Q = \frac{1}{N!}\int e^{-\beta U}\cdot e^{-\beta(V-U)}\, d\Omega_s\,, \tag{4}$$

and regard Eq. (4) as defining, apart from a constant multiplier, the average of $\exp[-\beta(V-U)]$ with a probability density $\exp(-\beta U)$. Since the exponential is a convex function, Jensen's inequality can be applied, and we find

$$Q \geqslant I = \frac{1}{N!}\int e^{-\beta U} d\Omega_s \cdot \exp(-\beta\langle V-U\rangle)\,, \tag{5}$$

where $\langle\ \rangle$ denotes an average using the probability density $\exp(-\beta U)$. Denoting the Helmholtz free energy calculated using the potential U by $A(U)$, Eq. (5) is equivalent to

$$A(V) \leqslant A(U) + \langle V-U\rangle\,.$$

If we choose the functional form of U with arbitrary parameters, these parameters can be chosen by adjusting them to make the right hand side of Eq. (5) a maximum[7]. A better procedure is to use Eq. (5) to choose the functional form. In particular, if U is a sum of independent particle potentials, then the stationary point of the functional I should give the best potential provided that I is then a maximum. If it is not a maximum, then in the neighborhood of the independent particle potential another potential exists which will give a better estimate of Q and a lower free energy. The system will then move away from the configuration defined by the stationary point of I.

The exact Hamiltonian of the system is

$$H = \sum_{j=1}^{N}\frac{p_j^2}{2m} + \tfrac{1}{2}\sum_{\substack{i=1\\ i\neq j}}^{N}\sum_{j=1}^{N} F(|\mathbf{r}_i - \mathbf{r}_j|)\,, \tag{6}$$

where F is a long range pair interaction. In order for Z to be well defined we assume, with Horwitz and Katz, that the system is enclosed within a sphere and the singularity at short range is smoothed out. Because F is long range we expect that the time averaged dynamics of a given particle is very closely approximated by its motion in a mean field. We therefore approximate H by

$$H' = \sum_{j=1}^{N}\frac{p_j^2}{2m} + \sum_{j=1}^{N}\phi(\mathbf{r}_j)\,, \tag{7}$$

so that

$$U = \sum_{j=1}^{N}\phi(\mathbf{r}_j)\,. \tag{8}$$

To find the stationary value of I we replace U by $U + \delta U$ and expand to first order. The first order change in I

 0022-2488/81/020348-04$1.00

is given by

$$(\delta I/I) = \beta \int e^{-\beta U}\, \delta U\, [\langle V-U\rangle - (V-U)]\, d\Omega_s\,. \quad (9)$$

The necessary condition for I to be a maximum is that $\delta I = 0$ for arbitrary δU. It is convenient to set

$$\delta U = \sum_{j=1}^{N} \delta\phi\,(\mathbf{r}_j)\,, \quad (10)$$

where $\delta\phi$ is arbitrary. Substituting Eq. (10) into Eq. (9) we find that $\delta I = 0$ implies

$$\int e^{-\beta\phi}\, \delta\phi\, \Gamma\, d\tau = 0\,, \quad (11)$$

where $d\tau$ is the ordinary volume element,

$$\Gamma = N\zeta^{N-1}\left\{\langle V-U\rangle - \frac{N-1}{\zeta}\int \exp(-\beta\phi')F(|\mathbf{r}-\mathbf{r}'|)\,d\tau' - \frac{(N-1)(N-2)}{2\zeta}\iint \exp[-\beta(\phi+\phi')]F(|\mathbf{r}-\mathbf{r}'|)\,d\tau\,d\tau' + \phi(\mathbf{r}) + \frac{(N-1)}{\zeta}\int \phi\exp(-\beta\phi)\,d\tau'\right\}, \quad (12)$$

and

$$\zeta = \int e^{-\beta\phi}\,d\tau\,,\ \phi \equiv \phi(\mathbf{r})\,,\ \text{and}\ \phi' \equiv \phi(\mathbf{r}')\,. \quad (13)$$

Since $\delta\phi$ is arbitrary, the variational principle requires $\Gamma = 0$. This equation is satisfied by choosing

$$\phi(\mathbf{r}) = \frac{N-1}{\zeta}\int e^{-\beta\phi(\mathbf{r}')}F(|\mathbf{r}-\mathbf{r}'|)\,d\tau'\,. \quad (14)$$

Equation (14) for $\phi(\mathbf{r})$ is the canonical average of the pair interaction at $\mathbf{r}$ due to the other $(N-1)$ particles. For a pure gravitational interaction Eq. (14) is equivalent to

$$\nabla^2\phi = 4\pi G m^2 (N-1)\, e^{-\beta\phi}/\zeta\,. \quad (15)$$

If we anticipate that the number density $n(\mathbf{r})$ is

$$n(\mathbf{r}) = N e^{-\beta\phi}/\zeta\,, \quad (16)$$

we find

$$\nabla^2\phi = 4\pi G\,\frac{(N-1)}{N}\, m^2 n(\mathbf{r})\,. \quad (17)$$

Apart from the factor $(N-1)/N$ which is very close to 1 for the systems we consider, Eq. (17) shows that ϕ/m is the Emden potential. The variational principle therefore establishes that the Emden potential is the best independent particle potential (in the thermodynamic sense) for a self-gravitating system.

III. STABILITY

The stability of the system is determined by the second order variation of I, which is easily shown to be

$$\tfrac{1}{2}\beta^2 \exp(-\beta\langle V-U\rangle)\zeta^N[\langle(\delta U)^2\rangle - \langle(\delta U)^2\rangle + \beta \times\{\langle(\delta U)^2\rangle\langle V-U\rangle - \langle(\delta U)^2(V-U)\rangle\}]\,. \quad (18)$$

The multiplying factor is always positive so the sign of the second variation is determined by

$$L = \langle\delta U\rangle^2 - \langle(\delta U)^2\rangle + \beta\{\langle(\delta U)^2\rangle\langle V-U\rangle - \langle(\delta U)^2(V-U)\rangle\}\,. \quad (19)$$

As before we take $\delta U = \Sigma_{j=1}^{N}\delta\phi(\mathbf{r}_j)$ and the averages in Eq. (19) reduce to integrations over ordinary volume. The resulting algebra can be reduced by expressing integrals as derivatives according to

$$\langle(\delta U)^2(V-U)\rangle = \frac{\partial^2}{\partial\alpha^2}\langle e^{-\alpha\delta U}(V-U)\rangle\,, \quad (20)$$

and noting that if

$$f(\alpha) = \left(\int e^{-\beta\phi-\alpha\delta\phi}\,d\tau\right)^N \quad (21)$$

then

$$f(0) = \zeta^N\,,\quad f'(0) = -N\zeta^N\langle\delta\phi\rangle\,,$$

and

$$f''(0) = \zeta^N[N\langle(\delta\phi)^2\rangle + N(N-1)\langle\delta\phi\rangle^2]\,. \quad (22)$$

The final expression for Eq. (19) can be written

$$\frac{L}{N} = \langle\delta\phi\rangle^2 - \langle(\delta\phi)^2\rangle - \beta[\langle\delta\phi\rangle\langle\delta\phi\rangle^2 - 2\langle\delta\phi\rangle\langle\phi\delta\phi\rangle + (N-1)\langle\delta\phi\delta\phi' F_{rr'}\rangle]\,, \quad (23)$$

where $\langle\ \rangle$ now denotes an average over the ordinary volume where $\langle\ \rangle$ now denotes an average over the ordinary volume with probability density $\propto\exp(-\beta\phi)$ and

$$\langle\delta\phi\delta\phi' F_{rr'}\rangle \equiv \iint e^{-\beta\phi-\beta\phi'}F(|\mathbf{r}-\mathbf{r}'|)\,d\tau\,d\tau'/\zeta^2\,. \quad (24)$$

Finally we note that if

$$q = \delta\phi - \langle\delta\phi\rangle\,, \quad (25)$$

then Eq. (23) can be written

$$(L/N) = -\langle q^2\rangle - \beta(N-1)\langle qq' F_{rr'}\rangle\,, \quad (26)$$

using Eq. (14).

If $L<0$, I is a maximum and the system will be stable. If $L>0$, the system will be unstable. In order to establish the point of instability it is convenient to turn to the associated eigenvalue problem.

IV. THE EIGENVALUE PROBLEM FOR SPHERICAL VARIATIONS

The right-hand side of Eq. (26) when written in full for gravitational interaction becomes

$$-\frac{1}{\zeta}\int e^{-\beta\phi} q\left[q - \frac{\beta(N-1)Gm^2}{\zeta}\int\frac{e^{-\beta\phi'}q(\mathbf{r}')\,d\tau'}{|\mathbf{r}-\mathbf{r}'|}\right]d\tau\,. \quad (27)$$

We now specialize to purely spherical variations so that q is a function of r alone. Equation (27) suggests that we consider the eigenvalue problem:

$$q_k(r) - \beta(N-1)\frac{Gm^2}{\zeta}\int\frac{e^{-\beta\phi'}q_k(r')d\tau'}{|\mathbf{r}-\mathbf{r}'|} - \frac{\lambda_k}{4\pi}\int\frac{e^{-\beta\phi'}q_k(r')d\tau'}{|\mathbf{r}-\mathbf{r}'|} = C_k\,, \quad (28)$$

where λ_k is an eigenvalue and C_k is a constant which can be chosen to allow us to satisfy one of the conditions on q. The integral equation (28) is equivalent to

$$\nabla^2 q_k + \left[\frac{4\pi\beta(N-1)Gm^2}{\zeta} + \lambda_k\right] e^{-\beta\phi} q_k = 0, \quad (29)$$

which is the differential equation examined by Horwitz and Katz[1], and applied by them to the Grand canonical ensemble using different boundary conditions. The equation Horwitz and Katz[2] use to determine the stability of the canonical ensemble is a different, more complicated equation; the equation we use is simpler because we work with $q = \delta\phi - \langle\delta\phi\rangle$.

We wish to expand q according to

$$q = \sum_{k=1}^{\infty} a_k q_k(r), \quad (30)$$

but since $\langle q\rangle = 0$ we require $\langle q_k\rangle = 0$. From Eq. (29) we find

$$\langle q_k\rangle = \frac{1}{\zeta}\int e^{-\beta\phi} q_k(r)\, d\tau = \frac{4\pi}{A_k}\left(\frac{dq_k}{dr}\right)_B r_B^2, \quad (31)$$

where r_B is the outer boundary of the sphere containing the particles and

$$A_k = 4\pi\beta(N-1)Gm^2 + \lambda_k\zeta. \quad (32)$$

We therefore require $(dq_k/dr)_B = 0$. In order for Eq. (28) to be consistent with this result we require $C_k = [d(rq_k)/dr]_B$, as may be easily seen by evaluating the integrals for $r = r_B$. It is convenient to work in terms of $X_k = rq_k$ so that the eigenvalue equation (29) becomes

$$\frac{d^2}{dr^2}X_k + \left[\frac{4\pi\beta(N-1)Gm^2}{\zeta} + \lambda_k\right] e^{-\beta\phi} X_k = 0, \quad (33)$$

with boundary conditions

$$q_k(0) \text{ finite } \Rightarrow X_k(0) = 0, \qquad \left(\frac{dq_k}{dr}\right)_B = 0 \Rightarrow r_B\left(\frac{dX_k}{dr}\right)_B - (X_k)_B = 0. \quad (34)$$

The boundary condition on q_k at $r = r_B$ has the further significance that since mass is conserved a spherically symmetric perturbation cannot change the external potential gradient. Therefore, at the boundary, the gradient of q must vanish and this is just the second of Eqs. (34).

Expanding q as in Eq. (30), substituting into Eq. (27), and using Eq. (28), we find Eq. (27) becomes

$$-\frac{1}{\zeta}\sum_k\sum_l a_k a_l \frac{\lambda_k}{4\pi}\iint \frac{e^{-\beta\phi} q_k q_l' e^{-\beta\phi'}}{|\mathbf{r}-\mathbf{r}'|}\, d\tau\, d\tau'; \quad (35)$$

the constant C_k gives no contribution because $\langle q\rangle = 0$. From Eq. (28) and the equivalent integral equation with k replaced by l we deduce

$$\iint \frac{e^{-\beta\phi-\beta\phi'} q_k q_l'}{|\mathbf{r}-\mathbf{r}'|}\, d\tau\, d\tau' = I_k\delta_{kl}, \quad (36)$$

which may easily be shown to be equivalent to the orthogonality condition deduced from Eq. (29). Using Eqs. (32) and (35) the expression for L in Sec. III becomes

$$\frac{L}{N} = -\frac{1}{4\pi\zeta}\sum\lambda_k I_k a_k^2. \quad (37)$$

Since I_k and a_k^2 are both positive, $L > 0$ only occurs when an eigenvalue becomes negative. Because of the form of Eq. (33) the eigenvalues are discrete and they can be ordered. The first to change sign is the lowest. Transforming to Emden variables we find λ_1 changes sign when the density contrast between center and boundary is 32.12. This is the now classical instability associated with a self-gravitating system in a heat bath. It is equivalent to C_v changing sign through infinity. The second eigenvalue changes sign at a density contrast of 5221.5 where C_v again changes sign through infinity and so on for the other eigenvalues.

V. THE EIGENVALUE PROBLEM FOR NONSPHERICAL VARIATIONS

When the variation is not spherically symmetric we expand in eigenfunctions

$$q_{\mathbf{k}} \equiv q_{klm} = (1/r) f_{kl}(r) Y_{lm}(\theta,\phi), \quad (38)$$

where Y_{lm} is a spherical harmonic. If $l \geqslant 1$, then the condition $\langle q_{klm}\rangle = 0$ is automatically satisfied. The eigenvalue equation (29) becomes

$$\frac{d^2}{dr^2}(f_{kl}) - \frac{l(l+1)}{r^2} f_{kl} + \left[\frac{4\pi\beta(N-1)Gm^2}{\zeta} + \lambda_{kl}\right] \times e^{-\beta\phi} f_{kl} = 0. \quad (39)$$

The boundary conditions are that $q_{\mathbf{k}}$ is bounded at $r = 0$, and matches a solution of Laplaces equation at $r = r_B$ which vanishes as $r \to \infty$, and has the same angular dependence as $q_{\mathbf{k}}$. These conditions are equivalent to

$$f_{kl}(0) \text{ and } \frac{df}{dr_B} = -\frac{l f_{kl}}{r_B} \text{ at } r = r_B, \quad l \geqslant 1. \quad (40)$$

If $l \geqslant 2$, it can be shown from the solution to the Emden equation that

$$\frac{l(l+1)}{r^2} > 4\pi\beta(N-1)Gm^2 e^{-\beta\phi}/\zeta. \quad (41)$$

Equation (39) can then be written (we omit the subscripts for convenience)

$$\frac{d^2 f}{dr^2} - Wf + \lambda P f = 0, \quad (42)$$

where W and P are > 0 for $l \geqslant 2$. Multiplying Eq. (42) by f, and integrating over r, shows that

$$\lambda = \left[\int_0^{r_B} W f^2\, dr + \int_0^{r_B}\left(\frac{df}{dr}\right)^2 dr - f(r_B)\frac{df}{dr_B}\right] \Big/ \int_0^{r_B} P f^2\, dr. \quad (43)$$

The boundary condition (40) now allows us to conclude from Eq. (43) that

$$\lambda_{kl} > 0 \text{ for } l \geqslant 2. \quad (44)$$

The integral equation from which Eq. (39) has been deduced is Eq. (28) with $C_k = 0$. Substitution into Eq. (27) shows that

$$\frac{L}{N} = -\frac{1}{\zeta}\sum_{k,l} a_{kl}^2 \lambda_{kl} I_{kl}\,, \tag{45}$$

where

$$I_{kl} = \int\int q_k(\mathbf{r}) q_k(\mathbf{r}') \frac{\exp(-\beta\phi)}{|\mathbf{r}-\mathbf{r}'|}\, d\tau\, d\tau'\,. \tag{46}$$

The system is therefore stable to any nonspherical perturbation with $l \geqslant 2$. This result agrees with that found by Horwitz and Katz[2]. For the case $l = 1$ numerical solution of the eigenvalue problem shows that the system is stable. This result may in fact be established directly by observing that, with an appropriate coefficient of proportionality, the boundary condition at $r = 0$ and Eq. (39) with $\lambda = 0$ are satisfied by

$$f_{11} \propto r\, d\phi/dr\,.$$

The outer boundary condition is never satisfied for a finite density ratio and we conclude that there is no solution with $\lambda = 0$. Since for density ratios close to 1 the system is stable, it remains stable. This result disagrees with that of Horwitz and Katz, because they use a boundary condition which we believe to be incorrect and find that the $l = 1$ mode is unstable. Since this mode is equivalent to the displacement of the center of mass, Horwitz and Katz remove the instability by fixing the center of mass. Our result shows that this device is unnecessary.

VI. RELATIONSHIP TO THE HELMHOLTZ FREE ENERGY

The canonical partition function is intimately related to the Helmholtz free energy because it represents a system of fixed volume in a heat bath. For such systems, if fluctuations about the temperature T_0 of the heat bath are considered, the stability of the system is determined by whether or not the activity

$$A = U - T_0 S\,, \tag{47}$$

is a minimum. In Eq. (47), U is the internal energy and S the entropy of the system of interest evaluated at the temperature T of the system. When $T = T_0$ the activity is just the Helmholtz free energy. The change of A with time for a viscous fluid with coefficient of thermal conductivity κ is

$$\frac{dA}{dt} = \frac{d}{dt}\left\{\frac{1}{2}\int P\left(\frac{1}{\gamma-1}\left(\frac{T'}{T}\right)^2 + \left(\frac{\rho'}{\rho}\right)^2\right) d\tau + \frac{1}{2}\int \rho u^2\, d\tau - \frac{G}{2}\int\int \frac{\rho'(\mathbf{r})\rho'(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, d\tau\, d\tau'\right\}, \tag{48}$$

where F' denotes the perturbation to the zero order quantity F, P is the pressure, ρ the density, G is the gravitational constant, γ the adiabatic index, $\mathbf{u}$ the velocity, and the integrations are over volume. We can therefore identify the quadratic terms in A produced by a perturbation away from equilibrium as being

$$\delta^2 A = \frac{1}{2}\int P\left[\frac{1}{\gamma-1}\left(\frac{T'}{T}\right)^2 + \left(\frac{\rho'}{\rho}\right)^2\right] d\tau - \frac{G}{2}\int\int \frac{\rho'(\mathbf{r})\rho'(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, d\tau\, d\tau' + \frac{1}{2}\int \rho u^2\, d\tau\,. \tag{49}$$

Assume now that we make a static perturbation with T constant. In Eq. (49) we set $T' = 0 = \mathbf{u}$. Noting that

$$\rho(\mathbf{r}) = M\frac{(N-1)}{N}\frac{e^{-\beta\phi}}{\int e^{-\beta\tau}\, dr}\,, \tag{50}$$

so that

$$\frac{\delta\rho'}{\rho} = -\beta(\delta\phi - \langle\delta\phi\rangle)\,, \tag{51}$$

we find

$$\delta^2 A = -L/2\,, \tag{52}$$

where L is defined by Eq. (26). The stability condition determined from the variational principle is therefore equivalent to requiring that the activity is a minimum. The statistical mechanical criteria is therefore identical to the thermodynamic criteria. If we construct the Lynden-Bell and Wood[4] activity for $T = T_0$ (so that $A \equiv$Helmholtz free energy) and perturb it, we find, as expected, that it is equal to $-L/2$.

VII. DISCUSSION AND CONCLUSIONS

A full account of the instabilities in a self-gravitating system placed in a heat bath is contained in the analysis of Lynden-Bell and Wood[4] and Horwitz and Katz.[2] Our results show that the description of the canonical ensemble from the statistical mechanical point of view does not require the complicated procedure used by Horwitz and Katz. Their analysis resembles ours in that it involves a search for a maximum of an integral. It differs from ours because they first make a transformation to a functional integral which greatly complicates the analysis of the canonical ensemble.

It is not clear why Horwitz and Katz do not use the boundary conditions (40) for the nonspherical perturbations. All potential perturbations are obliged to match a solution of the Laplace equation which vanishes at infinity. For the modes $l \geqslant 2$ the Horwitz–Katz boundary condition does not affect the result. However, for the $l = 1$ mode they predict an instability when the density ratio is 32.1, whereas we find the system is unconditionally stable for this perturbation.

The advantage of the procedure used by Horwitz and Katz is that it is general, and can therefore be used for all the common ensembles. Jensen's inequality can only be used when the probability density is convex and this excludes its use in the physically important microcanonical ensemble.

[1] G. Horwitz and J. Katz, Astrophys. J. **211**, 226 (1977).
[2] G. Horwitz and J. Katz, Astrophys. J. **222**, 941 (1978).
[3] J. Katz and G. Horwitz, Astrophys. J. Suppl. **33**, 251 (1977).
[4] D. Lynden-Bell and R. Wood, Mon. Not. R. Astron. Soc. **138**, 495 (1968).
[5] G.H. Hardy, J.E. Littlewood, and G. Polya, *Inequalities*, 2nd ed. (Cambridge U. P., Cambridge, England, 1959).
[6] J. Ipser, Astrophys. J. **193**, 463 (1974).
[7] J.J. Monaghan, Aust. J. Phys. **31**, 95 (1978).

A decomposition theorem for finite-dimensional Hermitian projections

M. J. Giannoni
Division de Physique Théorique, [a] *Institut de Physique Nucléaire 91406 Orsay Cedex, France*

(Received 13 April 1979; accepted for publication 2 October 1979)

We investigate conditions for the existence of a decomposition of a Hermitian projector ρ into two Hermitian and time-reversal invariant operators ρ_0 and χ under the form $\rho = e^{i\chi}\rho_0 e^{-i\chi}$. Sufficient conditions are given and an explicit construction of a decomposition is performed when they are fulfilled. A stronger theorem of existence and uniqueness is studied.

PACS numbers: 05.30.Ch, 02.30.Tb

INTRODUCTION

The purpose of this paper is to study the following problem: Given a Hermitian projector ρ defined on a Hilbert space of quantal states, under which conditions (necessary or sufficient) there do exist two Hermitian and time-reversal invariant operators ρ_0 and χ such that

$$\rho = e^{i\chi}\rho_0 e^{-i\chi}. \tag{1}$$

Such a decomposition was first introduced by Baranger and Véneroni[1] and in the formulation of these authors it constitutes the starting point of the "adiabatic time-dependent Hartree–Fock approximation" (ATDHF). In the framework of this formalism, the projector ρ is the reduced single-particle density operator of a system of independent spn-1/2 particles. In the adiabatic limit, the decomposition (1) seems to be crucial to provide a relation between the time-dependent Hartree–Fock approximation and phenomenological descriptions of collective motion such as the Copenhagen model.[2]

To our knowledge, the decomposition (1) has been used up to now only for single-particle density operators of nuclei.[3,4] The first studies of its validity[5,6] have also been restricted to this particular case. The proof of existence given here is valid for p-body reduced density matrices of a quantum system, provided they are projectors and they satisfy some conditions specified in Sec. 2. Conditions for the existence of the decomposition (1) are investigated in some details: Their study has been partly neglected in the preceding works. Indeed, they were focused on the ATDHF approximation, where the assumed smallness of χ makes the discussion much simpler.

For the sake of simplicity, the proofs given below are limited to a finite N-dimensional Hermitian space $\mathscr{H}$. However, most of the results can be extended to infinite-dimensional Hilbert spaces in the case where ρ has a finite trace. All the operators involved are linear operators defined on $\mathscr{H}$, except the antilinear time-reversal operator T.

Section 1 is devoted to the study of some results concerning quantum time-reversal which are used in Secs. 2 and 3. Section 2 is the central part of this work. There the decomposition theorem is demonstrated for the most general density projection operator satisfying some sufficient conditions. Finally, we investigate in Sec. 3 some special conditions for the existence of the decomposition (1) for the reduced single-particle density operator of a system of fermions.

1. SOME USEFUL PROPERTIES OF QUANTUM TIME-REVERSAL

In this preliminary section we present some theorems which are essential for the demonstration of the existence theorem of Sec. 2. These results can be easily obtained from elementary properties of the time-reversal operator T (see, e.g., Ref. 7).

In the following, the time-reversed of any linear operator A will be denoted by $A_T : A_T = T^+AT$, and the time-reversed of any vector $|u\rangle$ by $|\bar{u}\rangle : |\bar{u}\rangle = T|u\rangle$.

As is well known, the operator T is antiunitary (i.e., T is antilinear, and $T^+T = TT^+ = 1$) and satisfies:

$$T^2 = \pm 1.$$

In the latter equation, the plus sign applies to any system of bosons, or to systems containing an even number of fermions, the minus sign applies to an odd system of fermions.

The first results, stated in Theorems 1a and 1b, provide criteria to identify a time-even Hermitian operator by its spectral representation, respectively in the cases $T^2 = 1$ and $T^2 = -1$.

Theorem 1: (a) ($T^2 = 1$) A Hermitian operator is time-even iff it is diagonalizable in a real orthogonal basis.

(b) ($T^2 = -1$) A Hermitian operator A is time-even iff it is diagonalizable in an orthogonal basis of the type $\{|e_i\rangle, |e_i\rangle; i = 1,...,N/2\}$ with the same eigenvalue associated to $|e_i\rangle$ and $|\bar{e}_i\rangle$:

$$A = \sum_{i=1}^{N/2} \lambda_i (|e_i\rangle\langle e_i| + |\bar{e}_i\rangle\langle \bar{e}_i|), \quad \lambda_i \in \mathbb{R}. \tag{2}$$

Notice from Eq. (2) that the multiplicity of any eigenvalue of a time-even Hermitian operator A is an even number when T satisfies $T^2 = -1$. When A is the Hamiltonian of an odd system of fermions, this property is known as Kramers degeneracy. If A is a projector, Eq. (2) implies that its trace is an even number.

From Theorem 1, one gets readily the following properties of commuting Hermitian time-even operators:

Theorem 2: (a) ($T^2 = 1$) Two Hermitian time-even operators which commute have a common complete orthogonal set of real eigenvectors.

[a] Laboratoire associé au C.N.R.S.

 0022-2488/81/020352-05$1.00

(b) ($T^2 = -1$) Two Hermitian time-even operators A and B which commute can be simultaneously diagonalized in an orthogonal basis of the type $\{|e_i\rangle, |\bar{e}_i\rangle ; i = 1,...,N/2\}$, such that

$$A = \sum_{i=1}^{N/2} \lambda_i (|e_i\rangle \langle e_i| + |\bar{e}_i\rangle \langle \bar{e}_i|), \quad \lambda_i \in \mathbb{R},$$

$$B = \sum_{i=1}^{N/2} \mu_i (|e_i\rangle \langle e_i| + |\bar{e}_i\rangle \langle \bar{e}_i|), \quad \mu_i \in \mathbb{R}.$$

Theorems 2(a) and 2(b) permit us to obtain the following result, valid in both cases $T^2 = 1$ and $T^2 = -1$:

Theorem 3: ($T^2 = \pm 1$) Let U be a unitary operator.

(i) A necessary and sufficient condition for the existence of a time-even operator A satisfying

$$U = e^{iA} \tag{3}$$

is that

$$U_T = U^+ .$$

(ii) There exists a unique operator A satisfying Eq. (3) and having its eigenvalues λ_i lying in a given interval:

$$\lambda_i \in [\alpha, \alpha + 2\pi[.$$

2. THE DECOMPOSITION THEOREM

In the first part of this section we demonstrate the existence of two Hermitian and time-even operators ρ_0 and χ satisfying Eq. (1), provided some conditions are fulfilled. The proof of the existence theorem is performed by exhibiting a particular solution (ρ_0, χ), whose properties are studied in part B. Results of part B enable us to obtain, in Sec. (2 C), a new result, the decomposition theorem, which is stronger than the existence theorem proved in Sec. (2 A).

A. The existence theorem

Before stating the existence theorem, it is useful to mention a preliminary result which permits us to formulate the conditions for the existence of the decomposition (1) in several equivalent ways.

Preliminary result: Let ρ and ρ' be two Hermitian projections.

The following assumptions are *equivalent*:

(P_1) $\|\rho - \rho'\| < 1$.[8]

(P_2) The operator $(1 - R)$ is regular, where

$$R = (\rho - \rho')^2.$$

(P_3) The unitary operator

$$\tau\tau' = (2\rho - 1)(2\rho' - 1)$$

does not admit the eigenvalue $\nu = -1$.

(P_4) The projectors ρ and ρ' do not have any common eigenvector corresponding to different eigenvalues of ρ and ρ'.

The proof of these equivalences, being lengthy but rather straightforward, is left to the reader.

Existence theorem: Let ρ be a Hermitian projector and ρ_T its time-reversed. If ρ and ρ_T satisfy the equivalent hypotheses (P_i), there exist two operators ρ_0 and χ such that

$$\rho = e^{i\chi} \rho_0 e^{-i\chi}, \tag{1}$$

$$\rho_0 = \rho_0^\dagger = (\rho_0)_T, \tag{4}$$

$$\chi = \chi^\dagger = \chi_T . \tag{5}$$

Lemma: Let ρ be a Hermitian projector and ρ_T its time-reversed. If ρ and ρ_T satisfy the hypotheses (P_i), there exists a time-even Hermitian operator χ such that

$$\rho_T = e^{-2i\chi} \rho e^{2i\chi} . \tag{6}$$

Proof: The proof given here is suggested by results concerning pairs of projectors which can be found in Refs. 9, 10.

Consider the Hermitian and time-even operator:

$$1 - R = 1 - (\rho - \rho_T)^2 ,$$

which has the obvious property:

$$[1 - R, \rho] = [1 - R, \rho_T] = 0. \tag{7}$$

Since ρ and ρ_T satisfy (P_i), this operator is strictly positive, and one can define a Hermitian inverse square root $(1 - R)^{-1/2}$.

Let us define:

$$U = [\rho\rho_T + (1 - \rho)(1 - \rho_T)](1 - R)^{-1/2}$$
$$= (1 - R)^{-1/2}[\rho\rho_T + (1 - \rho)(1 - \rho_T)] . \tag{8}$$

One can easily show that

(i) U is unitary,
(ii) $U_T = U^\dagger$,
(iii) $\rho_T = U^\dagger \rho U$. (9)

According to Theorem 3, one can find an operator χ fulfilling Eqs. (5) and (6), and such that

$$U = e^{2i\chi} , \tag{10}$$

which proves the Lemma.

To achieve the proof of the theorem, it remains to verify that the Hermitian projector ρ_0 defined by

$$\rho_0 = e^{-i\chi} \rho e^{i\chi} \tag{11}$$

is time-even, which is a direct consequence of the time-reversal invariance of χ.

Before proceeding further, it is worth stressing the following point: We have shown the existence of a time-even Hermitian operator χ satisfying (6) by referring to Theorem 3. However, such an operator χ is not yet univocally defined, since: (i) the operator $(1 - R)^{-1/2}$ used in the construction of U can be defined in several different ways; (ii) it remains to specify the interval of definition of the eigenvalues of χ. In order to remove all these ambiguities, we first note that the operator U used to define χ by Eq. (10) has the property

$$U^2 = \tau\tau_T = e^{4i\chi} , \tag{12}$$

with $\tau = (2\rho - 1)$. Equation (12) can be obtained from the definition (8) of U, by use of Eq. (7) and of the involutive character of τ and τ_T. According to Theorem 3, there is a *unique* operator χ satisfying (12) and having all its eigenvalues in some given interval $[\alpha, \alpha + \pi/2[$. We define χ by

choosing $\alpha = -\pi/4$. Since the condition (P_3) excludes the eigenvalue $\nu_u = -1$ for the operator $e^{4i\chi}$, the eigenvalues λ_u of χ all belong to the *open* interval:

$$\lambda_u \in]-\pi//4, \quad \pi/4[\,. \tag{13}$$

Incidentally, we emphasize that the operator χ just defined is identical to that defined by Baranger and Véneroni.[1] This was not apparent up to now, but is clearly illustrated by Eqs. (12) and (13). Following the terminology of Ref. 1, we will call the decomposition defined by such an operator χ, and the operator ρ_0 constructed from χ by Eq. (11), the "natural" decomposition.[11]

Notice that in the above definition of χ, the condition (P_3) is dissimulated in Eq. (13), whereas the equivalent hypothesis (P_2) was needed at the very first to define U through Eq. (8). These conditions (P_i) were not mentioned in Ref. 1, whose authors were interested in the decomposition theorem for operators χ small compared to unity ($\lambda_u \ll 1$): this assumption guarantees the fulfilment of (P_i), and therefore ensures the existence of the decomposition (1) submitted to (4) and (5).

B. Further properties of the natural decomposition

Let us now investigate characteristic properties of the natural decomposition constructed in Sec. (2 A).

From the definition (12), (13) of χ, it is clear that for any eigenvector $|u\rangle$ of the diagonalizable operator $\tau\tau_T$ ($\tau\tau_T$ is unitary):

$$\tau\tau_T|u\rangle = \nu_u|u\rangle\,, \tag{14}$$

one has:

$$\chi|u\rangle = \lambda_u|u\rangle\,, \tag{15}$$

with λ_u defined by Eq. (13) and by

$$e^{4i\lambda_u} = \nu_u\,. \tag{16}$$

One can also easily show that

$$\tau\tau_T(\tau|u\rangle) = \nu_u^*(\tau|u\rangle) \tag{17}$$

for any $|u\rangle$ solution of Eq. (14), which leads to

$$\chi(\tau|u\rangle) = -\lambda_u(\tau|u\rangle). \tag{18}$$

As a consequence of Eqs. (15) and (18), the operator χ of the natural decomposition satisfies

$$\chi\tau + \tau\chi = 0, \tag{19}$$

which can be written equivalently as

$$\chi\rho + \rho\chi = \chi\rho_0 + \rho_0\chi = \chi \tag{20}$$

or as

$$\rho\chi\rho = (1-\rho)\chi(1-\rho) = 0. \tag{21}$$

If ρ is the one-body reduced density operator of a spin-1/2 particle system, Eq. (21) means that the operator χ has only particle-hole and hole-particle matrix elements.

To further study the natural decomposition, one can state the following lemma (whose proof is straightforward):

Lemma: Let χ be a Hermitian operator such that

$$\tau_T = e^{-2i\chi}\tau\, e^{2i\chi}\,,$$

$$\chi\tau + \tau\chi = 0.$$

Then χ satisfies $e^{4i\chi} = \tau\tau_T$. As an immediate consequence, the natural decomposition is the only one fulfilling (13) and having the property (19) (see Theorem 3).

C. Decomposition theorem

The main results established up to now can be summarized as follows:

(a) Assuming the conditions (P_i) satisfied by ρ and ρ_T, we have constructed a Hermitian operator χ solution of Eq. (6). This operator χ is shown to be time-even.

(b) We have seen that Eq. (19) [added to the condition (13)] is a characteristic property of this particular solution.

(c) The Hermitian operator ρ_0, constructed from χ and ρ by Eq. (11), is time-even.

These results can be collected into the following statement.

Decomposition theorem: Let ρ be a Hermitian projection, and T the time-reversal antiunitary operator satisfying $T^2 = \pm 1$.

The equivalent assumptions (P_i) for ρ and ρ_T are necessary and sufficient conditions for the existence of a unqiue set of Hermitian operators ρ_0 and χ such that

(i) $\chi\rho_0 + \rho_0\chi = \chi$, (20)

(ii) all the eigenvalues of χ lie within the interval $]-\pi/4, \ \pi/4[$, (13)

(iii) $\rho = e^{i\chi}\rho_0\, e^{-i\chi}$. (1)

The two operators ρ_0 and χ are time-even.

N.B.: (1) By rearranging the results (a)–(c) in different ways, one can obtain several other formulations (not completely equivalent) of the decomposition theorem (see, e.g., Ref. 12).

(2) In the statement given above, the conditions (P_i) appear as necessary. This is obvious, since the violation of (P_i) would lead to the occurrence of at least one eigenvalue $\lambda_u = \pm\pi/4$ for χ, which would be in contradiction with the requirement (ii). We would like to point out that a stronger result is available, by replacing the open interval in (ii) by a semiopen interval. The proof would go as follows: One shows that Eqs. (12) and (20) cannot be simultaneously fulfilled by an operator χ having an eigenvalue $\lambda_u = \pm\pi/4$.

3. SOME IMPLICATIONS FOR SYSTEMS OF SPIN-1/2 PARTICLES

Let us turn back to the *existence theorem*, studied in Sec. (2 A). This theorem has been proved under some *sufficient* conditions (P_i), expressed up to now in a mathematical language. By looking at systems of spin-1/2 particles, we will now investigate *necessary* conditions of existence, and also get some light about their physical content.

From now on ρ will be a one-body reduced density operator describing a system of spin-1/2 particles. If the decomposition (1) exists, the trace of the time-even Hermitian projection ρ_0 is an even number (see Sec. 1). Since ρ is deduced from ρ_0 by a unitary transformation, we conclude that

$$\mathrm{Tr}\rho = 2p, \quad p \in \mathbb{N}. \tag{22}$$

As a *necessary* condition for existence, Eq. (22) implies that no decomposition (1) fulfilling Eqs. (4) and (5) can be found for the reduced one-body density operator of an odd system of fermions. Property (22), which did not appear in the proof given in Sec. (2 A), has just been shown "a posteriori" to be a consequence of the sufficient conditions (P_i). We directly show in the Appendix that (P_i) implies (22).

We now consider the case where the single-particle states are normalized eigenstates $|\pm\rangle$ of the spin operator S_z:

$$\rho = \rho^u \otimes |+\rangle\langle+| + \rho^d \otimes |-\rangle\langle-|. \tag{23}$$

It might be of interest to ask under which conditions on the spatial parts ρ^u and ρ^d of ρ there exists a decomposition (1) satisfying (4) and (5), such that the spin states $|\pm\rangle$ are eigenvectors of χ. Added to the requirements for χ to be Hermitian and time-even, this condition constrains χ to be of the form:

$$\chi = \chi^u \otimes |+\rangle\langle+| + \chi^u_T \otimes |-\rangle\langle-|, \tag{24}$$

with

$$\chi^u = (\chi^u)^*. \tag{25}$$

We will now show that the property

$$n^u = n^d, \tag{26}$$

with $n^u = \mathrm{Tr}\rho^u$ and $n^d = \mathrm{Tr}\rho^d$, is a *necessary* condition[13] for the existence of a decomposition (1) satisfying the requirements (4), (5), (23), and (24). This is easily seen by computing the operator $e^{2i\chi}\rho_T e^{-2i\chi}$, which must be equal to ρ when a decomposition (6) submitted to (24) and (25) exists. One gets

$$\rho^u = \exp(2i\chi^u)\rho^d_T \exp(-2i\chi^u).$$

Since χ^u is Hermitian, this implies

$$\mathrm{Tr}\rho^u = \mathrm{Tr}\rho^d_T = \mathrm{Tr}\rho^d,$$

which shows (26).

To end up this section, let us see how the condition (26) is contained in the assumptions (P_i) for ρ and ρ_T. We first note that the conditions (P_i) are satisfied by ρ and ρ_T if, and only if they are satisfied by ρ^u and ρ^d_T. (The proof of this result is elementary.) In particular, the fulfilment of (P_i) by ρ and ρ_T requires that

$$\|\rho^u - \rho^d_T\| < 1. \tag{27}$$

As shown by Sz.-Nagy,[9] this property implies (26) (see also Ref. 14). We will not demonstrate this well known result, but simply mention that it can be derived by exactly the same procedure as used in the proof of the existence theorem: Equation (27) allows to define

$$V = [\rho^u\rho^d_T + (1-\rho^u)(1-\rho^d_T)][1-(\rho^u-\rho^d_T)^2]^{-1/2}.$$

The unitary operator V is easily shown to transform ρ^u into ρ^d_T:

$$\rho^d_T = V^*\rho^u V,$$

which leads to (26).

SUMMARY AND COMMENTS

The essential results obtained in the present work are the existence and the decomposition theorems given in Sec. 2.

The first theorem states the existence of a decomposition fulfilling Eqs. (1), (4), and (5) for any projection operator ρ which, together with its time-reversed ρ_T, satisfies the conditions (P_i). Since this result has been demonstrated by the explicit construction of a particular solution (the so-called natural decomposition), the conditions (P_i) appear as sufficient, but not *a priori* necessary. Investigation of necessary conditions for the existence has been made in Sec. 3 for the reduced one-body density operator of a system of spin-1/2 particles. We did not discuss the possible uniqueness of (1) submitted to (4), (5), and to the condition (13) for the eigenvalues of χ. Actually, it can be shown in some specific physical situations that such uniqueness is not true. A counterexample is given by the density operator obtained from a solution of the static Hartree–Fock equation by a Galilean transformation.[1,5]

In order to get an existence-plus-uniqueness theorem, we have investigated characteristic properties of the natural decomposition. These properties are the reduction (13) of the interval of definition for the eigenvalues of χ and the relation (20). Added as supplementary conditions to Eqs. (4) and (5), they ensure uniqueness and lead to the decomposition theorem.

To end up this summary of our mathematical study, we point out that the choice of a finite N-dimensional space, made here for the sake of simplicity, is not a real restriction, at least as far as the existence theorem for projection operators of finite trace is concerned. Indeed, as shown in the Appendix, the significant space to consider in this case is the linear sum $(\mathscr{R} + \mathscr{R}_T)$, which is a Hermitian finite space.

Coming back to physics, one can ask about the consequences of the nonuniqueness of (1) submitted to the conditions (4), (5), and (13). Does it induce ambiguities in the physical results? An answer to this question is known when the decomposition (1) is used in the framework of the ATDHF approximation. Indeed, it has been shown in Ref. 1 that all the possible sets of operators (ρ_0,χ) fulfilling (1), (4), and (5) lead to equivalent dynamics; thus condition (20) does not appear *a posteriori* as a physical limitation, and can be legitimately imposed in practical calculations, as done in Refs. 5, 15.

The proofs given above are not restricted to reduced one-body densities, although up to now all the applications concern the time-dependent Hartree–Fock approximation. It might be of some interest that these proofs apply to any finite-dimensional projections satisfying conditions (P_i), and therefore to a larger class of n-body operators.

ACKNOWLEDGMENT

I am indebted to M. Vénéroni for numerous stimulating discussions and constant interest in this work.

APPENDIX

We restrict ourselves to a space $\mathscr{H}$ generated by the p-body states of a system of fermions and such that the time-

reversal operator has the property $T^2 = -1$ (i.e., p is an odd number). If $\mathscr{H}$ is a single-particle space, what follows concerns the reduced single-particle density operator used for instance in the Hartree–Fock approximation.

We will show directly (another proof is given in Sec. 3) that the fulfilment of conditions (P_i) for ρ and its time-reversed ρ_T requires:

$$\mathrm{Tr}\rho = 2p, \quad p \in \mathbb{N} .$$

Let $\mathscr{R}$ and $\mathscr{R}_T$ be the subspaces on to which ρ and ρ_T project:

$$\mathscr{R} = \rho(\mathscr{H}), \quad \mathscr{R}_T = \rho_T(\mathscr{H}).$$

From its definition, it is clear that the operator

$$\tau\tau_T - 1 = 4\rho\rho_T - 2(\rho + \rho_T)$$

transforms each vector of $\mathscr{H}$ into a vector of the linear sum $(\mathscr{R} + \mathscr{R}_T)$:

$$(\tau\tau_T - 1)|u\rangle = |u_1\rangle + |u_2\rangle ,$$

with $|u_1\rangle \in \mathscr{R}$, $|u_2\rangle \in \mathscr{R}_T$. Hence, $(\mathscr{R} + \mathscr{R}_T)$ is an invariant subspace of $(\tau\tau_T - 1)$, and we can define the restriction Ω of the operator $(\tau\tau_T - 1)$ to the subspace $(\mathscr{R} + \mathscr{R}_T)$.

To study the spectrum of Ω, we note that if

$$\Omega|u\rangle = (e^{4i\lambda_u} - 1)|u\rangle ,$$

then

$$\Omega(\tau|u\rangle) = (e^{-4i\lambda_u} - 1)\tau|u\rangle ,$$

$$\Omega|\bar{u}\rangle = (e^{4i\lambda_u} - 1)|\bar{u}\rangle ,$$

and

$$\Omega(\tau_T|\bar{u}\rangle) = (e^{-4i\lambda_u} - 1)\tau_T|\bar{u}\rangle .$$

All the vectors $|u\rangle$, $\tau|u\rangle$, $|\bar{u}\rangle$, $\tau_T|\bar{u}\rangle$ belong clearly to $(\mathscr{R} + \mathscr{R}_T)$. The dimensionality of the subspace $\mathscr{S}_u$ generated by these four vectors (all generated from a given $|u\rangle$) can be different in the case where $e^{4i\lambda_u}$ is real and in the opposite case.

(i) If $e^{4i\lambda_u} \neq \pm 1$, these four vectors span a four-dimensional space $\mathscr{S}_u$, since the two sets $(|u\rangle, |\bar{u}\rangle)$ and $(\tau|u\rangle, \tau_T|\bar{u}\rangle)$ correspond to different eigenvalues of Ω, and the two vectors of each set are linearly independent [see Refs. 7].

(ii) The case $e^{4i\lambda_u} = -1$ cannot occur if the condition (P_3) is required.

(iii) If $e^{4i\lambda_u} = 1$, the four eigenvectors are associated with the same eigenvalue zero of Ω. To study the dimensionality of $\mathscr{S}_u$, we shall use the following property.

Lemma: Let ρ and ρ' be two Hermitian projectors. Each vector $|u\rangle$ satisfying

$$\rho|u\rangle = \rho'|u\rangle$$

is a common eigenvector of ρ and ρ':

$$\rho|u\rangle = \rho'|u\rangle = \mu|u\rangle .$$

Let us consider the eigenvector $|u\rangle$ of Ω associated with the eigenvalue $(e^{4i\lambda_u} - 1) = 0$. It is easily seen that such vectors satisfy

$$\rho|u\rangle = \rho_T|u\rangle ,$$

and the Lemma ensures that

$$\rho|u\rangle = \rho_T|u\rangle = |u\rangle . \tag{A1}$$

Therefore, the subspace $\mathscr{S}_u$ in the case $e^{4i\lambda_u} = 1$ is just:

$$\mathscr{S}_u = \mathscr{R} \cap \mathscr{R}_T .$$

This space is of dimensionality 2 or 4, since Eq. (A1) implies

$$\rho_T|\bar{u}\rangle = \rho|\bar{u}\rangle = |\bar{u}\rangle ,$$

with $\langle u|\bar{u}\rangle = 0$.

To summarize, we have obtained the results

$$\dim(\mathscr{R} \cap \mathscr{R}_T) = 2n, \quad n \in \mathbb{N} , \tag{A2}$$

$$\dim(\mathscr{R} + \mathscr{R}_T) = 2n + 4k, \quad k \in \mathbb{N} . \tag{A3}$$

As is well known,

$$\dim(\mathscr{R} + \mathscr{R}_T) + \dim(\mathscr{R} \cap \mathscr{R}_T) = \dim\mathscr{R} + \dim\mathscr{R}_T .$$

Since

$$\dim\mathscr{R} = \dim\mathscr{R}_T , \tag{A4}$$

one gets from (A2)–(A4) the final result

$$\mathrm{Tr}\rho = \dim\mathscr{R} = 2(k + n) = 2p, \quad p \in \mathbb{N} .$$

[1] M. Baranger and M. Véneroni, Ann. Phys. **114**, 123 (1978).
[2] A. Bohr and B.R. Mottelson, *Nuclear Structure*, (Benjamin, New York, 1975) Vol. 2.
[3] D.M. Brink, M.J. Giannoni, and M. Véneroni, Nucl. Phys. A **258**, 237 (1976); M.J. Giannoni, Preprint IPNO/TH 79-65, Orsay, December 1979.
[4] M.J. Giannoni, F. Moreau, P. Quentin, D. Vautherin, M. Véneroni, and D.M. Brink, Phys. Lett. B **65**, 305 (1976); M.J. Giannoni, and P. Quentin, Phys. Rev. C (to be published).
[5] P. Bonche and P. Quentin, Phys. Rev. C **18**, 1891 (1978).
[6] P. Ring and P. Schuck, Nucl. Phys. A **292**, 20 (1977).
[7] E.P. Wigner, J. Math. Phys. **1**, 409 (1960); E.P. Wigner, *Group Theory* (Academic, New York and London, 1959); A. Messiah, *Mécanique quantique* (Dunod, Paris, 1964), Vol. 2.
[8] The norm considered is defined by $\|A\| = \sup_{|u\rangle \in \mathscr{H}} \|A|u\rangle\| / \||u\rangle\|$.
[9] B. Sz.-Nagy, Comments Math. Helv. **19**, 347 (1947).
[10] T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966), pp. 32–34.
[11] Indeed in the adiabatic limit considered in Ref. 1, this decomposition is most convenient because it allows an explicit reduction of the number of dynamical variables to the number of degrees of freedom; moreover in the Hamiltonian formulation of the ATDHF formalism, the matrix elements of the natural ρ_0 and χ play respectively the roles of coordinates and momenta.[3] However, outside the adiabatic limit, the choice of the natural decomposition is no more dictated by physical reasons: For uniform translation, it happens that the Galilean invariance suggests another choice, which is more appropriate to this specific case.
[12] M.J. Giannoni, in *Time-Dependent Hartree–Fock Method*, edited by P. Bonche, B. Giraud, and P. Quentin (Editions de Physique, Orsay, 1979), p. 251.
[13] Authors of Ref. 5 focused their attention on the natural decomposition, fulfilling the condition (20). They showed that Eq. (26) is a necessary condition for the existence of this particular decomposition. One can easily see that Eqs. (20) and (5) imply (24) and (25), so that the property demonstrated here is stronger than the statement of Ref. 5.
[14] N.I. Akhiezer and I.M. Glazman, *Theory of Linear Operators in Hilbert Space* (Ungar, New York, 1961), Vol. 1, p. 70.
[15] P. Bonche, H. Doubre, and P. Quentin, Phys. Lett. B **82**, 5 (1979).

Massive spin-1 field in expanding universes and the quantum equivalence principle

Adriana Foussats, Roberto Laura, and Oscar Zandron[a)]
Departamento de Física, Facultad de Ciencias Exactas e Ingeniería, U.N.R., Av. Pellegrini 250, Rosario, Argentina

(Received 24 March 1980, accepted for publication 31 July 1980)

In this work, we develop the quantum field theory formalism in the curved space-time for the case of massive vector field, using the Quantum Equivalence Principle previously introduced. With this principle and for a particular model for expanding universe—spatially flat Robertson–Walker metric—an adequate particle model is obtained. The mean density of created particles for the corresponding Bogolyubov transformation is finite.

PACS numbers: 11.10. − z, 04.60. + n

1. INTRODUCTION

The Quantum Equivalence Principle (QEP) has been introduced in previous works (cf. Refs. 1 and 2). In the case of a scalar material field the QEP enabled us to find an adequate particle model in the curved space-time. With this model we obtained a Bogolyubov transformation, which led to the creation of a finite number of particles. Dealing in this way with the problem of generalizing the flat space-time field theory into the curved space-time field theory, essencially consists in treating the material field as a quantized one, and to introduce through the metric used, the gravitational field as an unquantized classical external one.

Lichnerowicz developed the mathematical formalism of the fields' theory in a curved space-time for the fundamental fields (cf. Refs. 3,4, and 5). Afterwards the problem of a real scalar field has been treated in various works.[6–10] The existence of ambiguities derived from the developing of a particle model in the curved space-time followed. These ambiguities lead to a possible existence of a mechanism of creating particles which is called the Bogolyubov transformation. Unfortunatly the Bogolyubov transformation for certain particle models produces an infinite number of particles.

As was shown in Refs. 1 and 2 for a real scalar field, one of the fundamental problems is that the so-called Lichnerowicz conditions do not determine in a unique way the biscalar kernel $G_1(x,x')$ of the curved space-time that is the generalization of the kernel $\Delta_1(x-x')$ of the flat space-time. One way to solve this problem is to suppose the existence of a different decomposition for the solutions of positive frequencies (and negative ones) for each normal Cauchy hypersurface Σ of the curved space-time. This curved space-time is supposed globaly hyperbolic and endowed with normal Cauchy hypersurface (cf. Ref. 2). This fact led us to suppose that there is a different kernel $G_1^{(\Sigma)}(x,x')$ for each normal Cauchy hypersurface Σ. In each normal Cauchy hypersurface Σ a proper kernel $G_1^{(\Sigma)}(x,x')$ will be determined, when their Cauchy data on each hypersurface Σ are given.

The QEP, which enabled us to solve this problem satisfactorily, essencially consists in choosing the Cauchy data in such a way so as to make the kernel $G_1^{(\Sigma)}(x,x')$ of the curved space-time as similar as possible to the kernel $\Delta_1(x-x')$ of the flat space-time, on the normal Cauchy Hypersurface Σ.

In the present work we are going to use this proposition for a real massive vector field with the spatially flat Robertson–Walker metric.

In Sec. 2, we are going to develop briefly, the Lagrangian formalism for a real massive vector field in the curved space-time.

In Sec. 3, we deal with the Cauchy problem for a vector field, and we show, generalizing a Lichnerowicz theorem (cf. Ref. 3), the existence of a bitensorial kernel $X^{\mu\nu'}(x,x')$ (1-tensor in x, 1-tensor in x') that in a unique way solves the Cauchy problem.

In Sec. 4, we study, for the particle model proposed in the curved space-time, the decomposition into positive and negative frequencies.

In Sec. 5, we outline the Bogolyubov transformation for the vector field.

In Sec. 6, we find the solution of the field equations for the particular case of the spatially flat Robertson–Walker metric.

In Sec. 7, making use of the QEP, we find an adequate bitensorial kernel $X_1^{(\Sigma)\mu\nu'}(x,x')$ for each normal Cauchy hypersurface Σ.

In Sec. 8, we find the initial conditions for the solutions of positive and negative frequencies which result from requiring the diagonalization of the Hamiltonian operator.

Finally, in Sec. 9 it is shown that the initial conditions for the solutions of positive and negative frequencies, resulting from the QEP, are those which originate a Bogolyubov transformation for which the density of created particles is finite. The initial condition obtained from the diagonalization of the Hamiltonian operator led to an infinite density of created particles; consequently they cannot be used.

2. LAGRANGIAN FORMALISM

With the aim of constructing a model of massive spin-1 particles in a curved space-time, we start from the following action integral[11]:

[a)]Fellow of Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina.

 0022-2488/81/020357-14$1.00

$$S=\int\sqrt{|g|}\{-\tfrac{1}{4}g^{\mu\rho}g^{\nu\sigma}(\varphi_{\nu;\mu}-\varphi_{\mu;\nu})(\varphi_{\sigma;\rho}-\varphi_{\rho;\sigma})$$
$$+\tfrac{1}{2}m^2g^{\mu\nu}\varphi_\mu\varphi_\nu\}\,d^4x. \tag{2.1}$$

The variation of the action integral (2.1) leads us to the following equation for the field φ_μ:

$$-g^{\alpha\rho}\varphi^{\ \beta}_{;\alpha;\rho}+R^{\ \beta}_{\alpha}\varphi^{\ \alpha}+\varphi^{\ \alpha\ \beta}_{;\alpha}-m^2\varphi^{\ \beta}=0. \tag{2.2}$$

Equation (2.2) involves the following two equations:

$$(\Delta-m^2)\varphi^{\ \beta}=0, \tag{2.3}$$

$$\varphi^{\ \alpha}_{;\alpha}=0, \tag{2.4}$$

where $\Delta\varphi^\beta=-g^{\alpha\rho}\varphi^{\ \beta}_{;\alpha;\rho}+R^{\ \beta}_{\alpha}\varphi^{\ \alpha}$ is the Rham Laplacian (cf. Ref. 3) for a vector field.

The energy-momentum tensor of the massive vector field, as obtained from (2.1) is

$$T_{\alpha\beta}=\frac{2}{\sqrt{|g|}}\frac{\delta S}{\delta g^{\alpha\beta}}$$
$$=\tfrac{1}{4}[g_{\alpha\beta}g^{\mu\rho}g^{\nu\sigma}-g^{\nu\sigma}(\delta^\mu_\alpha\delta^\rho_\beta+\delta^\mu_\beta\delta^\rho_\alpha)$$
$$-g^{\mu\rho}(\delta^\nu_\alpha\delta^\sigma_\beta+\delta^\nu_\beta\delta^\sigma_\alpha)]f_{\mu\nu}f_{\rho\sigma}$$
$$-\tfrac{1}{2}m^2g_{\alpha\beta}g^{\mu\nu}\varphi_\mu\varphi_\nu+\tfrac{1}{2}m^2(\varphi_\alpha\varphi_\beta+\varphi_\beta\varphi_\alpha), \tag{2.5}$$

where $f_{\mu\nu}=\partial_\mu\varphi_\nu-\partial_\nu\varphi_\mu$.

The energy-momentum tensor (2.5) satisfies the following two conditions:

$$T_{\alpha\beta}=T_{\beta\alpha}, \tag{2.6}$$

$$T^{\alpha}_{\ \beta;\alpha}=0. \tag{2.7}$$

From (2.5), and for the particular case of the Robertson–Walker metric, which is

$$ds^2=dt^2-a^2(t)[dx^2+dy^2+dz^2], \tag{2.8}$$

and which is going to be studied later on; the following expressions are obtained for the components T_{00} and T_{0i} of the energy-momentum tensor:

$$T_{00}=(1/2a^2)\sum_j f_{0j}f_{0j}+(1/2a^4)\sum_{ij}f_{ij}\partial_i\varphi_j$$
$$+(1/2a^2)m^2\sum_i\varphi^{\ 2}_i+\tfrac{1}{2}m^2\varphi^{\ 2}_0, \tag{2.9}$$

$$T_{0i}=(1/2a^2)\sum_j f_{0j}\partial_i\varphi_j. \tag{2.10}$$

Equation (2.7) for $\beta=i$ and for the metric (2.8) is

$$T^{\alpha}_{i;\alpha}=\left(1/\sqrt{|g|}\right)\frac{\partial}{\partial x^\alpha}\left(\sqrt{|g|}\,T^{\alpha}_{i}\right)=0. \tag{2.11}$$

The integration of Eqs. (2.11) in the four-dimensional space leads to the conservation of the quantity

$$P_i=a^3(t)\int_{t=\text{const.}}T^0_{\ i}\,d^3x. \tag{2.12}$$

To obtain the spin operator we use the infinitesimal transformation of rotation. The total variation of action (2.1) is

$$\delta S=\int_R d^4x\,\delta\left(\mathcal{L}\sqrt{|g|}\right)+\int_R\sqrt{|g|}\,\mathcal{L}\,\delta(d^4x), \tag{2.13}$$

where R is a finite region of V_4.

We make the following infinitesimal transformations of coordinates

$$x'^\mu=x^\mu+\delta x^\mu=x^\mu+\xi^\mu(x). \tag{2.14}$$

If the transformation is such that $\bar\delta g_{\alpha\beta}=-(\xi_{\beta;\alpha}+\xi_{\alpha;\beta})=0$, where $\bar\delta g_{\alpha\beta}$ indicates only the variation in the form of the function $g_{\alpha\beta}$, Eq. (2.13) results:

$$\delta S=\int_R\frac{\partial}{\partial x^\nu}\left[\frac{\partial(\mathcal{L}\sqrt{|g|})}{\partial(\partial\varphi_\mu/\partial x^\nu)}\bar\delta\varphi_\mu+\mathcal{L}\sqrt{|g|}\,\xi^\nu\right]d^4x, \tag{2.15}$$

where $\bar\delta\varphi_\mu$ indicates only the variation in the form of the function φ_μ. Let us consider an infinitesimal transformation of rotation, with $\xi^\mu=w^\mu_{\ \nu}x^\nu$, where $w^\mu_{\ \nu}$ are the infinitesimal constant parameters of rotation and, moreover, $w^{\mu\nu}=-w^{\nu\mu}$ ($w^{\mu\nu}$ not constant). Using (2.15), it is possible to define, as is usual, the zero component of the spin angular momentum tensor density $\mathcal{S}^{\nu}_{[\alpha,\beta]}$ in the following way:

$$\mathcal{S}^{0\ \tau}_{\alpha}=-\frac{\partial(\mathcal{L}\sqrt{|g|})}{\partial(\partial\varphi_\mu/\partial x^0)}(g_{\mu\alpha}g^{\beta\tau}\varphi_\beta-\delta^\tau_\mu\varphi_\alpha). \tag{2.16}$$

In particular we are interested in the component $\mathcal{S}^{0\ 2}_{1}$ which, for the metric (2.8), results in

$$\mathcal{S}^{0\ 2}_{1}=a^3g^{\mu\sigma}f_{0\sigma}(g_{\mu1}g^{\beta2}\varphi_\beta-\delta^2_\mu\varphi_1). \tag{2.17}$$

Therefore the spatial density of the third component of the spin vector **S** will be:

$$s_3=\int\mathcal{S}^{0\ 2}_{1}d^3x=a\int d^3x\,(f_{02}\varphi_1-f_{01}\varphi_2). \tag{2.18}$$

3. THE CAUCHY PROBLEM FOR MASSIVE VECTOR FIELDS

We are now going to obtain the solution for the equation

$$(\Delta-m^2)\varphi_\nu(x)=0 \tag{3.1}$$

that satisfies the following boundary conditions on an spatial hypersurface Σ:

$$\varphi_\nu(y)=\varphi^{(\Sigma)}_\nu,\quad n^\rho\nabla_\rho\varphi_\nu(y)=\dot\varphi^{(\Sigma)}_\nu, \tag{3.2}$$

where $y\in\Sigma$ and n^ρ is the unitary normal vector in each point of Σ (Cauchy problem). We are also going to show now, generalizing a Lichnerowicz theorem (cf. Ref. 3), the existence of a bitensorial distribution $G_{\mu\nu'}(x,x')$ (1-tensor in x, 1-tensor in x') that will enable us to solve the Cauchy problem. To do so, we write the solution in the following way:

$$\varphi_\mu(x)=\int_\Sigma[G_{\mu\nu'}(x,x')\nabla_{\rho'}\varphi^{\ \nu'}_{(x')}$$
$$-\varphi^{\ \nu'}_{(x')}\nabla_{\rho'}G_{\mu\nu'}(x,x')]n^{\rho'}d\sigma', \tag{3.3}$$

with $x'\in\Sigma$.

In Ref. 3, it is shown that if U and V are two vector fields, they satisfy

$$[(\Delta-m^2)V]_\mu U^\mu-V_\mu[(\Delta-m^2)U]^\mu$$
$$=-\nabla^\rho[U^\alpha\nabla_\rho V_\alpha-V^\alpha\nabla_\rho U_\alpha]. \tag{3.4}$$

It is known that there are two elementary kernels $E^{+}_{\mu\nu}(x,x')$ and $E^{-}_{\mu\nu}(x,x')$, whose support are respectively in the future and in the past of x'. These kernels are determined, in a unique way, by the following properties:

$$E^{+}_{\mu\nu}(x,x') = E^{-}_{\mu\nu}(x',x),$$
$$(\Delta_x - m^2)E^{+}_{\mu\nu}(x,x') = \delta_{\mu\nu}(x,x'), \tag{3.5}$$

where $\delta_{\mu\nu}(x,x') = \mathscr{T}_{\mu\nu}(x,x')\delta(x,x')$ and $\mathscr{T}_{\mu\nu}(x, x' = x) = g_{\mu\nu}(x)$.

If we define

$$\{A^{-}(\varphi)\}_{\rho\nu} = E^{-}_{\lambda\nu}(x,x')\nabla_\rho \varphi^{\lambda}_{(x)} - \varphi^{\lambda}_{(x)}\nabla_\rho E^{-}_{\lambda\nu}(x,x'), \tag{3.6}$$

and in (3.4) we let $V^\lambda \to \varphi^\lambda$ and $U_\lambda \to E^{-}_{\lambda\nu}$ it follows that

$$\begin{aligned} -\nabla^\rho\{A^{-}(\varphi)\}_{\rho\nu} &= [(\Delta_x - m^2)\varphi(x)]^\lambda E^{-}_{\lambda\nu}(x,x') \\ &\quad - \varphi^\lambda(x)[(\Delta_x - m^2)E^{-}(x,x')]_{\lambda\nu} \qquad (3.7) \\ &= [(\Delta_x - m^2)\varphi(x)]^\lambda E^{-}_{\lambda\nu}(x,x') \\ &\quad - \varphi^\lambda(x)\delta_{\lambda\nu}(x,x'). \end{aligned}$$

We consider in the four-dimensional space two regions, Ω' and Ω'' separated by a spatial hypersurface Σ. The η^ρ vector is orientated from Ω'' to Ω'. If φ^ν is the solution of the equation $(\Delta - m^2)\varphi^\nu = 0$, Eq. (3.7) involves

$$\nabla^\rho\{A^{-}(\varphi)\}_{\rho\nu} = \varphi^\lambda(x)\delta_{\lambda\nu}(x,x'). \tag{3.8}$$

Using Stokes thorem we write

$$\int_{\Omega''} \nabla^\rho\{A^{-}(\varphi)\}_{\rho\nu}\, d^4x = \mathrm{Flux}_\Sigma A^{-}(\varphi). \tag{3.9}$$

Taking into account (3.8) and (3.9) it follows that

$$\mathrm{Flux}_\Sigma A^{-}(\varphi) = \varphi_\nu(x')\epsilon'(x'), \tag{3.10}$$

where

$$\epsilon'(x') = \begin{cases} 1 & \text{if } x'\in\Omega'' \\ 0 & \text{if } x'\in\Omega'. \end{cases}$$

If we define

$$\{A^{+}(\varphi)\}_{\rho\nu} = E^{+}_{\lambda\nu}(x,x')\nabla_\rho\varphi^\lambda(x) - \varphi^\lambda(x)\nabla_\rho E^{+}_{\lambda\nu}(x,x')$$

we obtain

$$\mathrm{Flux}_\Sigma A^{+}(\varphi) = -\varphi_\nu(x')\epsilon''(x'), \tag{3.11}$$

where

$$\epsilon''(x') = \begin{cases} 1 & \text{if } x'\in\Omega' \\ 0 & \text{if } x'\in\Omega''. \end{cases}$$

Subtracting expression (3.10) from (3.11), there results

$$-\mathrm{Flux}(A^{+}(\varphi) - A^{-}(\varphi)) = \varphi_\nu(x'),$$

and taking into account the definitions of A^{+} and A^{-} we see that Eq. (3.3) holds, having as a result

$$G_{\mu\nu}(x,x') = E^{+}_{\mu\nu}(x,x') - E^{-}_{\mu\nu}(x,x').$$

It is easy to show that if the kernel $G_{\mu\nu}(x,x')$ satisfies (3.3), it must have the following properties:

$$G_{\mu\nu}(x,x') = 0,$$
$$n^\rho\nabla_\rho G_{\mu\nu}(x,x') = \mathscr{T}_{\mu\nu}\delta(x,x'),$$
$$n^\rho n^{\sigma'}\nabla_\rho\nabla_{\sigma'}G_{\mu\nu}(x,x') = 0,$$

for $x,x'\in\Sigma$.

We are going to deal now with the solution of the following system of equations:

$$(\Delta - m^2)\varphi_\nu = 0, \tag{3.12}$$
$$\nabla_\nu\varphi^\nu = 0. \tag{3.13}$$

It will be shown that if the following conditions are imposed:

$$\nabla_\mu\varphi^\mu(y) = 0, \quad n^\rho\partial_\rho(\nabla_\mu\varphi^\mu)(y) = 0, \quad y\in\Sigma \tag{3.14}$$

on a $\varphi_\mu(x)$ field that satisfies (3.12), then the field has covariant divergence equal to zero in all the points of the space-time.

Taking into account that

$$\nabla_\mu[(\Delta - m^2)\varphi]^\mu = (\Delta - m^2)(\nabla_\mu\varphi^\mu)$$

and as $\varphi^\mu(x)$ satisfies (3.12) it results that

$$(\Delta - m^2)(\nabla_\mu\varphi^\mu) = 0. \tag{3.15}$$

The covariant divergence $\nabla_\mu\varphi^\mu$ satisfies (3.15) and the boundary conditions (3.14). Using the Cauchy problem solution for a scalar field it follows that $\nabla_\mu\varphi^\mu = 0$ holds for all the points of the curved space-time. In this way we have shown that the system of equations (3.12) and (3.13) is equivalent to Eq. (3.12) with the boundary conditions (3.14) on the hypersurface Σ. The kernel $G_{\mu\nu}(x,x')$, defined above, satisfies Eq. (3.12) but has no covariant divergence equal to zero. It is easily shown that, in the case of a vector field with covariant divergence equal to zero, the kernel $G_{\mu\nu}(x,x')$ in (3.3) is replaced by:

$$X_{\mu\nu}(x,x') = G_{\mu\nu}(x,x') - \frac{\nabla_\mu\nabla_\nu}{m^2}G(x,x'), \tag{3.16}$$

where $G(x,x')$ is the kernel of the solution of the Cauchy problem for a scalar field.

Therefore we have the result

$$\begin{aligned} \varphi_\mu(x) = \int_\Sigma [&X_{\mu\lambda'}(x,x')\nabla_{\rho'}\varphi^{\lambda'}(x') \\ &- \varphi^{\lambda'}(x')\nabla_{\rho'}X_{\mu\lambda'}(x,x')]n^{\rho'}\, d\sigma'. \end{aligned} \tag{3.17}$$

Moreover, as $X_{\mu\nu}(x,x')$ satisfies the equations

$$(\Delta_x - m^2)X_{\mu\nu}(x,x') = g^{\rho\mu}\nabla_\rho X_{\mu\nu}(x,x') = 0,$$

this kernel has the proper properties to be considered the commutators of the massive vector field

$$[\varphi_\mu(x);\varphi_\nu(x')] = iX_{\mu\nu}(x,x').$$

4. BASE OF SOLUTIONS AND DECOMPOSITION IN POSITIVE AND NEGATIVE FREQUENCIES

In order to establish orthogonal conditions on the solutions of the field equations, it is necessary to define an inner product. Σ being a Cauchy hypersurface and u^ρ and v^ρ two vector fields, we define:

$$\begin{aligned} \langle u^\lambda, v_\lambda\rangle_\Sigma &= i\int_\Sigma \left[\left(\nabla_\rho \overset{*}{u}{}^\lambda\right)v_\lambda - \overset{*}{u}{}^\lambda(\nabla_\rho v_\lambda)\right]n^\rho\, d\sigma \\ &= i\int_\Sigma \theta_\rho n^\rho\, d\sigma. \end{aligned} \tag{4.1}$$

If u^ρ and v^ν are solutions of Eq. (3.1) it is easy to show that the inner product, defined in this way, does not depend

upon the hypersurface Σ. It is also shown that

$$\langle u^\lambda, v_\lambda \rangle = \langle v^\lambda, u_\lambda \rangle^*,$$

so that the product is Hermitian, and therefore $\langle u^\lambda, u_\lambda \rangle$ is real. It is important to emphasize that $\langle u^\lambda, u_\lambda \rangle$ is not a positive definite number. It can be shown using (4.1), that if

$$\langle u^\lambda, u_\lambda \rangle > 0 \text{ then } \left\langle \overset{*}{u}{}^\lambda, \overset{*}{u}_\lambda \right\rangle < 0.$$

We stress that θ_ρ in (4.1), with $u^\rho = v^\rho$ is the usual four-vector current when the field is complex.

Using (4.1), the Cauchy problem solution (3.17) is

$$\varphi_u(x) = i\langle X_{\mu\nu'}(x,x'); \varphi^{\nu'}_{(x')} \rangle, \quad x' \in \Sigma. \tag{4.2}$$

Let $\{ \phi^{\ \mu}_{(a)} \} \cup \{ \overset{*}{\phi}{}^{\ \mu}_{(a)} \}$ be a basis of the space of complex solutions of Eqs. (3.12) and (3.13), where (a) is a pair of indices, one, $\underline{k}$, continuous and the other s discrete (to denote momentum and spin, respectively, as we are going to see later on). This basis will be called orthonormal if it satisfies

$$\begin{aligned} &\langle \phi^{\ \mu}_{\underline{k}s}; \phi_{\mu \underline{k}'s'} \rangle = -\delta(\underline{k} - \underline{k}')\delta_{ss'}, \\ &\left\langle \overset{*}{\phi}{}^{\ \mu}_{\underline{k}s}; \overset{*}{\phi}_{\mu \underline{k}'s'} \right\rangle = \delta(\underline{k} - \underline{k}')\delta_{ss'}, \\ &\left\langle \phi^{\ \mu}_{\underline{k}s}; \overset{*}{\phi}_{\ \mu \underline{k}'s'} \right\rangle = 0. \end{aligned} \tag{4.3}$$

Any solution of the system of Eqs. (3.12) and (3.13) will be expanded in this basis in the following way:

$$\varphi^\mu(x) = \sum_s d^3\underline{k} \left(a_{\underline{k}s} \phi^{\ \mu}_{\underline{k}s}(x) + \overset{*}{a} \right)$$

where the coefficients of the expansion satisfy

$$\begin{aligned} a_{\underline{k}s} &= -\langle \phi^{\ \rho'}_{\underline{k}s}(x'); \varphi_{\rho'}(x') \rangle, \\ \overset{*}{a}_{\underline{k}s} &= \left\langle \overset{*}{\phi}{}^{\ \rho'}_{\underline{k}s}(x'); \varphi_{\rho'}(x') \right\rangle. \end{aligned} \tag{4.5}$$

In the quantization process the coefficients $a_{\underline{k}s}$ and $\overset{*}{a}_{\underline{k}s}$ become the operators that satisfy the following relations of commutation:

$$\begin{aligned} &[a_{\underline{k}s}; a_{\underline{k}'s'}] = [a^+_{\underline{k}s}; a^+_{\underline{k}'s'}] = 0, \\ &[a_{\underline{k}s}; a^+_{\underline{k}'s'}] = \delta(\underline{k} - \underline{k}')\delta_{ss'}. \end{aligned}$$

Replacing the expressions (4.5) in (4.4) results in

$$\begin{aligned} \varphi^{\ \mu}_{(x)} = \Big\langle \int d^3\underline{k} \sum_s \Big\{ &\phi^{\ \mu}_{\underline{k}s}(x) \overset{*}{\phi}{}^{\ \rho'}_{\underline{k}s}(x') \\ &- \overset{*}{\phi}{}^{\ \mu}_{\underline{k}s}(x) \phi^{\ \rho'}_{\underline{k}s}(x') \Big\}; \varphi_{\rho'}(x') \Big\rangle. \end{aligned} \tag{4.6}$$

Comparing (4.6) with (4.2) we see that we have obtained the development of the kernel $X^{\mu\rho'}(x,x')$ in the basis $\{ \phi^{\ \mu}_{\underline{k}s} \} \cup \{ \overset{*}{\phi}{}^{\ \mu}_{\underline{k}s} \}$

$$X^{\mu\rho'}(x,x') = i \int d^3\underline{k} \sum_s \left\{ \phi^{\ \mu}_{\underline{k}s}(x) \overset{*}{\phi}{}^{\ \rho'}_{\underline{k}s}(x') - \overset{*}{\phi}{}^{\ \mu}_{\underline{k}s}(x) \phi^{\ \rho'}_{\underline{k}s}(x') \right\}. \tag{4.7}$$

With the aim of performing the quantization in the curved space-time, it is necessary to generalize, in some way, the decomposition into positive and negative frequencies of the fields' theory of the flat space-time. Taking into account (4.6), the $\varphi^\mu(x)$ field can be decomposed in the following way:

$$\varphi^{\ \mu}(x) = \overset{\oplus}{\varphi}{}^{\ \mu}(x) + \overset{\ominus}{\varphi}{}^{\ \mu}(x),$$

where

$$\begin{aligned} \overset{\oplus}{\varphi}{}^{\ \mu}(x) &= \left\langle \int d^3\underline{k} \sum_s \left\{ \phi^{\ \mu}_{\underline{k}s}(x) \overset{*}{\phi}{}^{\ \rho'}_{\underline{k}s}(x') \right\}; \varphi_{\rho'}(x') \right\rangle, \\ \overset{\ominus}{\varphi}{}^{\ \mu}(x) &= \left\langle -\int d^3\underline{k} \sum_s \left\{ \overset{*}{\phi}{}^{\ \mu}_{\underline{k}s}(x) \phi^{\ \rho'}_{\underline{k}s}(x') \right\}; \varphi_{\rho'}(x') \right\rangle. \end{aligned} \tag{4.8}$$

Considering the orthogonality conditions (4.3) it is possible to find that

$$\begin{aligned} &\left\langle \overset{\oplus}{\varphi}{}^\mu(x), \overset{\oplus}{\varphi}_\mu(x) \right\rangle > 0, \\ &\left\langle \overset{\ominus}{\varphi}{}^\mu(x), \overset{\ominus}{\varphi}_\mu(x) \right\rangle < 0, \end{aligned}$$

therefore we shall assign $\overset{\oplus}{\varphi}$ and $\overset{\ominus}{\varphi}$ to the parts of positive and negative frequencies, respectively.

Now defining

$$\begin{aligned} \overset{\ominus}{X}{}^{\mu\rho'}(x,x') &= i \int d^3\underline{k} \sum_s \phi^{\ \mu}_{\underline{k}s}(x) \overset{*}{\phi}{}^{\ \rho'}_{\underline{k}s}(x'), \\ \overset{\oplus}{X}{}^{\mu\rho'}(x,x') &= -i \int d^3\underline{k} \sum_s \overset{*}{\phi}{}^{\ \mu}_{\underline{k}s}(x) \phi^{\ \rho'}_{\underline{k}s}(x'), \end{aligned}$$

it is possible to write

$$\begin{aligned} \overset{\oplus}{\varphi}{}^{\ \mu}(x) &= i \left\langle \overset{\ominus}{X}{}^{\mu\rho'}(x,x'); \varphi_{\rho'}(x') \right\rangle, \\ \overset{\ominus}{\varphi}{}^{\ \mu}(x) &= i \left\langle \overset{\oplus}{X}{}^{\mu\rho'}(x,x'); \varphi_{\rho'}(x') \right\rangle. \end{aligned}$$

We have to notice that the kernels $\overset{\oplus}{X}{}^{\mu\rho'}(x,x')$ and $\overset{\ominus}{X}{}^{\mu\rho'}(x,x')$ of the curved space-time, play a role similar to that of the kernels $\Delta^{\oplus\mu\rho}(x-x')$ and $\Delta^{\ominus\mu\rho}(x-x')$ of the field theory of the flat space-time.

It is possible to generalize the $\Delta_1^{\mu\rho}(x-x')$ kernel, defining

$$\begin{aligned} X_1^{\mu\rho'}(x,x') &= i \left(\overset{\oplus}{X}{}^{\mu\rho'} - \overset{\ominus}{X}{}^{\mu\rho'} \right)(x,x') \\ &= \int d^3\underline{k} \sum_s \left\{ \phi^{\ \mu}_{\underline{k}s}(x) \overset{*}{\phi}{}^{\ \rho'}_{\underline{k}s}(x') + \overset{*}{\phi}{}^{\ \rho'}_{\underline{k}s}(x') \phi^{\ \mu}_{\underline{k}s}(x) \right\}. \end{aligned} \tag{4.9}$$

There is an important difference between the properties of the $X^{\mu\rho'}(x,x')$ and $X_1^{\mu\rho'}(x,x')$ kernels. While the kernel $X^{\mu\rho'}(x,x')$ is unique, and therefore, when it is written in the form (4.7) results in being independent of the selected basis $\{ \phi^{\ \mu}_{\underline{k}s} \} \cup \{ \phi^{*\mu}_{\underline{k}s} \}$, the kernel $X_1^{\mu\rho'}(x,x')$ is not independent of the basis. This is due to the fact that the vector sets of positive and negative norm do not constitute a vectorial subspace.

In Refs. 1 and 2, the impossibility of determining a

unique kernel $G_1(x,x')$ for the case of the scalar field has been largely discussed. To solve the problem we suppose that a kernel $G_1^{(\Sigma)}(x,x')$ exists for each Cauchy hypersurface Σ. In a similar way we can solve the problem in the case of a vector field. We suppose that there is a different decomposition of solutions of positive (and negative) frequencies, for each Cauchy hypersurface Σ of the curve space-time. Consequently, there is going to be a kernel $X_1^{(\Sigma)\mu\rho'}(x,x')$, different for each Cauchy hypersurface Σ. This fact of nonuniqueness of the kernel $X_1^{(\Sigma)\mu\rho'}(x,x')$ is also shown when the Bogolyubov transformation is performed. While the kernel $X^{\mu\rho'}(x,x')$ is invariant under a transformation of this type, $X_1^{\mu\rho'}(x,x')$ is not (cf., e.g., Ref. 9). As was said in the introduction, we are going to consider the Bogolyubov transformation as a possible mechanism for particle creation. In the next section we shall analyze, for the case of a vector field, this transformation.

5. BOGOLYUBOV TRANSFORMATION

Let $\{\phi_{\underline{k}s}^{(\mathscr{T})\mu}\}\cup\{\phi_{\underline{k}s}^{(\mathscr{T})\mu *}\}$ be a basis of the space of solutions of the equations (3.12) and (3.13) that satisfy the initial conditions in the time $t=\mathscr{T}$; and $\{\phi_{\underline{k}s}^{(\mathscr{T}')\mu}\}\cup\{\phi_{\underline{k}s}^{*(\mathscr{T}')\mu}\}$ the basis of the space of solutions that satisfy the initial conditions in the time $t=\mathscr{T}'$ (cf. Ref. 12). As both sets of functions are basis solutions of the space of complex solutions of the same differential equations, there will have to be such adequate complex coefficients $\alpha_{\underline{k}s,\underline{k}'s'}$ and $\beta_{\underline{k}s,\underline{k}'s'}$ as will enable us to write a set of solutions of one function in terms of another, in the following way:

$$\phi_{\underline{k}s}^{(\mathscr{T})\mu}=\int d^3\underline{k}'\sum_{s'}\left(\alpha_{\underline{k}s,\underline{k}'s'}\phi_{\underline{k}'s'}^{(\mathscr{T}')\mu}+\beta_{\underline{k}s,\underline{k}'s'}\phi_{\underline{k}'s'}^{(\mathscr{T}')*\mu}\right). \tag{5.1}$$

This means that a classical solution, which is a solution of positive (or negative) frequencies in the "out" region, is a mixture of solutions of positive and negative frequencies in the "in" region and vice versa. As we shall see in Sec. 6, for the type of metric used, we are going to take as the basis of solutions those with well defined linear momentum $\underline{k}$. Equation (5.1) is then reduced to:

$$\phi_{\underline{k}s}^{(\mathscr{T})\mu}(x)=\sum_{s'}\left\{\alpha_{\underline{k}ss'}\phi_{\underline{k}s'}^{(\mathscr{T}')\mu}(x)+\beta_{-\underline{k}ss'}\phi_{-\underline{k}s'}^{(\mathscr{T}')*\mu}(x)\right\}, \tag{5.2}$$

and therefore

$$\phi_{\underline{k}s}^{(\mathscr{T})*\mu}(x)=\sum_{s'}\left(\alpha_{\underline{k}ss'}^{*}\phi_{\underline{k}s'}^{(\mathscr{T}')*\mu}(x)+\beta_{-\underline{k}ss'}^{*}\phi_{-\underline{k}s'}^{(\mathscr{T}')\mu}(x)\right). \tag{5.3}$$

From Eqs. (5.2), (5.3), and (4.4) the following Bogolyubov transformation on the $a_{\underline{k}s}$ and $a^{+}_{\underline{k}s}$ operators results:

$$a_{\underline{k}s'}^{(\mathscr{T}')}=\sum_{s}\left\{\alpha_{\underline{k}ss'}a_{\underline{k}s}^{(\mathscr{T})}+\beta_{\underline{k}ss'}^{*}a_{-\underline{k}s}^{+(\mathscr{T})}\right\}, \tag{5.4}$$

$$a_{\underline{k}s'}^{+(\mathscr{T}')}=\sum_{s}\left\{\alpha_{\underline{k}ss'}^{*}a_{\underline{k}s}^{+(\mathscr{T})}+\beta_{\underline{k}ss'}a_{-\underline{k}s}^{(\mathscr{T})}\right\}. \tag{5.5}$$

Using the orthogonality conditions (4.3) in $\mathscr{T}$ and in $\mathscr{T}'$ and taking into account (5.2) and (5.3) we arrive to the following equations for the $\alpha_{\underline{k}ss'}$ and $\beta_{\underline{k}ss'}$ coefficients:

$$\sum_{s}\left(\alpha_{\underline{k}ss'}\alpha_{\underline{k}s''s'}^{*}-\beta_{-\underline{k}ss'}\beta_{-\underline{k}s''s'}^{*}\right)=\delta_{ss''}, \tag{5.6}$$

$$\sum_{s'}\left(\beta_{-\underline{k}ss'}^{*}\alpha_{-\underline{k}s''s'}^{*}-\alpha_{\underline{k}ss'}^{*}\beta_{\underline{k}s''s'}^{*}\right)=0. \tag{5.7}$$

On the other hand, if we suppose that in the time $t=\mathscr{T}$ the universe is in the $|0\rangle_{\mathscr{T}}$ vacuum-state, we have the result

$$N_{\underline{k}s}^{(\mathscr{T})}|0\rangle_{\mathscr{T}}=a_{\underline{k}s}^{+(\mathscr{T})}a_{\underline{k}s}^{(\mathscr{T})}|0\rangle_{\mathscr{T}}=0,$$

where $N_{\underline{k}s}^{(\mathscr{T})}$ is the operator number of particles in the $\underline{k}$ mode and s spin component. Making use of Eqs. (5.4) and (5.5) it follows that the mean number of particles present, in the $\underline{k}$ mode, and with s spin components, in a $(La(t))^3$ volume and in a time $\mathscr{T}'>\mathscr{T}$ is,

$${}_{\mathscr{T}}\langle 0|N_{\underline{k}s}^{(\mathscr{T}')}|0\rangle_{\mathscr{T}}=\sum_{s'}|\beta_{\underline{k}s's}|^2.$$

Therefore the mean density of particles present in the time $\mathscr{T}'>\mathscr{T}$ is

$$N^{(\mathscr{T}')}=\lim_{L\to\infty}\frac{1}{[La(t)]^3}\sum_{\underline{k},s}{}_{\mathscr{T}}\langle 0|N_{\underline{k}s}^{(\mathscr{T}')}|0\rangle_{\mathscr{T}};$$

replacing $\Sigma_{\underline{k}}\to(L/2\pi)^3\int d^3\underline{k}$ leads to

$$N^{(\mathscr{T}')}=\frac{1}{[2\pi a(t)]^3}\int d^3\underline{k}\sum_{ss'}|\beta_{\underline{k}s's}|^2. \tag{5.8}$$

The necessary and sufficient condition to obtain finite mean density of particles must be

$$\int d^3\underline{k}\sum_{ss'}|\beta_{\underline{k}ss'}|^2<\infty. \tag{5.9}$$

6. SOLUTION OF THE FIELD EQUATIONS

We are going to realize our study of the massive vector field for the metric (2.8). The only nonvanishing components of the Christoffel symbols are

$$\Gamma_{ii}^{0}=a(t)\dot{a}(t),\quad \Gamma_{0i}^{i}=\Gamma_{i0}^{i}=\frac{\dot{a}(t)}{a(t)}.$$

In the previous expressions we do not have to sum on the repeated latin indices.

The components of the contracted curvature tensor $R_{\mu\nu\rho}{}^{\mu}$ are

$$R_{00}=-3\frac{\ddot{a}(t)}{a(t)},$$

$$R_{ii}=2\dot{a}^2(t)+a(t)\ddot{a}(t),$$

$$R_{\mu\nu}=0 \text{ if } \mu\neq\nu.$$

With the Christoffel symbols and the contracted curvature it is possible to develop Eq. (2.3) and (2.4). From Eq. (2.3) there results

$$\frac{\partial^2}{\partial t^2}\varphi^0-\frac{1}{a^2(t)}\nabla^2\varphi^0+5H(t)\frac{\partial}{\partial t}\varphi^0$$

$$\frac{\partial^2}{\partial t^2}\varphi^0 - \frac{1}{a^2(t)}\nabla^2\varphi^0 + 5H(t)\frac{\partial}{\partial t}\varphi^0 + \left(m^2 - \frac{R(t)}{2}\right)\varphi^0 = 0, \tag{6.1}$$

$$\frac{\partial^2}{\partial t^2}\varphi^j - \frac{1}{a^2(t)}\nabla^2\varphi^j + 5H(t)\frac{\partial}{\partial t}\varphi^j + \left[m^2 - \frac{R(t)}{3} + 2H^2(t)\right]\varphi^j - \frac{2H(t)}{a^2(t)}\frac{\partial}{\partial x^j}\varphi^0 = 0, \tag{6.2}$$

where $H(t) = \dot{a}(t)/a(t)$ is the Hubble coefficient and $R(t) = R^{\mu}_{\mu}(t) = -6(\ddot{a}(t)/a(t) + H^2(t))$ is the scalar curvature.

Equation (2.4) is written

$$\frac{\partial}{\partial t}\varphi^0 + \sum_{j=1}^{3}\frac{\partial}{\partial x^j}\varphi^j + 3H(t)\varphi^0 = 0. \tag{6.3}$$

A basis of solutions of the system (6.1), (6.2), and (6.3) may be obtained through the method of separation of variables. In this manner we obtain the solutions in the form

$$\phi^{\mu}_{\underline{h}}(t, \vec{x}) = f^{\mu}_{\underline{h}}(t)e^{-i\underline{h}\cdot\vec{x}} \tag{6.4}$$

The functions $f^{\mu}_{\underline{h}}(t)$ must satisfy the following system of ordinary differential equations:

$$\ddot{f}^0_{\underline{h}}(t) + 5H(t)\dot{f}^0_{\underline{h}}(t) + \left(\omega^2 - \frac{R(t)}{2}\right)f^0_{\underline{h}}(t) = 0, \tag{6.5}$$

$$\ddot{f}^j_{\underline{h}}(t) + 5H(t)\dot{f}^j_{\underline{h}}(t) + \left(\omega^2 - \frac{R(t)}{3} + 2H^2(t)\right)f^j_{\underline{h}}(t) = 2i\underline{h}^j H(t)f^0_{\underline{h}}(t), \tag{6.6}$$

$$\dot{f}^0_{\underline{h}}(t) - i\sum_{m=1}^{3}\underline{h}_m f^m_{\underline{h}}(t) + 3H(t)f^0_{\underline{h}}(t) = 0, \tag{6.7}$$

where $w = (m^2 + \underline{h}^2/a^2)^{1/2}$.

As we have demonstrated, Eq. (6.3) can be replaced by the conditions (3.14) on the Cauchy hypersurface Σ. Replacing (6.4) in the conditions (3.14) and using Eq. (6.5) we obtain

$$\dot{f}^0_{\underline{h}}(\mathscr{T}) - i\sum_{j=1}^{3}\underline{h}_j f^j_{\underline{h}}(\mathscr{T}) + 3H(\mathscr{T})f^0_{\underline{h}}(\mathscr{T}) = 0,$$

$$2H(\mathscr{T})\dot{f}^0_{\underline{h}}(\mathscr{T}) + [w^2(\mathscr{T}) + 6H^2(\mathscr{T})]f^0_{\underline{h}}(\mathscr{T}) + i\sum_{m=1}^{3}\underline{h}_m\dot{f}^m_{\underline{h}}(\mathscr{T}) = 0. \tag{6.8}$$

From here onwards the differential equation (6.7) will be replaced by the initial conditions (6.8). The system of equations of second order, (6.5) and (6.6), allow eight linearly independent solutions. Taking into account the condition of zero divergence or the equivalent initial conditions (6.8), the linearly independent solutions are reduced to six. Nevertheless, we only need to know three solutions, which we are going to call $f^{\mu}_{\underline{h}s}(t)$, with $s = 1, 2, 3$. We are also going to indicate $\phi^{\mu}_{\underline{h}s}(t, \vec{x}) = f^{\mu}_{\underline{h}s}(t)e^{-i\underline{h}\cdot\vec{x}}$. First it is shown that, if $f^{\mu}_{\underline{h}s}(t)$ satisfies (6.5), (6.6), and (6.8), $f^{\mu *}_{\underline{h}s}(t)$ also satisfies them. If the $\phi^{\mu}_{\underline{h}s}(t, \vec{x})$ functions satisfy the orthogonality conditions (4.3), it is easily demonstrated that the six functions $f^{\mu}_{\underline{h}s}(t)$ and $f^{\mu *}_{\underline{h}s}(t)$ are linearly independent. Any real solution of the system of equations

$$(\Delta - m^2)\varphi^{\nu} = 0, \quad \nabla^{\mu}\varphi_{\mu} = 0,$$

can be written as a linear combination of the functions $\phi^{\mu}_{\underline{h}s}(x)$ and $\phi^{\mu *}_{\underline{h}s}(x)$, that is to say,

$$\varphi^{\mu}(x) = \int d^3\underline{h}\sum_{s=1}^{3}\{a_{\underline{h}s}\phi^{\mu}_{\underline{h}s}(x) + a^*_{\underline{h}s}\phi^{*\mu}_{\underline{h}s}(x)\}.$$

We have justified in this way the assumed development in (4.4).

We have to emphasize that the orthogonality conditions (4.3) in $t = \mathscr{T}$ together with those of zero divergence (6.8) are not sufficient to determine the 24 initial conditions

$$f^{\mu}_{\underline{h}s}(\mathscr{T}) = A^{\mu}_{\underline{h}s}, \quad \dot{f}^{\mu}_{\underline{h}s}(\mathscr{T}) = B^{\mu}_{\underline{h}s}, \tag{6.9}$$

which are necessary to determine in a unique way a basis for the solutions of the problem. Additional conditions, such as the QEP (Sec. 7) or the diagonalization of the Hamiltonian operator (Sec. 8), are necessary to define the basis of solutions in each time $\mathscr{T}$.

7. QUANTUM EQUIVALENCE PRINCIPLE

To define a bitensorial kernel $\overset{(\Sigma)}{X}{}^{\mu\nu'}_{1}(x, x')$ for each Cauchy hypersurface Σ, the following Cauchy data are necessary

$$\overset{(\Sigma)}{X}{}^{\mu\nu'}_{1}(x,x'); \quad n^{\rho}\partial_{\rho}\overset{(\Sigma)}{X}{}^{\mu\nu'}_{1}(x,x'); \quad x, x' \in \Sigma. \tag{7.1}$$

We recall that in the flat space-time, and in Cartesian coordinates, the kernel $\Delta^{\mu\nu'}_{1}(x - x')$ is written

$$\Delta^{\mu\nu'}_{1}(x - x') = (g^{\mu\nu} - \partial^{\mu}\partial^{\nu'}/m^2)\Delta_1(x - x'), \tag{7.2}$$

where (cf e.g. Ref. 13):

$$\Delta_1(x - x') = \frac{1}{(2\pi)^3}\int\frac{\cos w_k(t - t')}{w_k}e^{-i\underline{k}\cdot(\vec{x}' - \vec{x})}d^3\underline{k} = -\frac{m^2}{4\pi}\,\mathrm{Im}\left(\frac{H^{(1)}_1[m((x'_{\mu} - x_{\mu})(x'^{\mu} - x^{\mu}))^{1/2}]}{m((x'_{\mu} - x_{\mu})(x'^{\mu} - x^{\mu}))^{1/2}}\right). \tag{7.3}$$

If we write the expression (7.2) in the flat space-time, but in curvilinear coordinates, we obtain

$$\Delta^{\mu\nu'}_{1}(x, x') = \left[t^{\mu\nu'}(x, x') - \frac{1}{m^2}g^{\mu\alpha}_{(x)} \times\frac{\partial}{\partial x^{\alpha}}g^{\nu'\beta'}(x')\frac{\partial}{\partial x^{\beta'}}\right]\Delta_1(s), \tag{7.4}$$

where S is the geodesic arc that joins the x and x' points, and $t^{\mu\nu'}(x, x')$ is the transport bitensor (cf. e.g., Ref. 3).

As a generalization of the Quantum Equivalence Principle used in Ref. 2 for the case of the scalar field, we are going to state for the case of the vector field the following QEP:

"The $\overset{(\Sigma)}{X}{}^{\mu\nu'}_{1}(x, x')$ kernel has on the normal Cauchy hypersurface Σ the following Cauchy data:

$$\overset{(\Sigma)}{X}{}^{\mu\nu'}_{1}(x, x') = \Delta^{\mu\nu'}_{1}(x, x'), \tag{7.5}$$

$$n^{\rho'}\partial_{\rho'}\overset{(\Sigma)}{X}{}_1^{\mu\nu'}(x,x') = n^{\rho'}\partial_{\rho'}\Delta_1^{\mu\nu'}(x,x'),$$

$$x, x' \in \Sigma, \tag{7.6}$$

where in (7.5) and (7.6) the kernel $\Delta_1^{\mu\nu'}(x,x')$ is the transcription of the expression (7.4) at the curve space-time."

Now, we shall make an evaluation of the Cauchy data for the metric (2.8). For simplicity we suppose that the normal hypersurface Σ is the hyperplane $t = \mathscr{T}$. To calculate explicitly $\Delta_1(S)$ and $t^{\mu\nu'}(x,x')$ and to obtain with them $\Delta_1^{\mu\nu'}(x,x')$ (cf. Ref. 2) we are going to make use of Riemannian normal coordinates (cf. e.g., Ref. 14). The use of such coordinates, as was largely discussed in Ref. 2, makes clear the global nature of the problem. The kernels $X_1^{(\Sigma)\mu\nu'}(x,x')$ and $\Delta_1^{\mu\nu'}(x,x')$ are equal on the whole normal Cauchy hypersurface. Then we can carry out the calculus to all orders of H when we write the Cauchy data (7.5) and (7.6) using normal coordinates.

Let a point O with coordinates $x_{(0)} = (\mathscr{T}, \mathbf{x}_0)$ and another point P with coordinates $x' = (t', \mathbf{x}')$, and $S =$ the length of the geodesic arc that joins O with P. We are going to give the name v^μ to the components of the unitary tangent vector to the geodesic arc at the point O, expressed in an orthonormal tetrad with its temporal axis normally orientated to the hypersurface Σ at the point O. We shall call normal coordinates of the point P, with an origin at the point O, those obtained with the following coordinates transformation:

$$\overset{(N)}{x}{}^{\mu'} = \delta_\mu^{\mu'} v_{(0)}^\mu S. \tag{7.7}$$

We have to notice that, in the system of reference, the geodesics passing through the point O are straight, and also they become:

$$S^2 = g_{\mu'\rho'}(x_{(0)})\overset{(N)}{x}{}^{\mu'}\overset{(N)}{x}{}^{\rho'}. \tag{7.8}$$

The transformation (7.7), for the metric (2.8), is written

$$\begin{aligned}\overset{(N)}{t}{}' &= \Delta t + \tfrac{1}{2}a^2 H(\Delta\bar{x})^2 + \tfrac{1}{3}a^2(\tfrac{1}{12}R + H^2)\\ &\quad\times(\Delta\bar{x})^2\Delta t + O(H^3, \Delta x^4),\end{aligned} \tag{7.9}$$

$$\begin{aligned}\overset{(N)}{\bar{x}}{}' &= \Delta\bar{x} + H\Delta t\Delta\bar{x} + \tfrac{1}{2}a^2H^2(\Delta\bar{x})^2\Delta\bar{x}\\ &\quad + \tfrac{1}{3}(R/6 - H^2)\Delta t^2\Delta\bar{x} + \bar{O}(H^3, \Delta x^4),\end{aligned}$$

where $\Delta t = t' - \mathscr{T}$ and $\Delta\bar{x} = \bar{x}' - \bar{x}_{(0)}$. If the deceleration parameter $q \cong 1$, the scalar curvature R is of the same order of magnitude as H^2. $O(H^3, \Delta x^4)$ and $\bar{O}(H^3, \Delta x^4)$ are terms that tend to zero like H^3 when $H \to 0$, and like Δx^4 when $\Delta x \to 0$.

Taking into account (7.8) and (7.3), and the metric (2.8), we can write

$$\begin{aligned}\Delta_1(s) &= \Delta_1\left(\overset{(N)}{t}{}'^2 - a^2(\mathscr{T})\overset{(N)}{\bar{x}}{}'^2\right)^{1/2}\\ &= \frac{1}{(2\pi)^3}\int\frac{\cos w_k t^{(N)'}}{w_k}e^{-i\underline{k}\cdot\overset{(N)}{\bar{x}}{}'/a(\mathscr{T})}d^3\underline{k}.\end{aligned}$$

If in the previous expression we replace $t'^{(N)}$ and $\bar{x}'^{(N)}$ [Eq. (7.9)], we retain the terms up to the H order, and we make the variable change $\underline{h} = a(\mathscr{T})\underline{k}$, we obtain

$$\begin{aligned}\Delta_1(S) &\cong \frac{1}{(2\pi a(\mathscr{T}))^3}\int\frac{\cos w(\Delta t + \tfrac{1}{2}a^2(\mathscr{T})H(\mathscr{T})\Delta\bar{x}^2)}{w}\\ &\quad\times e^{-i\underline{h}\cdot(\Delta\bar{x} + H(\mathscr{T})\Delta t\,\Delta\bar{x})}d^3\underline{h},\end{aligned} \tag{7.10}$$

where $w = (m^2 + \underline{h}^2/a^2(\mathscr{T}))^{1/2}$.

Now, we are going to obtain explicitly the transport bitensor $t^{\mu\nu'}(x,x')$. Let O and P be two points of the curved space-time and S the geodesic arc that joins them. Also let u^α be a vector at point O of coordinates x, and $u^{\alpha'}$ the vector at point P of coordinates x' obtained by parallel transport of u^α vector on the S geodesic arc. From its definition the transport bitensor satisfies the equation

$$u^{\alpha'} = t_\alpha^{\alpha'}(x,x')u^\alpha, \tag{7.11}$$

and also has the property

$$t_\alpha^{\alpha'}(x, x' = x) = \delta_\alpha^{\alpha'}. \tag{7.12}$$

According to (7.11) $t_\alpha^{\alpha'}(x,x')$ is a covariant 1-tensor at x, x, and a contravariant 1-tensor at x'. The $u^{\alpha'}$ vector at point P can be also written

$$\begin{aligned}u^{\alpha'}(x') &= \delta_\alpha^{\alpha'}u^\alpha(x) - \delta_\alpha^{\alpha'}\\ &\quad\times\int_0^P\Gamma_{\beta\gamma}^\alpha(y)u^\beta(y)\,dy^\gamma,\end{aligned} \tag{7.13}$$

where the integral of the second member has to be made on the S geodesic arc that joins O with P.

The covariant and contravariant components of the metric tensor, in normal coordinates, neglecting order H^3 terms are:

$$\begin{aligned}\overset{(N)}{g}{}_{o'o'}(x') &= 1 - \tfrac{1}{3}a^2(\tfrac{1}{6}R - H^2)\overset{(N)}{\bar{x}}{}'^2,\\ \overset{(N)}{g}{}_{o'i'}(x') &= \tfrac{1}{3}a^2(\tfrac{1}{6}R - H^2)\delta_{k'i'}\overset{(N)}{x}{}^{k'}\overset{(N)}{t}{}',\\ \overset{(N)}{g}{}_{i'j'}(x') &= -a^2\delta_{j'i'}\left[1 + \tfrac{1}{3}(\tfrac{1}{6}R - H^2)\overset{(N)}{t}{}'^2\right]\\ &\quad + \tfrac{1}{3}a^4H^2\left[\overset{(N)}{\bar{x}}{}'^2\delta_{j'i'} - \overset{(N)}{x}{}^{k'}\overset{(N)}{x}{}^{h'}\delta_{k'i'}\delta_{h'j'}\right],\\ \overset{(N)}{g}{}^{o'o'}(x') &= 1 + \tfrac{1}{3}a^2(\tfrac{1}{6}R - H^2)\overset{(N)}{\bar{x}}{}'^2,\\ \overset{(N)}{g}{}^{o'i'}(x') &= \tfrac{1}{3}(\tfrac{1}{6}R - H^2)\overset{(N)}{x}{}^{i'}\overset{(N)}{t}{}',\\ \overset{(N)}{g}{}^{i'j'}(x') &= -(1/a^2)\delta_{j'}^{i'}\left[1 - \tfrac{1}{3}(\tfrac{1}{6}R - H^2)\overset{(N)}{t}{}'^2\right]\\ &\quad - \tfrac{1}{3}H^2\left[\overset{(N)}{\bar{x}}{}'^2\delta_{j'}^{i'} - \overset{(N)}{x}{}^{i'}\overset{(N)}{x}{}^{j'}\right].\end{aligned} \tag{7.14}$$

From (7.14), and with the same approximation, the following components of the affine connection are obtained:

$$\begin{aligned}\overset{(N)}{\Gamma}{}_{j'k'}^{i'}(x') &= -\tfrac{1}{3}a^2H^2(\delta_{j'}^{i'}\delta_{m'}^{k'} + \delta_{k'}^{i'}\delta_{m'}^{j'}\\ &\quad - 2\delta_{j'}^{k'}\delta_{m'}^{i'})\overset{(N)}{x}{}^{m'},\end{aligned}$$

$$\overset{(N)}{\Gamma}{}^{0'}_{j'k'}(x') = \tfrac{2}{3}a^2(\tfrac{1}{6}R - H^2)\overset{(N)}{t}{}'\delta^{k'}_{j'},$$

$$\overset{(N)}{\Gamma}{}^{0'}_{0'k'}(x') = -\tfrac{1}{3}a^2(\tfrac{1}{6}R - H^2)\overset{(N)}{x}{}^{m'}\delta^{k'}_{m'},$$

$$\overset{(N)}{\Gamma}{}^{i'}_{0'k'}(x') = \tfrac{1}{3}(\tfrac{1}{6}R - H^2)\overset{(N)}{t}{}'\delta^{i'}_{k'},$$

$$\overset{(N)}{\Gamma}{}^{i'}_{0'0'}(x') = -\tfrac{2}{3}(\tfrac{1}{6}R - H^2)\overset{(N)}{x}{}^{i'},$$

$$\overset{(N)}{\Gamma}{}^{0'}_{0'0'}(x') = 0. \tag{7.15}$$

In (7.14) and (7.15) a, H, R, are evaluated in $t = \mathscr{T}$.

In Eq. (7.13) we can replace $u^\beta(y)$ by $u^\beta(y = x)$ if we retain only the terms in H^2. Then, in the normal system we have the result:

$$\overset{(N)}{u}{}^{\alpha'}(x') \cong \delta^{\alpha'}_{\alpha}\overset{(N)}{u}{}^{\alpha}(x) - \delta^{\alpha'}_{\alpha}\overset{(N)}{u}{}^{\beta}(x)\int_0^{P(N)}\Gamma^{\alpha}_{\beta\gamma}(y)\,\overset{(N)}{d}y^{\gamma}. \tag{7.16}$$

Comparing (7.16) to (7.11) we find that the transport bitensor is:

$$\overset{(N)}{t}{}^{\alpha'}_{\alpha} = \delta^{\alpha'}_{\alpha} - \delta^{\alpha'}_{\epsilon}\delta^{\gamma}_{\alpha}\int_0^{P(N)}\Gamma^{\epsilon}_{\gamma\nu}\overset{(N)}{d}y^{\nu} + O(H^4). \tag{7.17}$$

Replacing (7.15) in (7.17) and integrating we obtain:

$$\begin{aligned}
&\overset{(N)}{t}{}^{0'}_{0} = 1 + \tfrac{1}{6}a^2(\tfrac{1}{6}R - H^2)\overset{(N)}{\bar{x}}{}'^2,\\
&\overset{(N)}{t}{}^{0'}_{k} = -\tfrac{1}{6}a^2(\tfrac{1}{6}R - H^2)\overset{(N)}{t}{}'\overset{(N)}{x}{}^{k'},\\
&\overset{(N)}{t}{}^{i'}_{k} = \delta^{i'}_{k} - \tfrac{1}{3}(\tfrac{1}{6}R - H^2)\delta^{i'}_{k}\tfrac{1}{2}\overset{(N)}{t}{}'^2\\
&\qquad + \frac{a^2H^2}{3}\left\{\delta^{i'}_{k}\tfrac{1}{2}\overset{(N)}{\bar{x}}{}'^2 - \tfrac{1}{2}\overset{(N)}{x}{}^{i'}\overset{(N)}{x}{}^{k'}\right\},\\
&\overset{(N)}{t}{}^{i'}_{0} = \tfrac{1}{6}(\tfrac{1}{6}R - H^2)\overset{(N)}{x}{}^{i'}\overset{(N)}{t}{}'.
\end{aligned} \tag{7.18}$$

Using the inverse transformation of (7.9) we write the transport bitensor in the coordinates where the metric takes the form (2.8):

$$\begin{aligned}
&t^{00'}(x, x') = 1 + \tfrac{1}{2}a^2H^2(\Delta\bar{x})^2,\\
&t^{0i'}(x, x') = -\left[H + \tfrac{1}{2}(\tfrac{1}{6}R - 3H^2)\Delta t\right]\Delta x^i,\\
&t^{i0'}(x, x') = \left[H + \tfrac{1}{2}(\tfrac{1}{6}R - H^2)\Delta t\right]\Delta x^i,\\
&t^{ij'}(x, x') = -(1/a^2(\mathscr{T}))\left|\left[1 - H\Delta t - \tfrac{1}{2}(\tfrac{1}{6}R - H^2)\Delta t^2\right]\right.\\
&\qquad \times\delta^i_j + \tfrac{1}{2}a^2(\mathscr{T})H^2\Delta x^i\Delta x^j\Big\}.
\end{aligned} \tag{7.19}$$

In order to simplify the calculus, we are only going to retain linear terms in H. For computing

$$\begin{aligned}
&\Delta_1^{\mu\nu'}(x, x')\big|_{t = t' = \mathscr{T}}\\
&\quad = \left(t^{\mu\nu'}_{(x,x')} - \frac{1}{m^2}g^{\mu\alpha}_{(x)}g^{\nu'\beta'}_{(x')}\frac{\partial}{\partial x^\alpha}\frac{\partial}{\partial x^{\beta'}}\right)\Delta_1(s)\Big|_{t = t' = \mathscr{T}}
\end{aligned} \tag{7.20}$$

we use the expressions (7.10) and (7.19), and we take into account that in the sense of distribution it is verified that

$$\begin{aligned}
&\int d^3\underline{h}\,g(\underline{h})(\Delta x^i)^n e^{-i\underline{h}\cdot\overline{\Delta x}}\\
&\quad = (-i)^n\int d^3\underline{h}\,e^{-i\underline{h}\cdot\overline{\Delta x}}\frac{\partial^n}{\partial h_i^n}g(\underline{h}).
\end{aligned} \tag{7.21}$$

The components of the kernel $\Delta_1^{\mu\nu'}(x, x')$ on the Cauchy hypersurface $t = \mathscr{T}$ are:

$$\begin{aligned}
\Delta_1^{ji'}\big|_{\Delta t = 0} &= -\frac{1}{(2\pi a)^3}\int d^3\underline{h}\,e^{-i\underline{h}\cdot\overline{\Delta x}}\\
&\quad\times\left(\delta^i_j\frac{1}{a^2w} + \frac{h_ih_j}{a^4m^2w}\right),\\
\Delta_1^{0i'}\big|_{\Delta t = 0} &= \frac{-i}{(2\pi a)^3}\int d^3\underline{h}\,e^{-i\underline{h}\cdot\overline{\Delta x}}\\
&\quad\times\frac{Hh_i}{a^2}\left(\frac{1}{2w^3} - \frac{1}{m^2w}\right),\\
\Delta_1^{i0'}\big|_{\Delta t = 0} &= \frac{i}{(2\pi a)^3}\int d^3\underline{h}\,e^{-i\underline{h}\cdot\overline{\Delta x}}\\
&\quad\times\frac{Hh_i}{a^2}\left(\frac{1}{2w^3} - \frac{1}{m^2w}\right),\\
\Delta_1^{00'}\big|_{\Delta t = 0} &= \frac{1}{(2\pi a)^3}\int d^3\underline{h}\,e^{-i\underline{h}\cdot\overline{\Delta x}}\left(\frac{1}{w} - \frac{w}{m^2}\right).
\end{aligned} \tag{7.22}$$

In order to determine all the Cauchy data on the hyperplane $t = \mathscr{T}$, it is also necessary to know $\partial/\partial t\,\Delta_1^{\mu\nu'}(x, x')$. Deriving the equation (7.20) results in:

$$\begin{aligned}
\frac{\partial}{\partial t'}\Delta_1^{\mu\nu'}(x, x')\Big|_{\Delta t = 0} &= \left\{\left[\frac{\partial}{\partial t'}t^{\mu\nu'}(x, x')\right.\right.\\
&\quad - \frac{1}{m^2}g^{\mu\alpha}(x)\left(\frac{\partial}{\partial t'}g^{\nu'\beta'}(x')\right)\frac{\partial}{\partial x^\alpha}\frac{\partial}{\partial x^{\beta'}}\Big]\Delta_1(s)\\
&\quad + \left(t^{\mu\nu'}(x, x') - \frac{1}{m^2}g^{\mu\alpha}(x)\right.\\
&\quad\left.\left.\times g^{\nu'\beta'}(x')\frac{\partial}{\partial x^\alpha}\frac{\partial}{\partial x^{\beta'}}\right)\frac{\partial}{\partial t'}\Delta_1(s)\right\}_{\Delta t = 0}.
\end{aligned}$$

Using in this equation the expressions (7.19) and (7.10) and working similarly we obtain

$$\begin{aligned}
&\frac{\partial}{\partial t'}\Delta_1^{ji'}\Big|_{\Delta t = 0}\\
&\quad = \frac{1}{(2\pi a)^3}\int d^3\underline{h}\,e^{-i\underline{h}\cdot\Delta\bar{x}}\left[\delta^i_j\left(\frac{2H}{a^2w} + \frac{Hm^2}{2a^2w^2}\right)\right.\\
&\qquad\left. + \frac{h_ih_j}{m^2a^4}H\left(\frac{3}{w} + \frac{m^2}{2w^3}\right)\right],
\end{aligned} \tag{7.23}$$

$$\begin{aligned}
\frac{\partial}{\partial t'}\Delta_1^{0i'}\Big|_{\Delta t = 0} &= \frac{\partial}{\partial t'}\Delta_1^{i0'}\Big|_{\Delta t = 0}\\
&= -\frac{i}{(2\pi a)^3}\int d^3\underline{h}\,e^{-i\underline{h}\cdot\Delta\bar{x}}\left(\frac{wh_i}{a^2m^2}\right),\\
\frac{\partial}{\partial t'}\Delta_1^{00'}\Big|_{\Delta t = 0} &= \frac{1}{(2\pi a)^3}\int d^3\underline{h}\,e^{-i\underline{h}\cdot\overline{\Delta x}}\\
&\quad\times H\left[-\frac{3}{2w} + \frac{2w}{m^2} - \frac{m^2}{2w^3}\right].
\end{aligned}$$

We have seen that Eq. (4.9) gives us the kernel $X_1^{\mu\nu'}(x, x')$ in functions of the basis of solutions. In our case this basis of solutions is:

$$\phi^{\mu}_{\underline{hs}}(x) = f^{\mu}_{\underline{hs}}(t)e^{-i\underline{h}\cdot\bar{x}}. \tag{7.24}$$

Replacing (7.24) in (4.9) we obtain:

$$X_1^{\mu\rho'}(x, x') = \int d^3\underline{h} \sum_{s=1}^{3} \left\{ \dot{f}^{\mu}_{\underline{hs}}(\mathscr{T}) f^{\rho'}_{\underline{hs}}(t') + f^{\mu}_{-\underline{hs}}(\mathscr{T}) \dot{f}^{\rho'}_{-\underline{hs}}(t') \right\} e^{-i\underline{h}(\bar{x}-\bar{x}_{(0)})}. \tag{7.25}$$

Equation (7.25) involves:

$$X_1^{\mu\rho'}(x, x')\big|_{t'=\mathscr{T}} = \int d^3\underline{h} \sum_{s=1}^{3} \left\{ \dot{f}^{\mu}_{\underline{hs}}(\mathscr{T}) f^{\rho'}_{\underline{hs}}(\mathscr{T}) + f^{\mu}_{-\underline{hs}}(\mathscr{T}) \dot{f}^{\rho'}_{-\underline{hs}}(\mathscr{T}) \right\} e^{-i\underline{h}(\bar{x}-\bar{x}_{(0)})}, \tag{7.26}$$

$$\frac{\partial}{\partial t'} X_1^{\mu\rho'}(x, x')\Big|_{t'=\mathscr{T}} = \int d^3\underline{h} \sum_{s=1}^{3} \left[\dot{f}^{\mu}_{\underline{hs}}(\mathscr{T}) \dot{f}^{\rho'}_{\underline{hs}}(\mathscr{T}) + f^{\mu}_{-\underline{hs}}(\mathscr{T}) \dot{f}^{*\rho'}_{-\underline{hs}}(\mathscr{T}) \right] \times e^{-i\underline{h}(\bar{x}-\bar{x}_{(0)})}. \tag{7.27}$$

Introducing in (7.5) Eqs. (7.22) and (7.26) results in:

$$\sum_{s=1}^{3} \left\{ \dot{f}^{i}_{-\underline{hs}}(\mathscr{T}) f^{0}_{-\underline{hs}}(\mathscr{T}) + \dot{f}^{0}_{\underline{hs}}(\mathscr{T}) f^{i}_{\underline{hs}}(\mathscr{T}) \right\} = -\frac{i}{(2\pi a)^3} \frac{H}{a^2} h_i \left[\frac{1}{2w^3} - \frac{1}{m^2 w} \right],$$

$$\sum_{s=1}^{3} \left\{ \dot{f}^{i}_{-\underline{hs}}(\mathscr{T}) f^{j}_{-\underline{hs}}(\mathscr{T}) + \dot{f}^{j}_{\underline{hs}}(\mathscr{T}) f^{i}_{\underline{hs}}(\mathscr{T}) \right\} = -\frac{1}{(2\pi a)^3} \left[\frac{\delta^i_j}{a^2 w} + \frac{h_i h_j}{a^4 w m^2} \right], \tag{7.28}$$

$$\sum_{s=1}^{3} \left\{ \dot{f}^{0}_{-\underline{hs}}(\mathscr{T}) f^{i}_{-\underline{hs}}(\mathscr{T}) + \dot{f}^{i}_{\underline{hs}}(\mathscr{T}) f^{0}_{\underline{hs}}(\mathscr{T}) \right\} = \frac{i}{(2\pi a)^3} \frac{H}{a^2} h_i \left[\frac{1}{2w^3} - \frac{1}{m^2 w} \right],$$

$$\sum_{s=1}^{3} \left\{ \dot{f}^{0}_{-\underline{hs}}(\mathscr{T}) f^{0}_{-\underline{hs}}(\mathscr{T}) + \dot{f}^{0}_{\underline{hs}}(\mathscr{T}) f^{0}_{\underline{hs}}(\mathscr{T}) \right\} = \frac{1}{(2\pi a)^3} \left[\frac{1}{w} - \frac{w}{m^2} \right].$$

Similarly, introducing in (7.6) Eqs. (7.23) and (7.27) we obtain

$$\sum_{s=1}^{3} \left\{ \dot{f}^{i}_{-\underline{hs}}(\mathscr{T}) f^{0}_{-\underline{hs}}(\mathscr{T}) + \dot{f}^{0}_{\underline{hs}}(\mathscr{T}) f^{i}_{\underline{hs}}(\mathscr{T}) \right\} = -\frac{i}{(2\pi a)^3} \frac{w h_i}{a^2 m^2},$$

$$\sum_{s=1}^{3} \left\{ \dot{f}^{0*}_{-\underline{hs}}(\mathscr{T}) f^{i}_{-\underline{hs}}(\mathscr{T}) + \dot{f}^{i}_{\underline{hs}}(\mathscr{T}) f^{0}_{\underline{hs}}(\mathscr{T}) \right\} = -\frac{i}{(2\pi a)^3} \frac{w h_i}{a^2 m^2},$$

$$\sum_{s}^{3} \left\{ \dot{f}^{i*}_{-\underline{hs}}(\mathscr{T}) f^{j}_{-\underline{hs}}(\mathscr{T}) + \dot{f}^{j}_{\underline{hs}}(\mathscr{T}) \dot{f}^{i}_{\underline{hs}}(\mathscr{T}) \right\} = \frac{1}{(2\pi a)^3} \left\{ \delta^i_j \left[\frac{2H}{a^2 w} + \frac{Hm^2}{2a^2 w^3} \right] + \frac{h_i h_j}{m^2 a^4} H \left[\frac{3}{w} + \frac{m^2}{2w^3} \right] \right\}, \tag{7.29}$$

$$\sum_{s=1}^{3} \left\{ \dot{f}^{0*}_{-\underline{hs}}(\mathscr{T}) f^{0}_{-\underline{hs}}(\mathscr{T}) + \dot{f}^{0*}_{\underline{hs}}(\mathscr{T}) f^{0}_{\underline{hs}}(\mathscr{T}) \right\} = \frac{1}{(2\pi a)^3} H \left[\frac{2w}{m^2} - \frac{3}{2w} - \frac{m^2}{2w^3} \right].$$

The orthogonality conditions (4.3) imply the following equations on $f^{\mu}_{\underline{hs}}(t)$ and $\dot{f}^{\mu}_{\underline{hs}}(t)$, in $t = \mathscr{T}$:

$$\sum_{j=1}^{3} \left\{ \dot{f}^{j}_{\underline{hs}}(\mathscr{T}) \dot{f}^{j}_{\underline{hs}'}(\mathscr{T}) - \dot{f}^{*j}_{\underline{hs}}(\mathscr{T}) f^{j}_{\underline{hs}'}(\mathscr{T}) \right\} = \frac{1}{a^2} \left\{ \frac{i\delta^s_{s'}}{(2\pi a)^3} + \left(\dot{f}^{0}_{\underline{hs}}(\mathscr{T}) \dot{f}^{0}_{\underline{hs}'}(\mathscr{T}) - \dot{f}^{*0}_{\underline{hs}}(\mathscr{T}) f^{0}_{\underline{hs}'}(\mathscr{T}) \right) \right\}, \tag{7.30}$$

$$\sum_{j=1}^{3} \left\{ \dot{f}^{j*}_{\underline{hs}}(\mathscr{T}) f^{j*}_{-\underline{hs}'}(\mathscr{T}) - f^{j*}_{\underline{hs}}(\mathscr{T}) \dot{f}^{j*}_{-\underline{hs}'}(\mathscr{T}) \right\} = \frac{1}{a^2} \left\{ \dot{f}^{0*}_{\underline{hs}}(\mathscr{T}) f^{0*}_{-\underline{hs}'}(\mathscr{T}) - f^{0*}_{\underline{hs}}(\mathscr{T}) \dot{f}^{0*}_{-\underline{hs}'}(\mathscr{T}) \right\}.$$

The system of equations (7.28)–(7.30) and (6.8) for $f^{\mu}_{\underline{hs}}(\mathscr{T})$ and $\dot{f}^{\mu}_{\underline{hs}}(\mathscr{T})$ has the following solution:

$$f^{0}_{\underline{hs}}(\mathscr{T}) = -\frac{h_s}{a(2\pi a)^{3/2}\sqrt{2w}} \times \left[\frac{1}{m} + i\frac{Hm}{2w^3} \left(\frac{2w^2}{m^2} - 1 \right) \right],$$

$$\frac{d}{dt} f^{0}_{\underline{hs}}(\mathscr{T}) = +\frac{h_s}{a(2\pi a)^{3/2}\sqrt{2w}} \left[\frac{3H}{m} - i\frac{w}{m} \right], \tag{7.31}$$

$$f^{j}_{\underline{hs}}(\mathscr{T}) = -\frac{a}{(2\pi a)^{3/2}\sqrt{2w}} \left[\frac{\delta^j_s}{a^2} + \frac{h_j h_s}{a^4 m(m+w)} \right],$$

$$\frac{d}{dt} f^{j}_{\underline{hs}}(\mathscr{T}) = \frac{a}{(2\pi a)^{3/2}\sqrt{2w}} \left\{ \left(2H + \frac{Hm^2}{2w^2} - iw \right) \times \left(\frac{\delta^j_s}{a^2} + \frac{h_j h_s}{a^4 m(m+w)} \right) + \frac{H}{mw} \frac{h_j h_s}{a^4} \right\}.$$

We have to remember that in the initial conditions (7.31) the terms in H^2 had been ignored.

We are going to see later on (Sec. 9) that with the initial conditions (7.31) obtained from the QEP, the mean density of created particles for the Bogolyubov transformation (5.2) is finite.

8. HAMILTONIAN DIAGONALIZATION

The metric (2.8) is independent of the spatial coordinates $(\bar{x})$, therefore the P_i linear momentum becomes a constant of motion. If we develop the field in a basis $\{\phi^{(\mathscr{T})\mu}_{\underline{hs}}(x)\}$

$\cup\{\phi^{(\mathscr{T})}\mu^{*}_{\underline{hs}}\}$ that satisfies the orthogonality conditions (4.3), it is easy to demonstrate that the $\hat{P}_i$ operator is diagonal,

$$\hat{P}_i(\mathscr{T}) = a^3(\mathscr{T})\int d^3\bar{x}:\hat{T}^0_i:$$

$$= -\int d^3\underline{h}\, h_i \sum_{s=1}^{3} a^+_{\underline{hs}}(\mathscr{T})a_{\underline{hs}}(\mathscr{T}). \tag{8.1}$$

A basis that in $t=\mathscr{T}$ satisfies the orthogonality conditions (4.3) or (7.30), is orthogonal for all other times since the defined inner product (Sec. 4) is independent of the Cauchy hypersurface. For this reason Eq. (8.1) for the operator $\hat{P}_i$ is valid for any time $\mathscr{T}$, independently of any other additional conditions imposed on the chosen basis. On the other hand, as the metric (2.8) depends on the time (t) the energy of the massive vector field is not conserved, and if the operator

$$\hat{P}_0(\mathscr{T}) = a^3(\mathscr{T})\int d^3\bar{x}:\hat{T}^0_0(\mathscr{T},\bar{x}):, \tag{8.2}$$

is developed in an orthogonal basis the result is not generally diagonal.

The QEP is not the only way to complete the initial conditions $f^{\mu}_{\underline{hs}}(\mathscr{T})$ and $\dot{f}^{\mu}_{\underline{hs}}(\mathscr{T})$ to determine the basis of solutions of positive and negative frequencies, on the hypersurface $t=\mathscr{T}$. We can also obtain the additional conditions, if instead of using the QEP, we require that the Hamiltonian operator be diagonal in $t=\mathscr{T}$, that is to say,

$$\hat{P}_0(\mathscr{T}) = \int d^3\underline{h} w(\mathscr{T}) \sum_{s=1}^{3} a^+_{\underline{hs}}(\mathscr{T})a_{\underline{hs}}(\mathscr{T}). \tag{8.3}$$

For this purpose we use the following development of the field operator $\hat{\varphi}^{\mu}(x)$ in an orthogonal basis:

$$\hat{\varphi}^{\mu}(x) = \int d^3\underline{h} \sum_{s=1}^{3}\left\{a_{\underline{hs}}(\mathscr{T})\overset{(\mathscr{T})}{\phi}{}^{\mu}_{\underline{hs}}(x) + a^+_{\underline{hs}}(\mathscr{T})\overset{(\mathscr{T})}{\phi}{}^{\mu}_{\underline{hs}}{}^{*}(x)\right\}. \tag{8.4}$$

If we replace (2.9) in (8.2), and then use (8.4) we obtain for $P_0(\mathscr{T})$ an expression of the form

$$\hat{P}_0(\mathscr{T}) = \int d^3\underline{h} \sum_{ss'=1}^{3}\{c_{ss'}(\mathscr{T},\underline{h})a^+_{\underline{hs}}(\mathscr{T})a^+_{-\underline{hs}'}(\mathscr{T}) + c^*_{ss'}(\mathscr{T},\underline{h})a_{\underline{hs}}(\mathscr{T})a_{-\underline{hs}'}(\mathscr{T}) + D_{ss'}(\mathscr{T},\underline{h})a^+_{\underline{hs}}(\mathscr{T})a_{\underline{hs}'}(\mathscr{T})\}. \tag{8.5}$$

To diagonalize $\hat{P}_0(\mathscr{T})$ we must have $C_{ss'}(\mathscr{T},\underline{h})=0$. Then it follows that:

$$\sum_j\{w^2(\mathscr{T})f^*_{j\underline{hs}}(\mathscr{T})f^*_{j-\underline{hs}'}(\mathscr{T}) + \dot{f}^*_{j\underline{hs}}(\mathscr{T})\dot{f}^*_{j-\underline{hs}'}(\mathscr{T})\}$$
$$= -a^2(\mathscr{T})([w^2(\mathscr{T}) + 9H^2(\mathscr{T})]f^*_{0\underline{hs}}(\mathscr{T})f^*_{0-\underline{hs}'}(\mathscr{T}) + \dot{f}^*_{0\underline{hs}}(\mathscr{T})\dot{f}^*_{0-\underline{hs}'}(\mathscr{T}) + 3H(\mathscr{T})\{f^*_{0\underline{hs}}(\mathscr{T})\times\dot{f}^*_{0-\underline{hs}'}(\mathscr{T}) + \dot{f}^*_{0\underline{hs}}(\mathscr{T})f^*_{0-\underline{hs}'}(\mathscr{T})\}). \tag{8.6}$$

Equation (8.6), the orthogonality conditions (7.30), and the condition (6.8) of zero divergence determine, in a unique way, the following initial conditions:

$$f^j_{\underline{hs}}(\mathscr{T}) = \frac{-a}{(2\pi a)^{3/2}(2w)^{1/2}}\left(\frac{\delta^j_s}{a^2} + \frac{h_j h_s}{a^4 m(m+w)}\right)$$

$$\frac{d}{dt}f^j_{\underline{hs}}(\mathscr{T}) = \frac{a}{(2\pi a)^{3/2}(2w)^{1/2}}\left\{(2H - iw)\times\left(\frac{\delta^j_s}{a^2} + \frac{h_j h_s}{a^4 m(m+w)}\right)\right\}, \tag{8.7}$$

$$f^0_{\underline{hs}}(\mathscr{T}) = -\frac{h_s}{ma(2\pi a)^{3/2}(2w)^{1/2}},$$

$$\frac{d}{dt}f^0_{\underline{hs}}(\mathscr{T}) = -\frac{h_s}{a(2\pi a)^{3/2}m(2w)^{1/2}}(3H - iw).$$

The conditions $f^j_{\underline{hs}}(\mathscr{T})$ and $\dot{f}^j_{\underline{hs}}(\mathscr{T})$ are the same as those obtained in the quantization of the massive vector field in the flat space-time (cf. e.g. ref. 13).

It is important to emphasize that, in order to obtain Eq. (8.6) we have not supposed $H\Delta x \ll 1$, and therefore the solutions (8.7) are exact.

Obviously the basis compatible with the initial conditions (8.7), diagonalize the operator $\hat{P}_0$ only in $t=\mathscr{T}$.

We are going to see later on (Sec. 9) that with the initial conditions (8.7), the corresponding Bogolyubov transformation leads us to an infinite density of created particles.

9. PARTICLE CREATION

Finally, we want to analyze the particle creation when we use the theories based either on the QEP or on the diagonalization of the Hamiltonian. To do so, we are going to use the results obtained by Olver on the valuation of error in the approximation WKB.[15]

The first derivatives in the differential equations (6.5) and (6.6) can be eliminated, making the following change of variables:

$$\eta = \int_{\mathscr{T}}^{t} a^{-1}(t')\,dt', \quad g^{\mu}_{(\eta)} = a^2(t(\eta))f^{\mu}(t(\eta)). \tag{9.1}$$

Then, Eqs. (6.5) and (6.6) are:

$$\frac{d^2}{d\eta^2}g^0(\eta) + a^2\left[w^2 - \tfrac{1}{6}R - 2H^2\right]g^0(\eta) = 0, \tag{9.2}$$

$$\frac{d^2}{d\eta^2}g^j(\eta) + a^2w^2g^j(\eta) = -2iHh_jg^0(\eta). \tag{9.3}$$

Defining:

$$P_T(\underline{h},\eta) = (a^2/h^2)(w^2 - 2H^2 - \tfrac{1}{6}R),$$

Eq. (9.2) can be written so as to use the Olver results (Ref. 15 Theorem 4, p. 800),

$$\frac{d^2}{d\eta^2}g^0(\eta) + \underline{h}^2P_T(\underline{h},\eta)g^0(\eta) = 0. \tag{9.4}$$

There is a solution of Eq. (9.4) that is written:

$$g^T(\eta) = (2h)^{-1/2}P_T^{-1/4}\times\left[\exp\left(-ih\int_0^{\eta}P_T^{1/2}\,d\eta'\right) + \epsilon_T(\underline{h},\eta)\right], \tag{9.5}$$

which has the following derivative:

$$\frac{d}{d\eta}g^T(\eta) = -i\left(\frac{h}{2}\right)^{1/2}P_T^{1/4}\left\{\left[1 - \frac{i}{4h}P_T^{-3/2}\frac{d}{d\eta}P_T\right]\times\exp\left(-ih\int_0^{\eta}P_T^{1/2}\,d\eta'\right) - \frac{i}{4h}P_T^{-3/2}\frac{d}{d\eta}P_T\epsilon_T(\underline{h},\eta) + i\delta_T(\underline{h},\eta)\right\}. \tag{9.6}$$

According to Olver it is easy to show that

$$|\epsilon_T(\leftrightarrows h, \eta)|, \quad |\delta_T(\leftrightarrows h, \eta)|,$$
$$\left(h^{-1}P_T^{-3/2}\left|\frac{d}{d\eta}P_T\right|\right) \lesssim O(h^{-3}), \tag{9.7}$$
$$\epsilon_T(\underline{h}, 0) = \delta_T(\underline{h}, 0) = 0.$$

The general solution of (9.2) or (9.4) can be written

$$g^0(\eta) = A^0 g^T(\eta) + B^0 \dot{g}^T(\eta), \tag{9.8}$$

where A^0 and B^0 are two arbitrary constants.

The solutions $g^T(\eta)$ and $g^{T*}(\eta)$ of (9.4) satisfy

$$W(g^T, g^{T*})(\eta) = g^T \frac{d}{d\eta} g^{T*} - g^{T*} \frac{d}{d\eta} g^T = i. \tag{9.9}$$

Now we are going to solve Eq. (9.3). We can consider, in this equation, the second member as a inhomogeneity. The homogeneous equation associated to (9.3) is

$$\frac{d^2}{d\eta^2} g^j(\eta) + h^2 P_E(h, \eta) g^j(\eta) = 0, \tag{9.10}$$

with $P_E(\underline{h}, \eta) = a^2 w^2 / h^2$.

Using the Olver approximation a solution $g^E(\eta)$ can be obtained from (9.10). These $g^E(\eta)$ have properties similar to those of Eqs. (9.5)–(9.7), and (9.9), when we replaced in them, T for E.

In this way, we find that the general solution of (9.10) is:

$$g^j_{\text{homog}}(\eta) = A^j g^E(\eta) + B^j \dot{g}^E(\eta).$$

A particular solution of (9.3) can be obtained by the method of variation of the constants:

$$g^j_{\text{PART}}(\eta) = g^E(\eta) 2h_j [A^0 F_1(h, \eta) + B^0 F_2(h, \eta)]$$
$$- \dot{g}^E(\eta) 2h_j [A^0 F_2^*(h, \eta) + B^0 F_1^*(h, \eta)], \tag{9.11}$$

$$F_1(h, \eta) = \int_0^\eta H(\eta') \dot{g}^E(\eta') g^T(\eta')\, d\eta', \tag{9.12}$$

$$F_2(h, \eta) = \int_0^\eta H(\eta') \dot{g}^E(\eta') \dot{g}^T(\eta')\, d\eta'. \tag{9.13}$$

Then the general solution of (9.3) is

$$g^j(\eta) = A^j g^E(\eta) + B^j \dot{g}^E(\eta) + g^j_{\text{PART}}(\eta). \tag{9.14}$$

Among the general solutions (9.8) and (9.14), the useful ones will be those which satisfy certain intitial conditions. In $t = \mathscr{T}$ ($\eta = 0$) they are named $g^{(0)\mu}_{\underline{hs}}(\eta)$ and in $t = \mathscr{T}'(\eta = \xi)$ they are named $g^{(\xi)\mu}_{\underline{hs}}(\eta)$.

We write these conditions in the form

$$\overset{(0)}{g}{}^{\mu}_{\underline{hs}}(\eta = 0) = G^{\mu}_{\underline{hs}}(0), \quad \frac{d}{d\eta} \overset{(0)}{g}{}^{\mu}_{\underline{hs}}(\eta = 0) = \overset{\blacktriangledown}{G}{}^{\mu}_{\underline{hs}}(0), \tag{9.15}$$

$$g^{(\xi)\mu}_{\underline{hs}}(\eta = \xi) = G^{\mu}_{\underline{hs}}(\xi), \quad \frac{d}{d\eta} \overset{(\xi)}{g}{}^{\mu}_{\underline{hs}}(\eta = \xi) = \overset{\blacktriangledown}{G}{}^{\mu}_{\underline{hs}}(\xi). \tag{9.16}$$

Later on, it will be useful to suppose that the initial conditions (9.15) and (9.16) are

$$G^0_{\underline{hs}} = G^0_{\underline{h}} h_s, \quad G^{\blacktriangledown 0}_{\underline{hs}} = G^{\blacktriangledown 0}_{\underline{h}} h_s, \tag{9.17}$$

$$G^j_{\underline{hs}} = \gamma_h \delta^j_s + \mu_h h_j h_s, \quad G^{\blacktriangledown j}_{\underline{hs}} = \gamma^{\blacktriangledown}_h \delta^j_s + \mu^{\blacktriangledown}_h h_j h_s. \tag{9.18}$$

We can determine in (9.8) and (9.14) the values of the constants A^0, B^0, A^j, and B^j in such a way that the conditions (9.16) are satisfied. These values are:

$$A^0 = \overset{(\xi)}{A}{}^0_{\underline{hs}} = -ih_s \left[G^0_h(\xi) \frac{d}{d\eta} \dot{g}^T(\xi) - G^{\blacktriangledown 0}_h(\xi) \dot{g}^T(\xi) \right], \tag{9.19a}$$

$$B^0 = \overset{(\xi)}{B}{}^0_{\underline{hs}} = ih_s \left[G^0_h(\xi) \frac{d}{d\eta} g^T(\xi) - G^{\blacktriangledown 0}_h(\xi) g^T(\xi) \right], \tag{9.19b}$$

$$A^j = \overset{(\xi)}{A}{}^j_{\underline{hs}} = -i \Big\{ \frac{d}{d\eta} \dot{g}^E(\xi) [\gamma_h(\xi) \delta^j_s + \mu_h(\xi) h_j h_s]$$
$$- [\gamma^{\blacktriangledown}_h(\xi) \delta^j_s + \mu^{\blacktriangledown}_h(\xi) h_j h_s] \dot{g}^E(\xi) \Big\}$$
$$+ 2ih_j h_s \Big\{ F_1(h, \xi) \Big[G^0_h(\xi) \frac{d}{d\eta} \dot{g}^T(\xi)$$
$$- G^{\blacktriangledown 0}_h(\xi) \dot{g}^T(\xi) \Big] + F_2(h, \xi)$$
$$\times \Big[- G^0_h(\xi) \frac{d}{d\eta} g^T(\xi) + G^{\blacktriangledown 0}_h(\xi) g^T(\xi) \Big] \Big\}, \tag{9.20a}$$

$$B^j = \overset{(\xi)}{B}{}^j_{\underline{hs}} = i \Big\{ [\gamma_h(\xi) \delta^j_s + \mu_h(\xi) h_j h_s] \frac{d}{d\eta} g^E(\xi)$$
$$- [\gamma^{\blacktriangledown}_h(\xi) \delta^j_s + \mu^{\blacktriangledown}_h(\xi) h_j h_s] g^E(\xi) \Big\}$$
$$- i2h_j h_s \Big\{ \dot{F}_2(h, \xi) \Big[G^0_h(\xi) \frac{d}{d\eta} \dot{g}^T(\xi)$$
$$- G^{\blacktriangledown 0}_h(\xi) \dot{g}^T(\xi) \Big] + \dot{F}_1(h, \xi)$$
$$\times \Big[- G^0_h(\xi) \frac{d}{d\eta} g^T(\xi) + G^{\blacktriangledown 0}_h(\xi) g^T(\xi) \Big] \Big\}. \tag{9.20b}$$

With these constants we can write

$$\overset{(\xi)}{g}{}^0_{\underline{hs}}(\eta) = \overset{(\xi)}{A}{}^0_{\underline{hs}} g^T(\eta) + \overset{(\xi)}{B}{}^0_{\underline{hs}} \dot{g}^T(\eta), \tag{9.21}$$

$$\overset{(\xi)}{g}{}^j_{\underline{hs}}(\eta) = \overset{(\xi)}{A}{}^j_{\underline{hs}} g^E(\eta) + \overset{(\xi)}{B}{}^j_{\underline{hs}} \dot{g}^E(\eta) + \overset{(\xi)}{g}{}^j_{\text{PART}}(\eta). \tag{9.22}$$

If in (9.19)–(9.22) we set $\xi = 0$, we obtain the solutions that satisfy (9.15) and which we will to call $g^{(0)\mu}_{\underline{hs}}(\eta)$.

The initial conditions are selected in such a way that the orthogonality and zero divergence are obtained, and therefore, the six functions $g^{(0)\mu}_{\underline{hs}}(\eta), g^{(0)*}_{\underline{hs}'}(\eta)$ $(s, s' = 1, 2, 3)$ are linearly independent.

Then any solution of the system (9.2) and (9.3) with zero divergence can be written as a linear combination of those functions. In particular,

$$\overset{(\xi)}{g}{}^{\mu}_{\underline{hs}}(\eta) = \sum_{s'} \alpha_{\underline{hss}'} \overset{(0)}{g}{}^{\mu}_{\underline{hs}'}(\eta) + \beta_{-\underline{hss}'} \overset{(0)}{g}{}^{\mu *}_{-\underline{hs}'}(\eta). \tag{9.23}$$

If we remember (9.1) and (6.4) we see that (9.23) implies

$$\phi^{(\mathscr{T}')\mu}_{\underline{hs}}(x) = \sum_{s'} \alpha_{\underline{hss}'} \overset{(\mathscr{T})}{\phi}{}^{\mu}_{\underline{hs}'}(x) + \beta_{-\underline{hss}'} \phi^{(\mathscr{T})\mu *}_{-\underline{hs}'}(x),$$

which is the Bogolyubov transformation (5.2).

We now have to see if with the initial conditions (7.31) or with the initial conditions (8.7) the result is

$$\int d^3\underline{h} \sum_{ss'} |\beta_{\underline{h}ss'}|^2 < \infty. \tag{9.24}$$

We can suppose that $\alpha_{\underline{h}ss'}$ and $\beta_{\underline{h}ss'}$ can be written as follows:

$$\begin{aligned} \alpha_{\underline{h}ss'} &= C\,\delta^s_{s'} + Dh_s h_{s'}, \\ \beta_{\underline{h}ss'} &= E\,\delta^s_{s'} + Jh_s h_{s'}, \end{aligned} \tag{9.25}$$

where C, D, E, and J depend on the modulus of $\underline{h}$ and, of course, of $\mathscr{T}$ and $\mathscr{T}'$.

Replacing (9.25) in (9.24) we obtain

$$\int_0^\infty dh\, h^2\{2|E|^2 + |E + Jh^2|^2\} < \infty. \tag{9.26}$$

If we can show that $\{2|E|^2 + |E + Jh^2|^2\}$ behaves like $O(h^{-(3+\epsilon)})$ with $\epsilon > O$ when $h \to \infty$, we will have proved that the mean density of created particles is finite. Then we must try to find equations that let us determine the behavior of E and of $E + Jh^2$ when $h \to \infty$.

If we replace (9.21), and (9.21) with $\xi = O$, in the zero component of (9.23), and then we take into account the linear independence of $g^T(\eta)$ and $g^{T*}(\eta)$, we obtain

$$\begin{aligned} \overset{(\xi)}{A}{}^0_{\underline{h}s} &= \sum_{s'} \left\{ \alpha_{\underline{h}ss'} \overset{(0)}{A}{}^0_{\underline{h}s'} + \beta_{-\underline{h}ss'} \overset{(0)}{B}{}^{0*}_{-\underline{h}s'} \right\}, \\ \overset{(\xi)}{B}{}^0_{\underline{h}s} &= \sum_{s'} \left\{ \alpha_{\underline{h}ss'} \overset{(0)}{B}{}^0_{\underline{h}s'} + \beta_{-\underline{h}ss'} \overset{(0)}{A}{}^{0*}_{-\underline{h}s'} \right\}. \end{aligned} \tag{9.27}$$

If we now use in these equations the expressions (9.25), (9.19) and (9.19) with $\xi = 0$, we obtain a system of two equations with the two unknown quantities $(C + Dh^2)$ and $(E + Jh^2)$. For the latter we obtain

$$\begin{aligned} E + Jh^2 = \Bigg\{ & \left[G^0_h \frac{d}{d\eta} g^{T*} - G^{\blacktriangledown 0}_h g^T \right]_{(\xi)} \\ & \times \left[G^{\blacktriangledown 0}_h g^T - G^0_h \frac{d}{d\eta} g^T \right]_{(O)} \\ & - \left[G^0_h \frac{d}{d\eta} \overset{*}{g}{}^T - G^{\blacktriangledown 0}_h \overset{*}{g}{}^T \right]_{(O)} \\ & \times \left[G^{\blacktriangledown 0}_h g^T - G^0_h \frac{d}{d\eta} g^T \right]_{(\xi)} \Bigg\} \\ & \times \frac{1}{i} \left[G^0_h \overset{*}{G}{}^{\blacktriangledown 0}_h - \overset{*}{G}{}^0_h G^{0\blacktriangledown}_h \right]^{-1}_{(O)}. \end{aligned} \tag{9.28}$$

We can now replace (9.22) and (9.22) with $\xi = 0$ in the spatial components ($\mu = j$) of (9.23). Using (9.27) and considering that $g^E(\eta)$ and $g^{E*}(\eta)$ are linearly independent, we find:

$$\begin{aligned} \overset{(\xi)}{A}{}^j_{\underline{h}s} &= \sum_{s'} \left\{ \alpha_{\underline{h}ss'} \overset{(0)}{A}{}^j_{\underline{h}s'} + \beta_{-\underline{h}ss'} \overset{(0)}{B}{}^{j*}_{-\underline{h}s'} \right\}, \\ \overset{(\xi)}{B}{}^j_{\underline{h}s} &= \sum_{s'} \left\{ \alpha_{\underline{h}ss'} \overset{(0)}{B}{}^j_{\underline{h}s'} + \beta_{-\underline{h}ss'} \overset{(0)}{A}{}^{j*}_{-\underline{h}s'} \right\}. \end{aligned} \tag{9.29, 9.30}$$

We now replace Eqs. (9.25), (9.20a), and (9.20a) with $\xi = 0$, and (9.20b) with $\xi = 0$ in Eq. (9.29). If we consider likewise that δ^j_s and $h_j h_s$ are two tensors independent linearly, we obtain from the part of the equation that is proportional to δ^j_s, the relation

$$\begin{aligned} & \left\{ \gamma_h \frac{d}{d\eta} \overset{*}{g}{}^E - \gamma^{\blacktriangledown}_h \overset{*}{g}{}^E \right\}_{(\xi)} \\ & \quad = C \left\{ \gamma_h \frac{d}{d\eta} \overset{*}{g}{}^E - \gamma^{\blacktriangledown}_h \overset{*}{g}{}^E \right\}_{(0)} \\ & \qquad + E \left\{ \overset{*}{\gamma}_h \frac{d}{d\eta} \overset{*}{g}{}^E - \overset{*}{\gamma}{}^{\blacktriangledown}_h \overset{*}{g}{}^E \right\}_{(0)}. \end{aligned} \tag{9.31}$$

Working in a similar way with Eq. (9.30) the result is:

$$\begin{aligned} \left\{ \gamma_h \frac{d}{d\eta} g^E - \gamma^{\blacktriangledown}_h g^E \right\}_{(\xi)} &= C \left\{ \gamma_h \frac{d}{d\eta} g^E - \gamma^{\blacktriangledown}_h g^E \right\}_{(0)} \\ &+ E \left\{ \overset{*}{\gamma}_h \frac{d}{d\eta} g^E - \overset{*}{\gamma}{}^{\blacktriangledown}_h g^E \right\}_{(0)}. \end{aligned} \tag{9.32}$$

From (9.31) and (9.32) we can resolve E:

$$\begin{aligned} E = \Bigg\{ & \left[\gamma_h \frac{d}{d\eta} \overset{*}{g}{}^E - \gamma^{\blacktriangledown}_h \overset{*}{g}{}^E \right]_{(0)} \left[\gamma_h \frac{d}{d\eta} g^E - \gamma^{\blacktriangledown}_h g^E \right]_{(\xi)} \\ & - \left[\gamma_h \frac{d}{d\eta} \overset{*}{g}{}^E - \gamma^{\blacktriangledown}_h \overset{*}{g}{}^E \right]_{(\xi)} \left[\gamma_h \frac{d}{d\eta} g^E - \gamma^{\blacktriangledown}_h g^E \right]_{(0)} \Bigg\} \\ & \times \frac{1}{i} [\gamma^*_h \gamma^{\blacktriangledown}_h - \gamma_h \gamma^{\blacktriangledown *}_h]^{-1}_{(0)}. \end{aligned} \tag{9.33}$$

Now we have to make an evaluation of the asymptotic behavior ($h \to \infty$) of (9.28) and (9.33), for the initial conditions (7.31) of the QEP, and for the initial conditions (8.7) that diagonalize the Hamiltonian. Making the change of variables (9.1), both conditions can be written:

$$\begin{aligned} \overset{(\xi)}{g}{}^0_{\underline{h}s}(\eta = \xi) = G^0_h(\xi) h_s &= \left\{ \frac{-ah_s}{(2\pi a)^{3/2}\sqrt{2w}} \right. \\ & \left. \times \left[\frac{1}{m} + iH\sigma \left(\frac{1}{mw} - \frac{m}{2w^3} \right) \right] \right\}_{(\eta = \xi)}, \end{aligned} \tag{9.34}$$

$$\begin{aligned} \frac{d}{d\eta} \overset{(\xi)}{g}{}^0_{\underline{h}s}(\eta = \xi) = G^{\blacktriangledown 0}_h(\xi) h_s &= \left\{ \frac{-a^2 h_s}{(2\pi a)^{3/2}\sqrt{2w}} \right. \\ & \left. \times \left[-\frac{H}{m} + i\frac{w}{m} \right] \right\}_{(\eta = \xi)}, \end{aligned} \tag{9.35}$$

$$\begin{aligned} \overset{(\xi)}{g}{}^j_{\underline{h}s}(\eta = \xi) &= \gamma_h(\xi)\delta^j_s + \mu_h(\xi) h_j h_s \\ &= \left\{ \frac{-1}{a(2\pi a)^{3/2}\sqrt{2w}} \left(-\delta^j_s + \frac{h_j h_s}{m(m+w)} \right) \right\}_{(\eta = \xi)}, \end{aligned} \tag{9.36}$$

$$\begin{aligned} \frac{d}{d\eta} \overset{(\xi)}{g}{}^j_{\underline{h}s}(\eta = \xi) &= \gamma^{\blacktriangledown}_h(\xi)\delta^j_s + \mu^{\blacktriangledown}_h(\xi) h_j h_s \\ &= \left\{ \frac{1}{(2\pi a)^{3/2}\sqrt{2w}} \left[\left(\frac{\sigma H m^2}{2w^2} - iw \right) \right. \right. \\ & \left. \left. \times \left(-\delta^j_s + \frac{h_j h_s}{m(m+w)} \right) + \frac{\sigma H}{mw} h_j h_s \right] \right\}_{(\eta = \xi)}. \end{aligned} \tag{9.37}$$

In the previous expressions we have had to consider $\sigma = 1$ for the QEP, and $\sigma = 0$ for the diagonalization of the Hamiltonian. Similar expresions result for $\xi = 0$.

From the expressions (9.31) and (9.37) we easily obtained the constant G^0_h, $G^{\blacktriangledown 0}_h$, γ_h, μ_h, $\gamma^{\blacktriangledown}_h$, and $\mu^{\blacktriangledown}_h$. With these

constants and the expressions for g^E and $(d/d\eta)g^E$ obtained from Olver results, the equations (9.33) let us demonstrate that $E \lesssim O(h^{-3})$, when $h \to \infty$.

This result is worth indistinctly, for the conditions that diagonalize the Hamiltonian and for those of the QEP.

If we replace now the expressions (9.5), (9.6) and the values of G_h^0 and $G_h^{0\blacktriangledown}$ in (9.28), we obtain:

$$\begin{aligned} E + Jh^2 = &\frac{1}{4}\left[-2i + \frac{(\sigma+1)H}{w} \right]_{(\xi)} \\ &\times \left[\frac{(\sigma-1)H}{w} + \frac{i(2H^2 + R/6)}{2w^2} \right]_{(0)} \\ &\times \exp\left(-ih \int_0^\xi P_T{}^{1/2} d\eta' \right) \\ &- \tfrac{1}{4}\left[-2i + \frac{(\sigma+1)H}{w} \right]_{(0)} \\ &\times \left[\frac{(\sigma-1)H}{w} + \frac{i(2H^2 + R/6)}{2w^2} \right]_{(\xi)} \\ &\times \exp\left(+ih \int_0^\xi P_T{}^{1/2} d\eta' \right) + O(h^{-3}). \end{aligned} \tag{9.38}$$

We see here that if we use the conditions that diagonalize the Hamiltonian ($\sigma = 0$), it is $|E + Jh^2|^2 \sim O(h^{-2})$, and considering Eq. (9.26) the mean density of created particles is not finite.

If we use instead the conditions of the QEP ($\sigma = 1$) in (9.38), then we obtain:

$$\begin{aligned} E + Jh^2 = &\frac{1}{4}\left[\frac{2H^2 + R/6}{w^2} \right]_{(0)} \exp\left(-ih \int_0^\xi P_T{}^{1/2} d\eta' \right) \\ &- \tfrac{1}{4}\left[\frac{2H^2 + R/6}{w^2} \right]_{(\xi)} \\ &\times \exp\left(+ih \int_0^\xi P_T{}^{1/2} d\eta' \right). \end{aligned} \tag{9.39}$$

We emphasize that in Eq. (9.39) we can guarantee that $E + Jh^2$ has no terms in h^{-1}, but we cannot guarantee the exactness of the terms in h^{-2}, because in the calculus we have neglected terms in H^2. Consequently, we only can say that $|E + Jh^2|^2 \lesssim O(h^{-4})$.

Therefore

$$\begin{aligned} \sum_{ss'} |\beta_{\underline{h}ss'}|^2 &= 2|E|^2 + |E + Jh^2|^2 \\ &\lesssim O(h^{-4}). \end{aligned}$$

With the theory constructed from the QEP, the mean density of created particles in a finite time is finite.

Finally, we are going to give brief consideration to the mean value of the energy density. Suppose that the state of the massive vector field in the time $t = \mathscr{T}$, is the vacuum state $|0\rangle_{\mathscr{T}}$. We are interested in knowing which is going to be the mean value of the expression (2.9) for a time $\mathscr{T}' > \mathscr{T}$. So we have to calculate ${}_{\mathscr{T}}\langle 0|{:}T^0_0(\mathscr{T}', \vec{x}){:}|0\rangle_{\mathscr{T}}$ making use of the following development for the field:

$$\varphi^\mu(\mathscr{T}', \vec{x}) = \int d^3\underline{h} \sum_s \left[a_{\underline{h}s}(\mathscr{T}') \overset{(\mathscr{T}')}{\phi}{}^{\mu}_{\underline{h}s}(\mathscr{T}', \vec{x}) + a^+_{\underline{h}s}(\mathscr{T}') \overset{(\mathscr{T}')}{\phi}{}^{\mu *}_{\underline{h}s}(\mathscr{T}', \vec{x}) \right].$$

Using Eqs. (5.4) and (5.5) to write the operators $a^{(\mathscr{T}')}_{\underline{h}s}$ and $a^+_{\underline{h}s}(\mathscr{T}')$ as functions of the operators $a_{\underline{h}s}(\mathscr{T})$ and $a^+_{\underline{h}s}(\mathscr{T})$ we obtain:

$$\begin{aligned} &{}_{\mathscr{T}}\langle O|{:}T^0{}_0{:}|O\rangle_{\mathscr{T}} \\ &= \sum_{ss's''} \int d^3\underline{h} \left\{ \left[\frac{1}{2a^2} \sum_i \dot{f}_{i\underline{h}s'} \dot{f}_{i\,-\underline{h}s''} \right. \right. \\ &+ \tfrac{1}{2} w^2 f_{0\underline{h}s'} f_{0\underline{h}s''} + \frac{w^2}{2a^2} \sum_j f_{j\underline{h}s'} f_{j\,-\underline{h}s''} \\ &\left. - \frac{i}{a^2} \sum_i h_i \dot{f}_{i\underline{h}s'} f_{0\,-\underline{h}s''} - \frac{1}{2a^4} \sum_{ij} h_i h_j f_{j\underline{h}s'} f_{i\,-\underline{h}s''} \right]_{t=\mathscr{T}'} \\ &\times \sum_s \alpha_{\underline{h}ss'} \beta^*_{\,-\underline{h}ss''} + [\text{comp.conj.}]_{t=\mathscr{T}'} \\ &\times \sum_s \dot{\alpha}_{\,-\underline{h}ss''} \beta_{\underline{h}ss'} + \left[\frac{1}{a^2} \sum_i \dot{f}^*_{i\underline{h}s'} \dot{f}_{i\underline{h}s''} \right. \\ &+ w^2 f^*_{0\underline{h}s'} f_{0\underline{h}s''} + \frac{w^2}{a^2} \sum_j f^*_{j\underline{h}s'} f_{j\underline{h}s''} \\ &+ \frac{1}{a^2} \sum_i i h_i \dot{f}^*_{i\underline{h}s'} f_{0\underline{h}s''} - \frac{1}{a^2} \sum_i i h_i \dot{f}_{i\underline{h}s''} f^*_{0\underline{h}s'} \\ &\left. - \frac{1}{2a^4} \sum_{ij} h_i h_j f^*_{j\underline{h}s'} f_{i\underline{h}s''} - \frac{1}{2a^4} \sum_{ij} f_{j\underline{h}s''} f^*_{i\underline{h}s'} \right]_{t=\mathscr{T}'} \\ &\left. \times \sum_s \beta_{\underline{h}ss'} \beta^*_{\,\underline{h}ss''} \right\}. \end{aligned} \tag{9.40}$$

If in this equation we replace (9.25) and the conditions (7.31) we obtain finally:

$$\begin{aligned} &{}_{\mathscr{T}}\langle 0|{:}T^0{}_0(\mathscr{T}', \vec{x}){:}|0\rangle_{\mathscr{T}} \\ &= \int d^3\underline{h} \left\{ \frac{-iH}{(2\pi a)^3} \left[\frac{h^2}{2w^2a^2} (C + Dh^2) \left(\dot{E} + \dot{J}h^2 \right) \right. \right. \\ &\left. \left. + \frac{m^2}{4w^2} \left(2CE^* + (C + Dh^2) \left(\dot{E} + \dot{J}h^2 \right) \right) \right] \right. \\ &\left. + \frac{w}{(2\pi a)^3} [2|E|^2 + |E + Jh^2|^2] \right\}. \end{aligned} \tag{9.41}$$

We have to notice here that the last term in the integral, as we have already seen, behaves like $O(h^{-3})$, when $h \to \infty$. Nevertheless, in order to prove either the convergence or the divergence of the integral (9.41) it has been necessary to take into account the terms up to H^2 order.

CONCLUSION

The QEP formulated in this way for the case of a vector field, as we did in the case of the scalar field (cf. Refs 1 and 2), leads us to a model of particles for which the Bogolyubov transformation gives us a finite mean density of created particles. As far as the mean value of the component $T^0{}_0$ of the energy-momentum tensor is concerned, we are not able to assure neither its convergence nor its divergence. In spite of the fact that the last term of Eq. (9.41) makes the integral not

convergent, this is a term in H^2 and in our calculus we have not taken this term into consideration. This is going to be considered in a future work for the case of the massless vector field, which is more interesting from a physical point of view.

ACKNOWLEDGMENT

We wish to thank Dr. M. Castagnino for his interesting discussions and comments, and the "Observatorio Astronómico Municipal de Rosario" for its kind hospitality.

[1]M. Castagnino, R. A. Weder, "Quantum Equivalence Principle and Finite Particle Creation in Expanding Universes", Submitted to J. Math. Phys.
[2]M. Castagnino, A. Foussats, R. Laura, and O. Zandron. "The Quantum Equivalence Principle and the Particle Model in Curve Space-time", Nuovo Cimento A (to be published).
[3]A. Lichnerowicz, Inst. Hautes Études. Sci. Publ. Math. 10 (1961).
[4]A. Lichnerowicz, Ann. Inst. Henri Poincaré 233, **3** (1964).
[5]A. Lichnerowicz, Bull. Am. Soc. Math. **92**, 11 (1964).
[6]L. Parker, Phys. Rev. **183**, 1057 (1969).
[7]L. Parker, Phys. Rev. D **3**, 346 (1971).
[8]M. Castagnino, These d'Etat, Université de Paris (1974). To be published by Mathematicae Notae.
[9]M. Castagnino, Gen. Rel. Grav. **9**, 101 (1978).
[10]M. Castagnino, A. Verbeure, and R. A. Weder, Nuovo Cimento **26**, 396 (1975).
[11]In this work the metric tensor has signature -2. Moreover the greek indices $\alpha, \beta, \ldots = 0, 1, 2, 3$ and the latin indices $i, j, \ldots = 1, 2, 3$.
[12]C. J. Isham, *Eight Texas Symposium on Relativistic Astrophysics* edited by M. D. Papagiannis (Boston, 1976), in *Annals of the New York Academy of Sciences*, Vol. 302, (1977), pp. 114–157.
[13]P. Roman, *Introduction to Quantum Field Theory* (Wiley, New York, 1968).
[14]J. A. Schoutten, *Ricci Calculus* (Springer, Berlin, 1954).
[15]F. W. J. Olver, Proc. Cambridge Philos. Soc. **57**, 790 (1961).

Wick polynomials with a mass gap are not infinitely divisible

Klaus Baumann
Institut für Theoretische Physik, der Universität Göttingen, Bunsenstr. 9, D-3400 Göttingen, West Germany

(Received 16 July 1980; accepted for publication 15 August 1980)

We show that Wick polynomials of a generalized free field with a mass gap are not infinitely divisible.

PACS numbers: 11.10.Cd, 02.10.Nj

I. INTRODUCTION

In a paper[1] by Hegerfeldt it was shown that any field (relativistic or Euclidean) can be decomposed into prime fields, which are themselves indecomposable, and an infinitely divisible field. The free and generalized free fields are examples of infinitely divisible fields. Now we want to look into the class of Wick polynomials of generalized free fields to see whether there are other candidates for such fields.

A necessary and sufficient criterion for a field to be infinitely divisible is the conditionally positive definitness of its truncated n-point functions. In this paper we stay within the framework of the Wightman axioms. In the following we want to show that Wick polynomials, for which the underlying generalized free field has a mass gap, are not infinitely divisible. Without the cluster property it is known that there are many such fields. Take for example $\mathscr{W}^T = \tilde{\mathscr{W}} - 1 = (0, \tilde{\mathscr{W}}_1, \tilde{\mathscr{W}}_2, \cdots)$, where $\tilde{\mathscr{W}}$ can be any Wightman functional. It is evident that all assumptions except clustering are fulfilled. Therefore our proof will rely mainly on cluster properties in contrast to the paper[2] where we investigated the case of a Wick square using a special trick.

II. WICK POLYNOMIAL

Let us start with a generalized free field $\phi(x)$.[3] Its 2-point function $W(y-x) := (\Omega, \phi(x)\phi(y)\Omega)$ is given by

$$W(\zeta) = \frac{1}{(2\pi)^{d/2}} \int_{p \in V^+} e^{ip\zeta} \rho(p^2)\, d^d p,$$
$$\zeta \in \tau^+ := \mathbb{R}^d + iV^+,$$

where d is the dimension of space–time and $\rho(p^2)$ denotes the Lehmann spectral function. We assume a mass gap, i.e., $\rho(p^2) = 0$ for $p^2 < M^2$. Lorentz invariance implies $W(\zeta) = \hat{W}(\zeta^2)$. This together with positivity gives the estimates

$$\hat{W}(z) \leqslant \hat{W}(\tfrac{1}{2}(\sigma - |z|)), \quad z = \sigma + i\tau \in \mathbb{C} \setminus [0, \infty)$$

$$\hat{W}(-\sigma) \leqslant \frac{P(\sigma)}{\sigma^\alpha} e^{-M\sqrt{\sigma}}, \quad 0 < \sigma \in \mathbb{R},$$

for some positive constant α and some polynomial P.

A Wick polynomial[4] in its simplest form is an expression like

$$:P(\phi):(x) = \sum_{k=1}^{n} \alpha_k :\phi^k:(x), \quad \alpha_k \in \mathbb{R}, \quad n \geqslant 2.$$

The dots denote Wick ordering. But it is only a minor modification to replace $:\phi^k:(x)$ by $:P_k(\phi,\ldots,\phi):(x)$ where $P_k(X_1,\ldots,X_k)$ is linear in all its arguments and the coefficients are not only numbers but can be finite products of ∂^μ such that $:P_k(\phi,\ldots,\phi):(x)$ is a scalar—e.g., $\Box\phi(x)$, $:\phi^\mu\phi_\mu:(x)$, $:\phi^\mu{}^\nu\phi_\mu\phi_\nu:(x)$ ···. Therefore we call

$$:P(\phi):(x) = \sum_{k=1}^{n} :P_k(\phi,\ldots,\phi):(x) \quad \text{a Wick polynomial, too.}$$

We want to deal only with neutral, scalar fields though we expect the result to be true for other fields, too. But it is much more complicated to write down the positivity condition for general tensor fields.

Now we want to show that no Wick polynomial is infinitely divisible. This is equivalent to showing that the truncated n-point functions $\mathscr{W}_n^T(x_1,\ldots,x_n)$ are not conditionally positive definite. Let us recall that a functional $\mathscr{Y}$ over the Borchers algebra $\mathscr{L}$ is said to be conditionally positive definite iff for all $\underline{f} = (0, f_1, f_2, \ldots) \in \mathscr{L}$ $\mathscr{Y}(\underline{f}^* \otimes \underline{f}) \geqslant 0$.

In the following we consider only the truncated 4-point function $\mathscr{W}_4^T(x_1,\ldots,x_4) = W_4^T(\xi_1,\xi_2,\xi_3), \xi_i := x_{i+1} - x_i$. If $\mathscr{W}_4^T$ is not positive definite then $\mathscr{W}^T$ cannot be conditionally positive definite. $W_4^T(\xi_1,\xi_2,\xi_3)$ is the boundary value of the analytic function $W_4^T(\zeta_1,\zeta_2,\zeta_3), \zeta_i = \xi_i + i\eta_i$, $\eta_i \in V^+$. Lorentz invariance together with locality enlarge the domain of analyticity and imply that $W_4^T(\zeta_1,\zeta_2,\zeta_3)$ is an analytic function of the six independent variables ζ_1^2, $(\zeta_1+\zeta_2)^2$, $(\zeta_1+\zeta_2+\zeta_3)^2$, ζ_2^2, $(\zeta_2+\zeta_3)^2$, and ζ_3^2; i.e., $W_4^T(\zeta_1,\zeta_2,\zeta_3) = \hat{W}_4^T(\zeta_1^2,(\zeta_1+\zeta_2)^2,(\zeta_1+\zeta_2+\zeta_3)^2,\zeta_2^2,(\zeta_2+\zeta_3)^2,\zeta_3^2)$. For Wick polynomials we know the domain of analyticity. It is given by $H^6 : H \otimes \ldots \otimes H$, where $H \subset \mathbb{C}$ denotes the domain of analyticity for the 2-point function $\hat{W}(\zeta^2)$ of the underlying generalized free field. For our purposes it is enough to know that $H \supseteq \mathbb{C} \setminus [0, \infty)$.

III. PROOF

We want to give an indirect proof. Assume $\mathscr{W}_4^T$ to be positive definite. Then the Cauchy–Schwarz inequality tells us that for $z, z', \zeta_1, \zeta_3 \in \tau^+$

$$|\mathscr{W}_4^T(\bar{z} + \bar{\zeta}_1, \bar{z}, z', z' + \zeta_3)|^2$$
$$\leqslant \mathscr{W}_4^T(\bar{z} + \bar{\zeta}_1, \bar{z}, z, z + \zeta_1)\mathscr{W}_4^T(\bar{z}' + \bar{\zeta}_3, \bar{z}', z', z' + \zeta_3),$$

or written in translation invariant manner

$$|W_4^T(-\bar{\xi}_1, \xi_2, \xi_3)|^2$$
$$\leqslant W_4^T(-\bar{\xi}_1, i\eta_2, \xi_1) W_4^T(-\bar{\xi}_3, i\eta_2, \xi_3),$$

with $\zeta_2 = z' - \bar{z} = \xi_2 + i\eta_2 \in \tau^+$, where we have chosen the imaginary parts of z' and z to be equal.

 0022-2488/81/020371-03$1.00

Let us attack the problem in $d = 3$ dimensions first. For $d \geqslant 4$ we put zeros at all the remaining free places of the d-dimensional vectors ζ.

A special choice for the ζ_i, $i = 1,2,3$ is

$$\zeta_1 = \lambda\,(0,1,0) + i\eta_1, \quad \lambda > 0, \eta_1 \in V^+,$$

$$\zeta_2 = \lambda\,\tfrac{1}{2}(\sqrt{2},1,-1) + i(\mu/\lambda)(1,0,0), \quad \mu > 0,$$

$$\zeta_3 = \lambda\,(0,0,1) + i\eta_3, \quad \eta_3 \in V^+.$$

Let us write the above inequality in the invariants

$$\begin{aligned}
&|\hat{W}_4^T((-\bar{\zeta}_1)^2,(-\bar{\zeta}_1+\zeta_2)^2,(-\bar{\zeta}_1+\zeta_2+\zeta_3)^2,\zeta_2^2,(\zeta_2+\zeta_3)^2,\\
&\quad \leqslant \hat{W}_4^T((-\bar{\zeta}_1)^2,(-\bar{\zeta}_1+i\eta_2)^2,(-\bar{\zeta}_1+i\eta_2+\zeta_1)^2,(i\eta_2)^2,(i\eta_2+\zeta_1)^2,\zeta_1^2)\\
&\qquad \times \hat{W}_4^T((-\bar{\zeta}_3)^2,(-\bar{\zeta}_3+i\eta_2)^2,(-\bar{\zeta}_3+i\eta_2+\zeta_3)^2,(i\eta_2)^2,(i\eta_2+\zeta_3)^2,\zeta_3^2).
\end{aligned}$$

By continuity and because the real parts of ζ_1 and ζ_3 are spacelike we can set $\eta_1 = \eta_3 = 0$ as long as $\eta_2 \in V^+$. The values of the invariants are

$$\begin{aligned}
&(-\bar{\zeta}_1)^2 = \zeta_1^2 = (-\bar{\zeta}_3)^2 = \zeta_3^2 = -\lambda^2,\\
&(-\bar{\zeta}_1+\zeta_2)^2 = (-\bar{\zeta}_1+\zeta_2+\zeta_3)^2 = \zeta_2^2 = (\zeta_2+\zeta_3)^2 = i(\sqrt{2})\mu - (\mu/\lambda)^2,\\
&(-\bar{\zeta}_1+i\eta_2)^2 = (i\eta_2+\zeta_1)^2 = (-\bar{\zeta}_3+i\eta_2)^2 = (i\eta_2+\zeta_3)^2 = -\lambda^2 - (\mu/\lambda)^2,\\
&(-\bar{\zeta}_1+i\eta_2+\zeta_1)^2 = (i\eta_2)^2 = (-\bar{\zeta}_3+i\eta_2+\zeta_3)^2 = -(\mu/\lambda)^2,
\end{aligned}$$

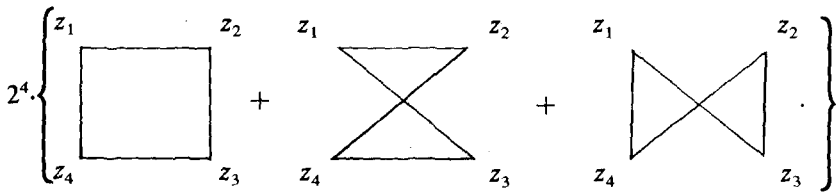
and therefore the following inequality must be fulfilled:

$$\begin{aligned}
&|\hat{W}_4^T(-\lambda^2, i(\sqrt{2})\mu - (\mu/\lambda)^2, i(\sqrt{2})\mu - (\mu/\lambda)^2, i(\sqrt{2})\mu - (\mu/\lambda)^2, i(\sqrt{2})\mu - (\mu/\lambda)^2, -\lambda^2)|\\
&\quad \leqslant \hat{W}_4^T(-\lambda^2, -\lambda^2 - (\mu/\lambda)^2, -(\mu/\lambda)^2, -(\mu/\lambda)^2, -\lambda^2 - (\mu/\lambda)^2, -\lambda^2).
\end{aligned} \tag{1}$$

IV. STRUCTURE OF $W_4^T(\zeta_1,\zeta_2,\zeta_3)$

Now we have to study the structure of $W_4^T(\zeta_1,\zeta_2,\zeta_3)$ or $\mathscr{W}_4^T(z_1,z_2,z_3,z_4)$.

For a moment let us consider a very simple Wick polynomial, namely, $:\phi^2:(x)$. The truncated 4-point function corresponds to a diagram[5] like



Such diagrams are self-explanatory:

$$\overbrace{z_i \qquad z_j} = \begin{cases} W(z_j - z_i) & j > i \\ W(z_i - z_j) & i > j \end{cases}.$$

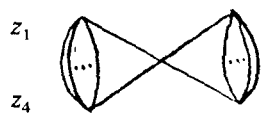
$$\begin{aligned}
&\mathscr{W}_4^T(z_1,z_2,z_3,z_4)\\
&\quad = 16\{W(z_2-z_1)W(z_4-z_1)W(z_3-z_2)W(z_4-z_3)\\
&\qquad + W(z_2-z_1)W(z_3-z_1)W(z_4-z_2)W(z_4-z_3)\\
&\qquad + W(z_3-z_1)W(z_4-z_1)W(z_3-z_2)W(z_4-z_2)\}\\
&\quad = 16\{W(\zeta_1)W(\zeta_1+\zeta_2+\zeta_3)W(\zeta_2)W(\zeta_3)\\
&\qquad + W(\zeta_1)W(\zeta_1+\zeta_2)W(\zeta_2+\zeta_3)W(\zeta_3)\\
&\qquad + W(\zeta_1+\zeta_2)W(\zeta_1+\zeta_2+\zeta_3)W(\zeta_2)W(\zeta_2+\zeta_3)\}.
\end{aligned}$$

The left-hand side of the inequality (1) converges in the limit $\lambda \to \infty$ to $16|\hat{W}(i(\sqrt{2})\mu)|^4$, whereas for the right-hand side we get

$$\begin{aligned}
&\lim_{\lambda\to\infty} 16\{[\hat{W}(-\lambda^2)\hat{W}(-(\mu/\lambda)^2]^2\\
&\qquad + [\hat{W}(-\lambda^2)\hat{W}(-\lambda^2-(\mu/\lambda)^2)]^2\\
&\qquad + [\hat{W}(-\lambda^2-(\mu/\lambda)^2)\hat{W}(-\mu/\lambda)^2)]\} = 0,
\end{aligned}$$

because $\hat{W}(-\lambda^2)$ falls off exponentially and $\hat{W}(-(\mu/\lambda)^2)$ explodes only like an inverse power of $(\mu/\lambda)^2$. So to get a contradiction we only have to adjust $\mu > 0$ such that $\hat{W}(i(\sqrt{2})\mu) \neq 0$. But $\hat{W}$ is an analytic function in $\mathbb{C}\setminus[0,\infty)$ and cannot be identically zero on the half-line $\{i\mu, \mu > 0\}$.

In the case of $:P(\phi):(x) = \Sigma_{k=1}^n \alpha_k :\phi^K:(x)$, $\alpha_n \neq 0$, $n > 2$, the following diagram



$$= W(\zeta_1+\zeta_2)W(\zeta_2+\zeta_3)[W(\zeta_1+\zeta_2+\zeta_3)W(\zeta_2)]^{n-1},$$

is one contribution to the left-hand side of inequality (1) which remains finite in the limit $\lambda \to \infty$. On the other hand the right-hand side converges again to zero, because a term which contains only ζ_2 and $\zeta_1 + \zeta_2 + \zeta_3$ violates the cluster condition for W_4^T and therefore there is at least one factor $W(-\lambda^2)$ which kills any finite power of $\hat{W}(-(\mu/\lambda)^2)$ in the limit $\lambda \to \infty$.

For a general Wick polynomial $:P(\phi):(x) = \Sigma_{k=1}^{n} :P_k(\phi,...,\phi):(x)$ we remind the reader that each $:P_k(\phi,...,\phi):(x)$ can be obtained as a limit of $:P_k(\phi(x_1),...,\phi(x_k)):$—all the singularities between the x_i's are removed by Wick ordering!—where we set $x_1 = \cdots = x_k = x$. For a diagram contributing to W_4^T this means the following:

Instead with one point z, we start with the points $u_1,...,u_k$, $u_i = u_j$ for $i = j$, in the vicinity of z_1. To each point u_i there is attached a line representing a factor $W(y - u_i)$ or $W(u_i - y)$. Because of Wick ordering there is no line between any two u_i's. Now we have to apply some differential operators ∂^μ and contractions to the u_i's prescribed by $:P_k(\phi(u_1),...,\phi(u_k)):$. Finally we put $u_1 = \cdots = u_k = z_1$. For the points z_2, z_3, and z_4 we proceed in an analogous way. The result is that some of the factors $\hat{W}$ have to be replaced by derivatives $\hat{W}^{(1)}$ multiplied with some polynomial in the independent invariants, e.g.,

$$\frac{\partial}{\partial z_{1\mu}} \hat{W}((z_2 - z_1)^2) \frac{\partial}{\partial z_1^\mu} \hat{W}((z_3 - z_1)^2)$$
$$= \hat{W}'(\zeta_1^2) \cdot 2\zeta_1^\mu \cdot \hat{W}'((\zeta_1 + \zeta_2)^2) 2(\zeta_1 + \zeta_2)_\mu$$
$$= \hat{W}'(\zeta_1^2) \hat{W}'((\zeta_1 + \zeta_2)^2) \cdot 2[\zeta_1^2 + (\zeta_1 + \zeta_2)^2 - \zeta_2^2].$$

Of course all the derivatives of $\hat{W}(z)$ obey similar bounds like $\hat{W}(z)$ itself does, i.e., $\hat{W}^{(1)}(-\lambda^2)$ still falls off exponentially and $\hat{W}^{(1)}(-(\mu/\lambda)^2)$ grows atmost like an inverse power. So the conclusion is the same. The left-hand side of inequality (1) for W_4^T has a limit greater than zero, whereas the right-hand side coverges to zero.

V. RESULTS

This result can be extended in many respects.

(a) It is not really necessary to assume a mass gap for $\phi(x)$. All we need is that $\hat{W}(-\lambda^2)$ and all its derivatives fall off stronger than any inverse power for λ going to infinity. Sufficient for this is to assume that $\rho(M)M^{-N}\,dM$ is a bounded measure for all sufficiently large N.

(b) We dealt only with the case of 3-or more dimensional space–time. What happens in two dimensions? By choosing

$$\zeta_1 = \lambda(0,1), \quad \zeta_2 = \frac{\lambda}{2}(1,!) + i\frac{\mu}{\lambda}(1,0), \quad \zeta_3 = i\frac{\mu}{\lambda}(1,0) = i\eta_2,$$

we get the same result. Of course in the inequality

$$|W_4^T(-\bar{\zeta}_1,\zeta_2,\zeta_3)|^2$$
$$\leqslant W_4^T(-\bar{\zeta}_1,i\eta_2,\zeta_1) W_4^T(i\eta_2,i\eta_2,i\eta_2),$$

the second factor of the right-hand side grows like a power of λ but this is killed by the fall off of the first factor. For the left-hand side we can argue as above, because there are terms with remain finite or even grow with λ going to infinity.

(c) If the generalized free field does not fall of fast enough in spacelike directions we were sucessful only in special cases, e.g., for the Wick square we can show that it is not infinitely divisible as long as the growth for small distances is not much stronger than the decrease for large spacelike arguments of the 2-point function.

ACKNOWLEDGMENT

I would like to thank W. D. Garber, G. C. Hegerfeldt, H. Krause, and J. Yngvason for many helpful discussions and comments.

[1]G. C. Hegerfeldt, "Prime Field Decomposition and Infinitely Divisible States on Borchers' Tensor Algebra," Commun. Math. Phys. **45**, 137–151 (1975).

[2]K. Baumann and G. C. Hegerfeldt, "Is the Wick square infinitely divisible?," J. Math. Phys. **18**, 171–172 (1977).

[3]R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, RI, 1965).
R. Streater and A. S. Wightman, PCT, *Spin and Statistics and All That* (New York, Benjamin, 1964).

[4]L. Garding and A. S. Wightman, "Fields as operator-valued distributions in relativistic quantum theory," Ark. Fys. **28**, 129–184 (1965).

[5]C.-D. Mayer, Über den Zusammenhang zwischen *s*-,*p*- und Wickprodukten freier Felder (Diploma thesis, Universität Göttingen, 1977).

Reduction of the $\mathbb{C}P^N$ nonlinear sigma model

H. Eichenherr and J. Honerkamp
Fakultät für Physik, der Universität Freiburg, Hermann-Herder-Str. 3, 7800 Freiburg, West Germany

(Received 20 December 1979; accepted for publication 11 January 1980)

Using the concept of gauge equivalence of two-dimensional classical field theories, we reduce the $\mathbb{C}P^N$ nonlinear σ model to a system of $2N-1$ independent real fields obeying relativistic field equations.

PACS numbers: 11.10.Lm, 11.10.Np

1. INTRODUCTION

Gauge transformations of the linear Lax system of differential equations for an exactly integrable field theoretic model represent a simple and elegant tool to constitute a field-coordinate transformation within a model. They exhibit that various models occurring in the discussion of classical exactly integrable field theories are—in spite of their different physical interpretations—mathematically equivalent, i.e., they differ only in the choice of field coordinates.

If the linear Lax system is written as

$$\begin{aligned}\partial_\xi \phi(\xi,\eta;\zeta) &= U(\xi,\eta;\zeta)\phi(\xi,\eta;\zeta),\\ \partial_\eta \phi(\xi,\eta;\zeta) &= V(\xi,\eta;\zeta)\phi(\xi,\eta;\zeta),\end{aligned} \tag{1}$$

where U and V are meromorphic functions of the parameter $\zeta\in\mathbb{C}$, the gauge transformation is defined by

$$\phi = g\phi';\quad \begin{aligned}U' &= g^{-1}Ug - g^{-1}\partial_\xi g,\\ V' &= g^{-1}Vg - g^{-1}\partial_\eta g.\end{aligned} \tag{2}$$

$\xi = (x^0 + x^1)/2$ and $\eta = (x^0 - x^1)/2$ are the usual light-cone coordinates, ϕ and g take values in a certain Lie group G and U,V take values in the corresponding Lie algebra $\mathcal{G}$. The structure of U,V and U',V', respectively, suggests at times parametrizations, which in general are different, so that one is led in this way to different coordinate systems.

The first clear demonstration of this was given by Zakharov and Takhtadzyan[1] and by D. Chudnovsky and G. Chudnovsky.[2] They pointed out that the nonlinear Schrödinger equation and the Heisenberg equation for a continuous chain of spins in the isotropic case are related by a gauge transformation. Afterwards it was shown,[3] that the same relation holds for the $m\cdot n$ component nonlinear Schrödinger equations and a generalized Heisenberg model for a matrix S, being an element of a Grassmann manifold $\mathrm{U}(m+n)/\mathrm{U}(m)\times\mathrm{U}(n)$ and also for the $\mathrm{O}(n)$ invariant nonlinear σ model and certain generalized sine-Gordon equations.

It is a natural question to ask whether this technique can be further exploited to construct the generalized sine-Gordon models equivalent to the $\mathbb{C}P^N$ σ models.[4] In this paper we will just discuss this question. But instead of gauge transforming the Lax system of the σ models we will demonstrate that by assuming a certain structure of the matrices U,V and a special dependence on the parameter ζ one is lead in a natural way to a class of models which are generalizations of the sine-Gordon model. The gauge transformation of the Lax system of these models then is very simple—the transformation g is identical with $\phi(\zeta=1)$—and leads to the $\mathbb{C}P^N$ σ models if U,V are elements of the Lie algebra of $\mathrm{U}(N+1)$ in a way to be specified in Sec. 2.

2. REDUCED $\mathbb{C}P^N$ σ MODELS

The special dependence of U and V on the parameter ζ will be

$$\begin{aligned}U &= \zeta A + C,\\ V &= \zeta^{-1}B + D.\end{aligned} \tag{3}$$

The consistency condition

$$\partial_\eta U - \partial_\xi V + [U,V] = 0,$$

then decomposes into

$$\partial_\eta A + [A,D] = 0, \tag{4a}$$

$$\partial_\xi B + [B,C] = 0, \tag{4b}$$

$$\partial_\eta C - \partial_\xi D + [A,B] + [C,D] = 0. \tag{4c}$$

Relativistic covariance requires D to be an η component, C to be a ξ component of a Lorentz two-vector. We now consider a decomposition of the Lie algebra $\mathcal{G}$ of the group $\mathrm{U}(N+1)$ into the Lie algebra $\mathcal{L}$ of the subgroup $\mathrm{U}(1)\times\mathrm{U}(N)$ and its orthogonal complement k with respect to the Killing form of $\mathrm{U}(N+1)$:

$$\mathcal{G} = \mathcal{L}\oplus k.$$

Note that

$$\begin{aligned}&[\mathcal{L},\mathcal{L}]\subset\mathcal{L},\\ &[\mathcal{L},k]\subset k,\\ &[k,k]\subset\mathcal{L}.\end{aligned} \tag{5}$$

Let us assume that $A,B\in k$ then it follows from (4) and (5) that $C,D\in\mathcal{L}$. Because of the very structure of $\mathcal{L}$ and k we represent A,B and C,D by complex block-off-diagonal and block-diagonal matrices respectively, in the following way:

$$A = \begin{pmatrix} 0 & \bar\psi_2\cdots\bar\psi_{N+1}\\ -\psi_2 & \\ \vdots & 0\\ -\psi_{N+1} & \end{pmatrix},$$

$$B = \begin{pmatrix} 0 & \bar\varphi_2\cdots\bar\varphi_{N+1}\\ -\varphi_2 & \\ \vdots & 0\\ -\varphi_{N+1} & \end{pmatrix}, \tag{6a}$$

 0022-2488/81/020374-03$1.00

$$C=\begin{pmatrix} -\operatorname{tr}X & 0 \\ 0 & X \end{pmatrix}, \quad X^{\dagger}=-X, \quad D \text{ analogously.} \tag{6b}$$

Note that also from our final intention it is natural to choose the following starting point:

If $Z\in\mathbb{C}^{N+1}$ is the usual $\mathbb{C}P^N$ field and Y a complex $N\times(N+1)$ matrix such that $g=(Z,Y)$ is a moving orthonormal frame of $\mathbb{C}^{N+1}$, then

$$A=g^{\dagger}D_{\xi}\,g=\begin{pmatrix} 0 & Z^{\dagger}\partial_{\xi}Y \\ Y^{\dagger}\partial_{\xi}Z & 0 \end{pmatrix},$$

$$B=g^{\dagger}D_{\eta}\,g=\begin{pmatrix} 0 & Z^{\dagger}\partial_{\eta}Y \\ Y^{\dagger}\partial_{\eta}Z & 0 \end{pmatrix},$$

$$C=A_{\xi}, \quad D=A_{\eta},$$

with

$$A_{\xi(\eta)}=\begin{pmatrix} Z^{\dagger}\partial_{\xi(\eta)}Z & 0 \\ 0 & Y^{\dagger}\partial_{\xi(\eta)}Y \end{pmatrix},$$

$$D_{\xi(\eta)}g=\partial_{\xi(\eta)}g-gA_{\xi(\eta)}$$

(see Ref. 5).

Since in the $\mathbb{C}P^N$ model we may normalize the coordinates such that

$$-\tfrac{1}{2}\operatorname{tr}(g^{\dagger}D_{\xi}\,g)^2=-\tfrac{1}{2}\operatorname{tr}(g^{\dagger}D_{\eta}\,g)^2=1,$$

we require in addition

$$\sum_{j=2}^{N+1}|\varphi_j|^2=\sum_{j=2}^{N+1}|\psi_j|^2=1. \tag{6c}$$

Proposition 1: By a gauge transformation $h(\xi,\eta)\in\mathrm{U}(1)\times\mathrm{U}(N)$ one may reduce (3) to a standard form

$$U=\zeta A+C, \tag{7a}$$

$$V=\zeta^{-1}B, \tag{7b}$$

with

$$A=\begin{pmatrix} 0 & \cdots & 1 \\ \vdots & & \\ 0 & 0 & \\ -1 & & \end{pmatrix}, \quad B \text{ as in (6a)}, \tag{7c}$$

$$C=\begin{pmatrix} c_1 & 0\cdots 0 & 0 \\ 0 & & c_2 \\ \vdots & 0 & \vdots \\ 0 & & c_N \\ 0 & -\bar{c}_2\cdots-\bar{c}_N & -c_1 \end{pmatrix}, \tag{7d}$$

where

$$c_1=\frac{\varphi_{N+1}\partial_{\xi}\bar{\varphi}_{N+1}+\sum_{k=2}^{N}\partial_{\xi}\varphi_k\cdot\bar{\varphi}_k}{3|\varphi_{N+1}|^2-1}, \tag{7e}$$

$$c_k=\frac{\partial_{\xi}\varphi_k}{\varphi_{N+1}}+\frac{\varphi_k}{\varphi_{N+1}}c_1. \tag{7f}$$

Note that indeed $c_1^{\dagger}=-c_1$.

Proof: The first step towards such a standard form is a gauge transformation $h_1(\xi,\eta)\in\mathrm{U}(1)\times\mathrm{U}(N)$ by which D is reduced to zero:

$$U'=\zeta h_1^{-1}Ah_1+h_1^{-1}Ch_1-h_1^{-1}\partial_{\xi}h_1,$$

$$V'=\zeta^{-1}h_1^{-1}Bh_1+h_1^{-1}Dh_1-h_1^{-1}\partial_{\eta}h_1. \tag{8}$$

Choosing $h_1(\xi,\eta)$ so that

$$\partial_{\eta}h_1=Dh_1, \tag{9}$$

leads to

$$V'=\zeta^{-1}B^1, \tag{10}$$

and (4a) and (4c), are reduced to

$$\partial_{\eta}A'=0, \quad \text{i.e.,} \quad A'_1=A'_1(\xi), \tag{11}$$

$$\partial_{\eta}C'=[A',B']=0. \tag{12}$$

We are yet free to apply a gauge transformation $h_2(\xi)\in\mathrm{U}(1)\times\mathrm{U}(N)$ which rotates the ξ-dependent unit vector of A'.

$$\begin{pmatrix} 0 \\ \psi_2 \\ \vdots \\ \psi_{N+1} \end{pmatrix} \text{ into } \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix},$$

leaving the condition $D=0$ invariant. $h_2(\xi)$ is determined up to transformations $h_3(\xi)\in\mathrm{U}(1)\times\mathrm{U}(N-1)$ which leave

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

invariant, and by inspection of Eq. (12) we see that $h_3(\xi)$ can be used finally to reach the structure of C indicated in (7d). So $h(\xi,\eta)=h_1(\xi,\eta)h_2(\xi)h_3(\xi)$ transforms (3) into (7a)–(7d). Inserting B,C into Eq. (4b) and solving for the c_j, we arrive at (7e), and (7f). Thus Eqs (4a) and (4b) have been solved, and the remaining Eqs. (4c) are the field equations for the φ_j:

$$\partial_{\eta}c_1+\bar{\varphi}_{N+1}-\varphi_{N+1}=0,$$

$$\partial_{\eta}c_k+\varphi_k=0, \quad k=2\cdots N. \tag{13}$$

They involve $2N-1$ independent real scalar fields.

3. THE GAUGE TRANSFORMATION

Proposition 2: The model defined by the linear system

$$\partial_{\xi}\phi=U\phi,$$

$$\partial_{\eta}\phi=V\phi,$$

where U,V are given by (7) is gauge equivalent to the $\mathbb{C}P^N$ model.

Proof: We choose $g=\phi\,(\zeta=1)\in\mathrm{U}(N+1)$, then

$$\partial_{\xi}\,g=(A+C)g,$$

$$\partial_{\eta}\,g=Bg,$$

and

$$U'=g^{-1}(\zeta A+C)g-g^{-1}(A+C)g$$
$$=(\zeta-1)g^{-1}Ag,$$

$$V'=\zeta^{-1}g^{-1}Bg-g^{-1}Bg$$
$$=(\zeta^{-1}-1)g^{-1}Bg.$$

Let us define

$$S = g^{-1}Pg \quad \text{where} \quad P = \begin{pmatrix} +1 & & & 0 \\ & -1 & & \\ & & \ddots & \\ 0 & & & -1 \end{pmatrix}.$$

As g runs through $U(N)$, S runs through $\mathbb{C}P^N$. Then

$$\begin{aligned} S\partial_\xi S &= g^{-1}Pg(\partial_\xi(g^{-1})Pg + g^{-1}P\partial_\xi g) \\ &= g^{-1}P(-(A+C)Pg + g^{-1}\partial_\xi g). \end{aligned}$$

Because of $PCP = C$, $PAP = -A$, we obtain

$$S\partial_\xi S = 2g^{-1}Ag,$$

and similarly

$$S\partial_\eta S = 2g^{-1}Bg.$$

On the other hand we have $S^2 = 1$ and therefore $S\cdot\partial_\xi S = -\partial_\xi S\cdot S$, hence

$$\begin{aligned} U' &= \tfrac{1}{4}(\zeta - 1)[S, \partial_\xi S], \\ V' &= \tfrac{1}{4}(\zeta^{-1} - 1)[S, \partial_\eta S]. \end{aligned} \tag{14}$$

U', V' are just the operators of the linear system of the $\mathbb{C}P^N$ model, i.e., of the dual symmetry, and the integrability condition for (14) is the field equation of the $\mathbb{C}P^N$ model.

4. EXAMPLES

(i) $N = 1$. The $\mathbb{C}P^1$ model is the O(3) σ model, the equivalent linear system (7) is then

$$U = \zeta\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \begin{pmatrix} c_1 & 0 \\ 0 & -c_1 \end{pmatrix},$$

$$V = \zeta^{-1}\begin{pmatrix} 0 & \bar{\varphi}_2 \\ -\varphi_2 & 0 \end{pmatrix},$$

where $c_1 = \varphi_2\partial_\xi\bar{\varphi}_2/2|\varphi_2|^2$. With $\varphi_2 = e^{-i\alpha}$ we obtain $c_1 = (i/2)\partial_\xi\alpha$. Hence

$$U = \zeta\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \frac{i}{2}\begin{pmatrix} \partial_\xi\alpha & 0 \\ 0 & -\partial_\xi\alpha \end{pmatrix},$$

$$V = \zeta^{-1}\begin{pmatrix} 0 & e^{i\alpha} \\ -e^{-i\alpha} & 0 \end{pmatrix}.$$

This is again the linear system of the sine-Gordon model—up to a rescaling of ξ and η.

(ii) The reduced $\mathbb{C}P^2$ system can be parametrized in terms of three independent real scalar fields α, ϕ, ψ:

$$B = \begin{pmatrix} 0 & \sin\alpha e^{-i\psi} & \cos\alpha e^{-i\phi} \\ -\sin\alpha e^{i\psi} & 0 & 0 \\ -\cos\alpha e^{i\phi} & 0 & 0 \end{pmatrix},$$

$$C = \begin{pmatrix} c_1 & 0 & 0 \\ 0 & 0 & c_2 \\ 0 & -\bar{c}_2 & -c_1 \end{pmatrix}.$$

It turns out to be convenient to introduce a new variable β by

$$c_1 = \frac{\varphi_3\partial_\xi\bar{\varphi}_3 + \bar{\varphi}_2\partial_\xi\varphi_2}{3|\varphi_3|^2 - 1} =: \frac{i}{2}\partial_\xi\beta.$$

Then

$$c_2 = e^{i(\psi-\phi)}(\partial_\xi\alpha + i\cot\alpha\partial_\xi(\beta+\phi)),$$

and ψ is determined by α, β, ϕ by the equations

$$\partial_\xi\psi = -\tfrac{1}{2}\partial_\xi\beta + \cot^2\alpha\partial_\xi(\beta+\phi),$$

$$\partial_\eta\psi = -\frac{1}{\sin^2\alpha}(\partial_\eta\beta + \partial_\eta\phi\cos^2\alpha).$$

α, β, and ϕ solve the hyberbolic field equations

$$\Box\alpha + \frac{\cos\alpha}{\sin^3\alpha}\partial_\xi(\beta+\phi)\partial_\eta(\beta+\phi) + \sin\alpha\cos\phi = 0,$$

$$\Box\beta - 4\cos\alpha\sin\phi = 0,$$

$$\Box\phi - \frac{1}{\sin\alpha\cos\alpha}\{\partial_\eta\alpha\partial_\xi(\beta+\phi) + \partial_\xi\alpha\partial_\eta(\beta+\phi)\} + \frac{\sin\phi}{\cos\alpha}(1+3\cos^2\alpha) = 0,$$

which can be derived from the Lagrangian density

$$\mathcal{L} = \tfrac{1}{2}[\partial_\mu\alpha\partial^\mu\alpha + \tfrac{1}{4}\partial_\mu\beta\partial^\mu\beta + \cot^2\alpha\partial_\mu(\beta+\phi)\times\partial^\mu(\beta+\phi)] + \cos\alpha\cos\phi - 1.$$

We finally remark that α, β, and ϕ are related to the $\mathbb{C}P^2$ field Z in normal coordinates by

$$\overline{(D_\xi Z\cdot D_\eta Z)} = \cos\alpha e^{i\phi},$$

$$i\partial_\xi\beta = \overline{(D_\xi Z\cdot D_\xi D_\xi Z)}, \quad -i\partial_\eta\beta = \overline{(D_\eta Z\cdot D_\eta D_\eta Z)}.$$

For $\beta = \phi = 0$, the above system reduces to the sine-Gordon equation with the Lax pair written in the spin 1 representation corresponding to the O(3) subgroup of U(3) generated by the Gell-Mann matrices $\lambda_2, \lambda_5, \lambda_7$.

[1] V.E. Zakharov and L.A. Takhtadzhyan, Teor. Mat. Fiz. **38**, 26 (1979).
[2] D.V. Chudnovsky and G.V. Chudnovsky, Preprint Saclay DPh-P/79, p. 110.
[3] J. Honerkamp, J. Math. Phys. **22**, 277 (1981).
[4] H. Eichenherr, Nucl. Phys. B **146**, 215 (1978); see also Erratum, Nucl. Phys. B **155**, 544 (1979); A. D'Adda, M. Lüscher, and P. Di Vecchia, Nucl. Phys. B **146**, 63 (1978); V.L. Golo and A.M. Perelomov, Phys. Lett. B **79**, 112 (1978); E. Witten, Nucl. Phys. B **149**, 285 (1979).
[5] H. Eichenherr and M. Forger, Nucl. Phys. B **155**, 381 (1979).

Quantum tachyons in Schwarzschild space–time

B. Talukdar and M. Sen
Department of Physics, Visva-Bharati University, Santiniketan 731235, West Bengal, India

D. Sen
Department of Physics, Burdwan University, Burdwan, West Bengal, India

(Received 30 May 1979; accepted for publication 11 January 1980)

The wave equation of a spinless tachyon is studied in Schwarzchild space–time. In contrast to earlier approaches to the problem, it is shown that tachyonic static solutions satisfy a simple second-order linear differential equation regardless of the mass of the black hole and the mass parameter of the tachyon. Physical implication of the present approach is discussed. Using Langer modification of the WKB (Wentzel-Kramers-Brillouin) boundary condition an expression similar to the Bohr-Sommerfeld quantization condition is derived.

PACS numbers: 11.10.Qr, 14.80.Kx

Following a line of research initiated by Barashenkov,[1] by Davies,[2] by Honig *et al.*,[3] and by Narlikar and Sudarshan,[4] Dhurandhar[5] in a recent publication contemplates analyzing tachyonic scalar waves in Schwarzschild space–time by using a semiclassical treatment. Here the tachyons are regarded as quantum wave packets, but the gravitational field is unquantized. Thus the tachyonic wave function ψ satisfies the Klein–Gordon equation in free space. Expanding ψ into partial waves and separating out its time dependence by means of Fourier analysis, one arrives at a differential equation which does not permit simple analytical solutions. To deal with the situation Dhurandhar[5] makes specific assumptions with regard to the mass of the black hole and mass parameter of the tachyon, obtains a relatively simple equation, and constructs solutions by the WKB method,[6] which are valid everywhere in space, save certain exceptional regions. This treatment provides, in a natural way, physical information about interaction of quantum tachyons with a potential barrier produced by a Schwarzschild black hole. As is typical of the WKB approximation, the Airy function is exploited to make a smooth transition between the oscillatory and exponential domains which lie on the negative and positive energy sides of the classical turning point.

The object of the present note is two-fold.

(i) To obtain the soluble equation of Dhurandhar by a strict mathematical procedure and thereby avoid the mass restriction imposed in Ref. 5.

(ii) To obtain the WKB eigenvalue formula via the Bohr–Sommerfeld quantization condition for tachyonic scalar waves in the background of Schwarzschild space–time. We also present astrophysical implications of (i) and (ii) wherever possible.

The tachyonic wave function $\psi(R,\theta,\varphi,t)$ in empty space statisfies the Klein–Gordon equation

$$(\Box^2 - M_0^2)\psi(R,\theta,\varphi,t) = 0\,, \tag{1}$$

where M_0 is the mass parameter of the tachyon and the operator

$$\Box^2 = \frac{1}{(-g)^{1/2}}\frac{\partial}{\partial x^j}\left((-g)^{1/2}g^{ij}\frac{\partial}{\partial x^i}\right). \tag{2}$$

Here g^{ij} is the metric tensor for Schwarzschild space–time and $-g = \det g^{ij}$. The components of g_{ij} are

$$g_{00} = \left(1 - \frac{2MG}{R}\right),\quad g_{11} = -\left(1 - \frac{2MG}{R}\right)^{-1},$$
$$g_{22} = -R^2,\quad \text{and}\quad g_{33} = -R^2\sin^2\theta\,, \tag{3}$$

M being the mass of the black hole and G the gravitational constant.

With the metric (3), the scalar field in Eq. (1) can be split into partial waves and time-dependence separated. To that end one uses the partial wave and Fourier decompositions

$$\psi(R,\theta,\varphi,t) = \sum_{l=0}^{\infty}\frac{\psi_l(R,t)}{R}P_l(\cos\theta)\,, \tag{4a}$$

and

$$\psi_l(R,t) = \int_{-\infty}^{\infty} A(\omega)\psi_l^{\Omega}(R)e^{-i\Omega t}\,d\Omega\,, \tag{4b}$$

and thus arrives at

$$\frac{d^2\psi_l^{\Omega}}{dr^2} + \frac{2m}{r(r-2m)}\frac{d\psi_l^{\Omega}}{dr} + \frac{r^2}{(r-2m)^2}\left(\omega^2 + \frac{r-2m}{r} - \frac{2m(r-2m)}{r^4} - \frac{l(l+1)}{r^3}(r-2m)\right)\psi_l^{\Omega} = 0\,, \tag{5}$$

with

$$r = RM_0,\quad m = GMM_0,\quad \text{and}\quad \omega = \Omega/M_0\,. \tag{6}$$

Note that because of azimuthal symmetry of the problem, the function ψ is independent of φ. It is clear from Eq. (4b) that the square of Ω corresponds to the energy associated with a particular partial wave l.

Dhurandhar[5] observes that Eq. (5) is fairly complicated but it assumes a relatively simple form if the mass of the black hole is of the order of the mass of the sun and the tachyon mass parameter equals the mass of an electron. We point out that such an assumption is not necessary to reduce Eq. (5) to a tractable form. To see that we transform the dependent variable by substituting

$$\psi_l^{\Omega}(r) = \Phi_l^{\Omega}(r)\exp\left(-m\int^r\frac{dx}{x(x-2m)}\right) \tag{7a}$$

 0022-2488/81/020377-03$1.00

$$= \left(\frac{r}{r-2m}\right)^{1/2} \Phi_l^{\Omega}(r)\,. \tag{7b}$$

This yields

$$\frac{d^2\Phi_l^{\Omega}}{dr^2} + \frac{r^2}{(r-2m)^2}\left(\omega^2 + 1 - \frac{2m}{r} - \frac{l(l+1)}{r^3} \times (r-2m) + \frac{m^2}{r^4}\right)\Phi_l^{\Omega}\,. \tag{8}$$

The term m^2/r^4 inside the large parentheses represents a singular region centered about $r = 0$. In place of this term Dhurandhar found $-2m(r-2m)/r^4$. Thus the region of coordinate singularity appears to be over estimated by a factor of four in the approximate treatment made in Ref. 5. Anyway, if we look for solutions of Eq. (8) outside the small sphere centered about $r = 0$, we can drop the term m^2/r^4 and arrive at the desired equation. For the s-wave case this equation reads

$$\frac{d^2\Phi}{dr^2} + \frac{r^2}{(r-2m)^2}\left(k^2 - \frac{2m}{r}\right)\Phi = 0\,, \tag{9}$$

where $k^2 = \omega^2 + 1$. We have, for brevity, omitted the superscript Ω and subscript $l = 0$. At very high energies the s-wave is expected to sample the region of space near the singularity. The term m^2/r^4 will no longer be negligible. Thus at distances close to $r = 0$ one will have to deal with an equation

$$\frac{d^2\Phi}{dr^2} + \frac{r^2}{4m^2}\left[k^2 - v^2(r)\right]\Phi = 0\,, \tag{10}$$

where

$$v^2(r) = \frac{2m}{r} - \frac{m^2}{r^4}\,. \tag{11}$$

Equation (10) is identical with Eq. (25) of Ref. 5. The reason for this coincidence is fairly straightforward. In obtaining Eq. (10) from the behavior of Eq. (6) near $r = 0$, Dhurandhar has chosen to work with the right kind of transformation, namely,

$$\psi = r^{1/2}\Phi\,. \tag{12}$$

The transformation (12) is a special instance of (7a). In particular, when r is very small one can replace $x(x-2m)$ by $-2mx$ to obtain Eq. (12) from Eq. (7a).

Looking at Eq. (7b) we see that the partial wave solution of Eq. (1) breaks down at $r = 2m$, the event horizon for the Schwarzschild black hole. In this region, curvature effects come into play. In Ref. 5, WKB method has been used to examine the nature of the solution here. The point $r_1 = 2m/k^2$ is a classical turning point of Eq. (9). Therefore, the situation in the vicinity of $r = r_1$ deserves to be more closely examined. Since Dhurandhar has treated this point in some detail, we pose a slightly different problem.

It is well known that as the particles of real mass cross the Schwarzschild radius, they cannot come out. In contrast to this the gravitational field of the black hole tends to oppose the infall of tachyons. A classical tachyon is bounced by the black hole and emerges from inside the event horizon. Besides this, a quantum tachyon also tunnels through the potential barrier and hits the singularity.[5] In the jargon of the trade, could one visualize the possibility of a black hole producing and ejecting tachyons? It may be that the peculiar space–time effect inside the Schwarzschild radius generates tachyons, which may form bound or quasibound states and eventually come out of the Schwarzschild barrier due to any small perturbation whatsoever. Thus any experiment which attempts to detect tachyons should be directed towards black holes rather than a laboratory set up. It will, therefore, be of some general interest to formulate the eigenvalue problem for the interaction of a quantum tachyon with Schwarzschild space–time within the framework of the WKB approximation.

At the classical turning point r_t in a repulsive potential field, like ours, the WKB function has a singular amplitude. This prevents formulation of the boundary condition. The difficulty can, however, be circumvented by using the well-known work of Langer.[7] The method consists in replacing the WKB differential equation by another differential equation which (i) agrees with the Schrödinger equation near the classical turning point and (ii) agrees with the WKB differential equation elsewhere. Before we proceed to use the basic philosopy of Langer's work, it will be worthwhile to look closely into Dhurandhar's formulation of the boundary condition.

Equation (8) (with m^2/r^4 deleted) exhibits that the l-wave problem has two classical turning points determined by the positive real roots of the equation

$$\omega^2 - v_l^2(r) = 0\,, \tag{13}$$

with

$$v_l^2(r) = \left(1 - \frac{2m}{r}\right)\left(\frac{l(l+1)}{r^2} - 1\right). \tag{14}$$

It is clear from Eqs. (13) and (14) that unlike the l-wave case, the s-wave problem has only one turning point. Thus for $l > 0$ one can derive an expression similar to the Bohr–Sommerfeld quantization condition. The s-wave case cannot be treated similarly. It appears that Dhurandhar has overlooked this point.

To treat the lowest partial wave on equal footing with the higher ones we replace $l(l+1)$ by $(l+\frac{1}{2})^2$ in Eq. (14). Such a replacement is consistent with the Langer modification of the WKB boundary condition.[8] The turning point r_1 and r_2 are now determined from

$$\omega^2 - v_l'^2(r) = 0, \tag{15}$$

with

$$v_l'^2(r) = \left(1 - \frac{2m}{r}\right)\left(\frac{(l+\frac{1}{2})^2}{r^2} - 1\right). \tag{16}$$

Clearly, the s-wave problem has also two classical turning points.

If Φ_l^{Ω} is to be bounded for $r < r_1$, then in $r_1 < r < r_2$

$$\Phi_l^{\Omega}(r) = Q^{-1/2}(r)\cos\left(\int_{r_1}^{r} dr\, Q(r) - \frac{\pi}{4}\right). \tag{17}$$

Similarly if $\Phi_l^{\Omega}(r)$ is to be bounded for $r < r_2$, then in $r_1 < r < r_2$

$$\Phi_l^{\Omega}(r) = Q^{-1/2}(r)\cos\left(\int_{r_2}^{r} dr\, Q(r) + \frac{\pi}{4}\right). \tag{18}$$

The quantity $Q(r)$ in Eqs. (17) and (18) determined from Eqs. (8), (15), and (16) is given by

$$Q^2(r) = \frac{r^2}{(r-2m)^2}\left[\omega^2 - v_l'^2(r)\right]. \tag{19}$$

The two expression for Φ_l^{Ω} given above must be the same. Thus

$$\int_{r_1}^{r_2} dr\, Q(r) = (n+\tfrac{1}{2})\pi. \tag{20}$$

The result is very similar to the Bohr–Sommerfeld quantization condition. Derivation of Eqs. (17)–(20) is fairly straightforward. Reference 7 will be useful in working out the details. In the general case one cannot use simple analytical methods to determine the eigenvalues via Eq. (20) since neither the solution of Eq. (15) nor the integral $\int_{r_1}^{r_2} dr\, Q(r)$ can be obtained in closed form. However, it is possible to treat a special case for low frequency waves, which are expected to lie at very large distances from the singularity. With this approximation, Eq. (15) yields the following expressions for the classical turning points.

$$r_1 = m - [m^2 + (l+\tfrac{1}{2})^2(\omega^2+1)]^{1/2}, \tag{21a}$$

and

$$r_2 = m + [m^2 + (l+\tfrac{1}{2})^2(\omega^2+1)]^{1/2}. \tag{21b}$$

In this case

$$Q = [(r-r_1)(r_2-r)]^{1/2} \tag{22}$$

Thus the determination of energy eigenvalues for low frequency waves by combining Eqs. (20)–(22) is a simple problem of elementary calculus.

One of the authors (MS) acknowledges the receipt of a junior research fellowship of the Indian Space Research Organization, Government of India.

[1]V.S. Barashenkov, Usp. Fiz. Nauk **114**,133 (1974).
[2]P.C.W. Davies, Nuovo Cimento B **25**, 571 (1975).
[3]E. Honig, K. Lake, and R.C. Roeder, Phys. Rev. D **10**, 3155 (1974).
[4]J.V. Narlikar and E.C.G. Sudarshan, Mon. Not. R. Astron. Soc. **175**, 105 (1976).
[5]S.V. Dhurandhar, J. Math. Phys. **19**, 5611 (1978).
[6]J. Mathews and R.L. Walker, *Mathematical Method of Physics*, 2nd ed. (Benjamin, New York, 1970).
[7]R.E. Langer, Phys. Rev. **51**, 669 (1937).
[8]N.F. Mott and H.S.W. Massey, *The Theory of Atomic Collisions*, 2nd ed. (Oxford U.P., New York, 1949).

A new expression for the high-energy upper bound on the scattering amplitude

Masud Chaichian
Research Institute for Theoretical Physics, University Of Helsinki, Finland

Jan Fischer
Institute of Physics, Czechoslovak Academy of Sciences, Prague, Czechoslovakia

S. M. Molodensky and N. F. Nelipa
Institute of Nuclear Research, Moscow State University, Moscow, USSR

Ivo Vrkoč
Mathematical Institute, Czechoslovak Academy of Sciences, Prague, Czechoslovakia

(Received 8 March 1979; accepted for publication 11 December 1979)

The two Froissart–Martin high-energy upper bounds for forward and nonforward scattering are combined into one formula under the additional assumption that the scattering amplitude is polynomially bounded in energy for all scattering angles inside the Lehmann–Martin ellipse. The method used presents a modification of that of Kinoshita, Loeffel, and Martin. The analogous bound for the scattering of particles with spin is obtained as well. Using the same method, a bound for the case of complex scattering angles is also derived and ways leading to its improvement by using the solution of the Dirichlet problem are suggested.

PACS numbers: 11.20.Dj

1. INTRODUCTION

Froissart[1] was the first to obtain high-energy bounds on the scattering amplitude from the requirements of unitarity, polynomial boundeness, and the Mandelstam representation. The same results were rederived later by Martin,[2] who used the analytic properties of the scattering amplitude which follow from axiomatic field theory instead of the Mandelstam representation. The bounds were found separately for the forward (and backward) direction and for all other scattering angles:

$$|f(s,\cos\theta)| < C_1 s \ln^2 s, \quad \text{for } \theta = 0 \text{ or } \pi, \tag{1}$$

$$|f(s,\cos\theta)| < C_2 \frac{s^{3/4}\ln^{3/2} s}{|\sin\theta|^{1/2}}, \quad \text{for } \theta \neq 0 \text{ and } \pi. \tag{2}$$

These two bounds cover all physical values of the scattering angle θ, but they do not transform into one another for θ approaching 0 or π. They also say nothing about $f(s,\cos\theta)$ outside the physical interval $-1 \leqslant \cos\theta \leqslant 1$.

The aim of the present paper is to extend (1) and (2) to unphysical angles and also to the case of particles with spin. The method used here presents a generalization of that developed by Kinoshita, Loeffel, and Martin in order to derive an upper bound on the scattering amplitude from the Mandelstam representation.[3,4] In Sec. 2, following the approach of Ref. 3, we derive some useful mathematical relations and estimates. [In this approach,it is additionally assumed that the scattering amplitude $f(s,z)$ is bounded by a power s^N (polynomial boundedness) in s for all energies above a certain value at all z inside the Lehmann–Martin ellipse.] In particular, we find that the function $g(s,w)$ defined below by formula (4) is bounded by the power $s^{N+1/4}$ at large energies, thereby improving the bound s^{N+3} of Refs. 3 and 4. Then, In Sec. 3, we derive a bound combining (1) and (2), i.e., all physical angles, into one formula, Eq. (25). Let us remark in this connection that there is, of course, a much simpler formula which combines (1) and (2) (in a continuous manner), namely,

$$|f(s,\cos\theta)| < \inf\left(C_1 s \ln^2 s, \quad C_2 \frac{s^{3/4}\ln^{3/2} s}{|\sin\theta|^{1/2}}\right).$$

But the method used in Sec. 3 to combine (1) and (2), later allows us to obtain a bound on the amplitude at unphysical angles (Sec. 4). The generalization of the results of Sec. 3 to the case of particles with spin is given in Sec. 5. A further improvement of the former bound using the solution of the Dirichlet problem is in progress.

2. GENERAL RELATIONS

The amplitude $f(s,z)$ for the scattering of two spinless particles can be expanded into the Legendre series

$$f(s,z) = \frac{\sqrt{s}}{2k}\sum_{l=0}^{\infty}(2l+1)a_l(s)P_l(z), \tag{3}$$

where s is the total energy squared, $z = \cos\theta$, θ is the scattering angle, and k is the momentum of the particle, all in the center-of-mass system.

Simultaneously, consider the auxiliary function $g(s,w)$ defined by

$$g(s,w) = \frac{\sqrt{s}}{2k}\sum_{l=0}^{\infty}(2l+1)a_l(s)w^l. \tag{4}$$

The series (3) converges inside the ellipse E_ρ with the semimajor axis ρ equal to

$$\rho = 1 + a/s, \tag{5}$$

and with the foci at the points $z = \pm 1$, a being a positive constant, which depends on the process considered. This implies the following relation for the expansion coefficients $a_l(s)$:

 0022-2488/81/020380-05$1.00

$$\frac{\sqrt{s}}{2k}(2l+1)a_l(s) = \frac{2l+1}{2\pi i}\int_{E_\rho} Q_l(\mu) f(s,\mu)\, d\mu,$$

where the integral is taken along the ellipse E_ρ (cf. Ref. 5, Sec. 39). Denoting by $M(s)$ the maximum of $|f(s,\mu)|$ on the ellipse, we have

$$\frac{\sqrt{s}}{2k}|a_l(s)| \leqslant \frac{M(s)}{2\pi}\int_{E_\rho} |Q_l(\mu)|\,|d\mu|\,.$$

The Legendre function $Q_l(\mu)$ can be estimated as follows (cf. Ref. 5, Sec. 38):

$$|Q_l(\mu)| \leqslant (\pi/l)^{1/2} R^{-(l+1)}(1-R^{-2})^{-1/2},$$

here $R = \rho + (\rho^2-1)^{1/2}$. Thus,

$$|a_l(s)| \leqslant \frac{k}{\sqrt{s}} M(s)L(s)\frac{1}{(\pi l)^{1/2}} R^{-(l+1)}(1-R^{-2})^{-1/2}, \quad (6)$$

$$|a_0(s)| \leqslant 1,$$

where $L(s)$ is the length of the ellipse. Using the expression $\lim_{l\to\infty} \inf |a_l(s)|^{-1/l}$ for the convergence radius $\tilde{R}$ of (4), we obtain that

$$\tilde{R} \geqslant R = \rho + (\rho^2-1)^{1/2}. \quad (7)$$

We now use the following well-known representation of the Legendre polynomials

$$P_l(z) = \frac{1}{\pi i}\int_\Gamma \frac{w^l}{(w^2-2wz+1)^{1/2}}\, dw, \quad (8)$$

where Γ is a curve connecting the points $z-(z^2-1)^{1/2}$ and $z+(z^2-1)^{1/2}$ (see Ref. 5, Sec. 156). We choose z so that the points $z \pm (z^2-1)^{1/2}$ lie inside the convergence circle $C_{\tilde{R}}$ of (4), and choose Γ so that it lies fully inside it, too. Then, using (3), (4), and (8), we can relate $f(s,z)$ to $g(s,w)$ by the following formula:

$$f(s,z) = \frac{1}{\pi i}\int_\Gamma \frac{g(s,w)}{(w^2-2wz+1)^{1/2}}\, dw. \quad (9)$$

We assume now that the scattering amplitude $f(s,z)$ is bounded by a polynomial in s for all energies higher than a certain value,

$$|f(s,z)| < s^N, \quad s > s_0, \quad (10)$$

at all z inside the Lehmann–Martin ellipse, N being independent of z.

The function $g(s,w)$ can be estimated as follows. Relation (4) and inequality (6) imply

$$|g(s,w)| \leqslant \frac{M(s)L(s)}{2\sqrt{\pi}}(1-R^{-2})^{-1/2}\sum_{l=1}^{\infty}\frac{2l+1}{\sqrt{l}} R^{-(l+1)}|w|^l + 1 + O\left(\frac{1}{s}\right)$$

$$\leqslant \frac{M(s)L(s)}{[2\pi(R^2-1)]^{1/2}}\sum_{l=0}^{\infty}(2l+1)\left|\frac{w}{R}\right|^l.$$

Using the well-known formula $\Sigma_{l=0}^{\infty} l u^l = u(1-u)^{-2}$ for $|u|<1$ we obtain

$$|g(s,w)| \leqslant \frac{M(s)L(s)R(R+|w|)}{[2\pi(R^2-1)]^{1/2}(R-|w|)^2} \leqslant \frac{2M(s)L(s)R^2}{[2\pi(R^2-1)]^{1/2}(R-|w|)^2}, \quad (|w|<R).$$

We also can derive an asymptotic formula $(R^2-1)^{1/2} \sim (2a/s)^{1/4}\sqrt{2}$ for $s\to\infty$ because of (5) and $R = \rho + (\rho^2-1)^{1/2}$. Since $L(s)$ is the length of the ellipse, we have $L(s)\to 4$ for $s\to\infty$. Recall that $M(s) < s^N$ for sufficiently large s because of (10). Consequently, the last inequality yields

$$\limsup_{s\to\infty}\frac{|g(s,w)|}{h(s,w)} \leqslant 1,$$

where

$$h(s,w) = \frac{4}{(2a)^{1/4}\sqrt{\pi}}\frac{s^{N+1/4}}{(R-|w|)^2}.$$

This implies the following high-energy upper bound on $g(s,w)$:

$$|g(s,w)| < \xi\frac{s^{N+1/4}}{(R-|w|)^2}, \quad \text{for } |w|<R, \quad (11)$$

for any ξ larger than $4/[(2a)^{1/4}\sqrt{\pi}]$.

Besides this polynomial bound, an energy-independent upper bound on $g(s,w)$ holds inside the unit circle, $|w|<1$. It follows from the unitarity condition, $|a_l(s)| \leqslant 1$, and has the form

$$|g(s,w)| < \frac{c_1}{(1-|w|)^2}, \quad \text{for } |w|<1. \quad (12)$$

These two bounds can be used to improve the bound on $g(s,w)$ for $|w| \geqslant 1$ by applying the following Theorem.[6] Let $\phi(z)$ be analytic in a domain limited by two circular arcs A_1 and A_3 (see Fig. 1), which intersect at the points A and B. Let $\phi(z)$ be bounded from above by M_1 and M_3 on A_1 and A_3, respectively. Then $\phi(z)$ is bounded on the intermediate circular arc A_2 connecting the points A and B by M_2, where

$$M_2 = M_1^{\beta/(\alpha+\beta)} M_3^{\alpha/(\alpha+\beta)}. \quad (13)$$

Here α and β are the angles of intersection of A_1, A_2 and A_2, A_3, respectively.

We will use this theorem to extend the bound (12) in the complex w plane to a point denoted by $re^{i\theta}$ which lies outside the unit circle $|w|<1$. The three arcs must be chosen so that

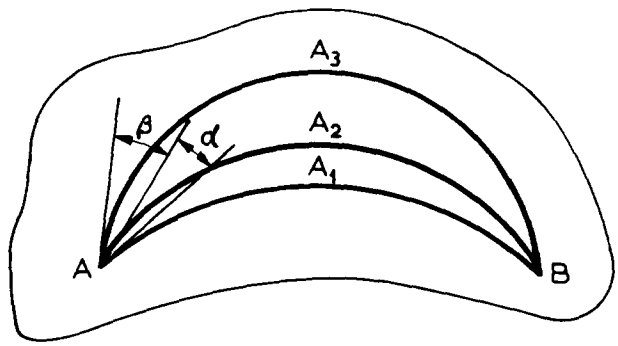


FIG. 1. The domain, in which $\phi(z)$ is holomorphic, is a sickle-shaped region between the arcs of circles A_1 and A_3.

A_1 and A_3 lie fully in $|w|<1$ and in $|w|<R$, respectively, the point $re^{i\theta}$ being situated on the arc A_2. This leads to the following situation (see Fig. 2):

(i) A_1 is the arc of a circle of radius $r-\epsilon$ centered at the origin (where $\epsilon>0$ will be determined later);

(ii) A_2 is the arc of the circle passing through the points $re^{i\theta}$ and $\pm(r-\epsilon)e^{i(\pi/2+\theta)}$;

(iii) A_3 is the arc of the circle passing through the points $(R-\eta)e^{i\theta}$ and $\pm(r-\epsilon)e^{i(\pi/2+\theta)}$ ($\eta>0$ will be determined later). Here, R is the radius defined by relation (7).

Some elementary trigonometrical calculations lead to the following relations:

$$\sin\alpha=\frac{\epsilon(2r-\epsilon)}{r^2+(r-\epsilon)^2}, \tag{14}$$

$$\begin{aligned}\sin\beta=2\{&r(r-\epsilon)[(R-\eta)^2-(r-\epsilon)^2]\\&-\epsilon(2r-\epsilon)(r-\epsilon)(R-\eta)\}\\&\times\{[r^2+(r-\epsilon)^2]\,[(R-\eta)^2+(r-\epsilon)^2]\}^{-1}.\end{aligned} \tag{15}$$

Since the ellipse E_ρ approaches the segment $[-1,1]$ with increasing s, we have to require

$$\epsilon\to0,\quad \eta\to0,\quad R-1,\quad r\to1,$$

with $r\geqslant1$, $r-\epsilon<1$. This means that

$$\alpha\to0,\quad \beta\to0,$$

and, consequently,

$$\frac{\alpha}{\beta}\frac{R-r}{\epsilon}\to1, \tag{16}$$

if, additionally,

$$\frac{\epsilon+\eta}{R-r}\to0. \tag{17}$$

Let us choose β/α such that

$$\beta/\alpha=\ln s \tag{18}$$

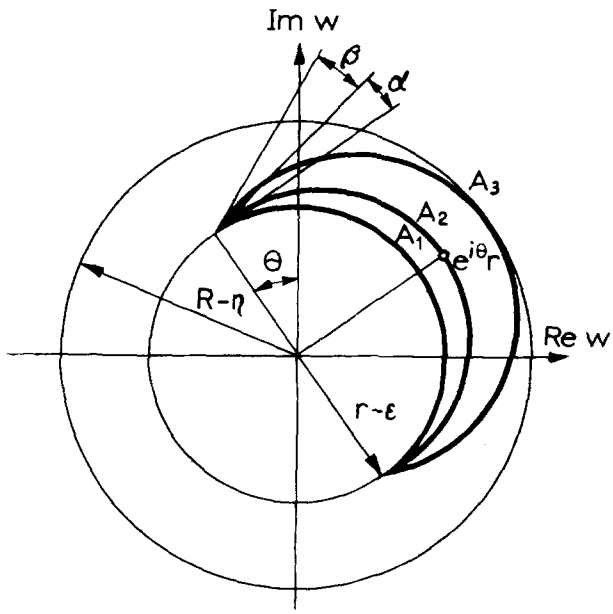


FIG. 2. Application of the Hadamard–Nevanlinna theorem to the function $g(s,w)$.

and apply the formula (13) by setting

$$M_1=\frac{c_1}{(1-|w|)^2},\quad \text{and}\quad M_3=\xi\frac{s^{N+1/4}}{\eta^2},$$

with $\eta=R-|w|$ [see (12) and (11), respectively]. Then, formula (13) takes the following form:

$$M_2=\left(\frac{c_1e^{N+1/4}}{(1-r+\epsilon)^2}\right)^{\ln s/(1+\ln s)}\left(\frac{\xi}{\eta^2}\right)^{1/(1+\ln s)}. \tag{19}$$

Whereas ϵ is determined from (16) and (18), η can be chosen in such a way that its influence in formula (19) is suppressed. In particular, we demand

$$\eta^{2/(1+\ln s)}\to e^{-2c_0},$$

for sufficiently large energies. If we choose, for instance, $\eta(s)=s^{-c_0}$ then the condition (17) induces the inequality $c_0>1/2$.

Further, we shall need a uniform estimate for $g(s,w)$ in a region $|w|\leqslant\tilde r$. We use again the condition (17) and obtain

$$\tilde Y=1+\frac{1}{1+\ln s}\left(\frac{2a}{s}\right)^{1/2}. \tag{20}$$

In this way, we are led to the following form of estimate in the region $|w|\leqslant\tilde r$:

$$|g(s,w)|<\frac{c_2}{[1-r+(R-r)/\ln s]^2}, \tag{21}$$

where $|w|=r$ and $c_2=c_1e^{N+1/4+2c_0}$ both for physical and for unphysical scattering angles.

Using this result for $g(s,w)$ and the relation (9) between $f(s,z)$ and $g(s,w)$ we can obtain the bound on the scattering amplitude $f(s,z)$. Its form is different for the physical and for the unphysical scattering angles.

3. THE CASE OF PHYSICAL SCATTERING ANGLES

For $\theta=0$ or π the integrand in Eq. (9) has poles at the points $w=z=\pm1$. Finding the value (9) at the points $z=\pm1$, inserting Eq. (21) into the result obtained and taking into account Eqs. (7) and (5), we obtain the following bound on the amplitude $f(s,z)$ for $s\gg a$:

$$f(s,\pm1)=g(s,\pm1)<c_2\frac{\ln^2 s}{(R-1)^2}\simeq\frac{c_2 s\ln^2 s}{2a}, \tag{22}$$

i.e., the Froissart bound (1) for forward scattering.

In the case of a scattering away from the forward or the backward direction, the integrand in (9) has no poles but has a cut in w along the interval $[e^{-i\theta},e^{i\theta}]$. We can write from (9)

$$|f(s,\cos\theta)|\leqslant\frac{1}{\pi}\int_\Gamma\left|\frac{g(s,w)}{(w^2-2w\cos\theta+1)^{1/2}}\right||dw|, \tag{23}$$

where the contour Γ connects the points $e^{i\theta}$ and $e^{-i\theta}$ and does not leave the interior of the circle $C_{\tilde r}$.

We choose the integration contour to be composed of (see Fig. 3):

(i) the interval AB: $w=re^{i|\theta|}$, $1-|\sin\theta|\leqslant r\leqslant1$;

(ii) the arc BC: $w=(1-|\sin\theta|)e^{i\phi}$, $-|\theta|\leqslant\phi\leqslant|\theta|$ (for $|\theta|>\pi/2$, ϕ runs from $|\theta|$ to π and from $-\pi$ to $-|\theta|$);

(iii) the interval CD: $w=re^{-i|\theta|}$, $1-|\sin\theta|\leqslant r\leqslant1$. Inserting relation (21) into (23) we see that the integrands over

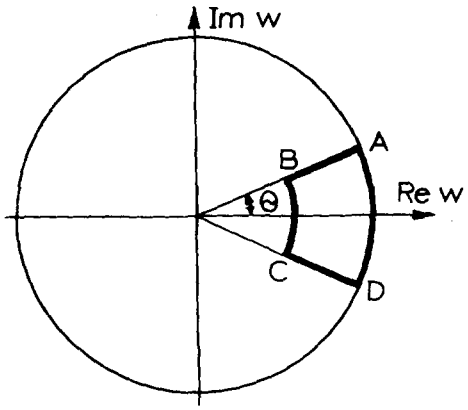


FIG. 3. Integration contour ABCD for the estimate of the right-hand side of Eq. (23) in the case of physical scattering angles.

the intervals AB and CD are equal and that the integral over BC does not exceed the bound on the integral over AB. We have

$$|w^2-2wz+1|^{1/2}=(1-r)^{1/2}\,[4r\sin^2\theta+(1-r)^2]^{1/4}.$$

Substituting (21) into (23) we obtain

$$|f(s,\cos\theta)| < \frac{3}{\pi}\int_{1-\sin\theta}^{1}\frac{|g(s,re^{i\theta})|}{(1-r)^{1/2}[4r\sin^2\theta+(1-r)^2]^{1/4}}\,dr$$
$$< c_3\int_0^{|\sin\theta|^{1/2}}\frac{dv}{(\delta+v^2)^2|\sin\theta|^{1/2}}, \tag{24}$$

where

$$v=(1-r)^{1/2},\quad \delta=\frac{R(s)-1}{\ln s}\simeq\frac{(2a)^{1/2}}{\sqrt{s}\,\ln s}\quad \text{for } s\gg a,$$

$$c_3=\frac{6c_2}{\pi}=\frac{12}{\pi}e^{N+1/4+2c_0},\ \text{since } c_1=2.$$

Evaluating this integral, we finally obtain the following upper bound on the scattering amplitude:

$$|f(s,\cos\theta)| < \frac{c_3}{2}\left(\frac{1}{\delta^{3/2}|\sin\theta|^{1/2}}\arctan\frac{|\sin\theta|^{1/2}}{\delta^{1/2}}+\frac{1}{\delta(\delta+|\sin\theta|)}\right). \tag{25}$$

This formula gives, apart from constant factors, the bound (1) for $\theta\ll\delta$ and the bound (2) for $\theta\gg\delta$.

4. UNPHYSICAL SCATTERING ANGLES

If z lies outside the physical interval $[-1,1]$, all formulas derived in Sec. 2 remain valid. We represent z in the form

$$z=\gamma\cos\theta+i(\gamma^2-1)^{1/2}\sin\theta, \tag{26}$$

where $\gamma>1$ and $\theta\in(-\pi,\pi]$, but θ has no longer the meaning of the scattering angle. Using Eq. (9) and the bound (21) we obtain

$$|f(s,z)|\leqslant\frac{c_2}{\pi}\int_{\Gamma'}\frac{1}{(1-r+\epsilon)^2}\frac{|dw|}{|w^2-2wz+1|^{1/2}}, \tag{27}$$

where $\epsilon=(R-r)/\ln s$ and Γ' connects the point $w=z+(z^2-1)^{1/2}$ with $w=z-(z^2-1)^{1/2}$ and lies fully in the interior of $C_{\tilde\gamma}$. Evidently,

$$z\pm(z^2-1)^{1/2}=\gamma_{\pm}\,e^{\pm i\theta},$$

and

$$w^2-2wz+1=(re^{i\varphi}-\gamma_+e^{i\theta})(re^{i\varphi}-\gamma_-e^{-i\theta}), \tag{28}$$

where

$$\gamma_{\pm}=\gamma\pm(\gamma^2-1)^{1/2},$$

and

$$w=re^{i\varphi}.$$

The contour Γ' must lie inside $C_{\tilde\gamma}$ and the following inequalities must be satisfied:

$$\gamma_+<\tilde r.$$

The right-hand side of (27) can be estimated in various ways and we shall not look for the best estimate here. One possibility would be to follow closely the approach of Sec. 3. We show another way, choosing for Γ' in (27) the segment connecting the points $\gamma_-e^{-i\theta}$ and $\gamma_+e^{i\theta}$. Along this line we have $|w^2-2wz+1|=b_2(1-\lambda^2)$, $-1\leqslant\lambda\leqslant1$, $r^2=b_0+b_1\lambda+b_2\lambda^2$ with $b_0=\gamma^2-\sin^2\theta$, $b_1=2\gamma(\gamma^2-1)^{1/2}$, $b_2=\gamma^2-\cos^2\theta$, $|dw|=(b_2)^{1/2}\,d\lambda$, and the right-hand side of (27) becomes

$$\frac{c_2}{\pi}\int_{-\pi/2}^{\pi/2}\frac{dx}{[1+R/\ln s-\Lambda(b_0+b_1\sin x+b_2\sin^2x)^{1/2}]^2},$$
$$\Lambda=1+1/\ln s.$$

This integral can easily be shown to be bounded by c_2I/π, where

$$I=\int_{-\pi/2}^{\pi/2}\frac{dx}{[1+R/\ln s-\Lambda\gamma-\Lambda(\gamma^2-1)^{1/2}\sin x]^2}.$$

Thus

$$|f(s,z)|\leqslant c_2\frac{r_R-\Lambda\gamma}{(r_R-\Lambda\gamma_+)^{3/2}(r_R-\Lambda\gamma_-)^{3/2}},$$
$$r_R=1+R/\ln s. \tag{29}$$

This is a high-energy upper bound on $f(s,z)$ for z inside $E(s)$ which is the ellipse with foci $+1$, -1 and with the semimajor axis $1+[1/(1+\ln s)^2]\,a/s$. The bound depends on s through R and Λ. Moreover, in contrast to relation (25), which holds for physical angles, z in (29) outside the physical segment $[-1,1]$ cannot be kept fixed. With increasing s, the ellipse $E(s)$ shrinks to the segment $[-1,1]$. The parameter γ tends to 1 so as to preserve $r_R-\Lambda\gamma_+$ positive at all energies above some value.

It is to be mentioned that the bound (29) is not the best one and may be improved in two different ways. One of them is a better estimate of the right-hand side of (9), of course. The other one consists in solving the Dirichlet problem for the doubly connected domain whose boundary is formed by two disjoint curves Γ_1 and Γ_2, where Γ_1 is the interval $[-1,1]$ and Γ_2 is the boundary of $E(s)$.

As (1) and (2) are better bounds than (29) at the points $z\in\Gamma_1$, the solution of the Dirichlet problem for $f(s,z)$ bounded by (1), (2), and by (29) at $z\in\Gamma_1$ and at $z\in\Gamma_2$ respectively, would yield an improvement of (29) inside $E(s)$. A detailed analysis of this problem will be given later.

5. THE CASE OF ARBITRARY SPINS

For the case of the scattering of particles with spins the series (3) will be rewritten as follows[7]:

$$F_i \equiv F^{\lambda_a\lambda_d}_{\lambda_c\lambda_b}(s,z) = \sum_{J=M}^{\infty} (2J+1) f^J_{\lambda_a\lambda_b\lambda_c\lambda_d}(s)\, d^J_{\lambda\mu}(z), \quad (30)$$

where λ_j is the helicity of the jth particle,

$$\lambda = \lambda_a - \lambda_b, \quad \mu = \lambda_c - \lambda_d, \quad M = \max(|\lambda|,|\mu|).$$

The series (30) converges in Martin's ellipse E_{ρ_i} [8] for the fixed energy $s^{1/2}$ and helicities λ_i.

The functions $G^{\lambda_a\lambda_d}_{\lambda_c\lambda_b}(s,w)$, similar to the function involved in Eq. (4), are defined as

$$G^{\lambda_a\lambda_d}_{\lambda_c\lambda_b}(s,w) = \sum_{J=M}^{\infty} (2J+1) f^J_{\lambda_a\lambda_b\lambda_c\lambda_d}(s)\, \mathrm{sign}(\lambda,\mu)(-i)^{|\lambda-\mu|} \times \left(\frac{\Gamma(J+m+1)\Gamma(J-m+1)}{\Gamma(J+M+1)\Gamma(J-M+1)}\right)^{1/2} w^{J-M}, \quad (31)$$

where

$$m = \min(|\lambda|,|\mu|)$$

and they satisfy the conditions (11) and (12) (see Ref. 9).

Using the integral representation for the functions $d^J_{\lambda\mu}(z)$,[10] we obtain the following relation between the functions $F^{\lambda_a\lambda_d}_{\lambda_c\lambda_b}(s,z)$ and $G^{\lambda_a\lambda_d}_{\lambda_c\lambda_b}(s,w)$:

$$F^{\lambda_a\lambda_d}_{\lambda_c\lambda_b}(s,\cos\theta) = \frac{1}{\pi i}\int_\Gamma \frac{dw\, G^{\lambda_a\lambda_d}_{\lambda_c\lambda_b}(s,w) h(\theta,t,a,b)}{(1-2\cos\theta w + w^2)^{1/2}}, \quad (32)$$

where

$$t = \frac{w - \cos\theta - (1-2\cos\theta w + w^2)^{1/2}}{i\sin\theta},$$

$$a = |\lambda - \mu|, \quad b = |\lambda + \mu|,$$

$$h(\theta,t,a,b) = \frac{[\cos(\theta/2) + it\sin(\theta/2)]^{a+b}}{t^a}, \quad (33)$$

and the contour Γ is the same as that for the spinless case (see Fig. 3).

For fixed values λ,μ the function $h(\theta,t,a,b)$ is majorized on the contour Γ by a constant which is independent of θ. Indeed, let us express w from Eq. (33) in terms of θ,t:

$$w = \cos\theta + (i/2)(t + t^{-1})\sin\theta$$

On the contour Γ the function $(w - \cos\theta)/\sin\theta$ is bounded by a constant for any angles θ, therefore t,t^{-1} and $h(\theta,t,a,b)$ are bounded by constants which are independent of θ. Therefore the integral (32) is reduced to Eq. (23) and, in analogy with Eq. (25), we obtain the upper bound on the amplitude $F_i(s,\cos\theta)$:

$$F_i(s,\cos\theta) < \mathrm{const}\left(\frac{1}{\delta_i^{3/2}|\sin\theta|^{1/2}}\arctan\frac{|\sin\theta|^{1/2}}{\delta_i^{1/2}} + \frac{1}{\delta_i(\delta_i + |\sin\theta|)}\right), \quad (34)$$

where

$$\delta_i = \frac{R_i(s)-1}{\ln s} \simeq \frac{(2a_i)^{1/2}}{\sqrt{s}\ln s} \quad \text{for } s \gg a_i.$$

If $\theta \ll \delta_i$ or $\theta \gg \delta_i$ formula (34) transforms to Eq. (1) or Eq. (2). It agrees with the result of Refs. 11 and 12. Analogous considerations, as in Refs. 13 and 14 for the helicity amplitudes, can be used also here to obtain bounds similar to (25).

ACKNOWLEDGMENTS

The authors would like to thank C. Cronström, V. Glaser, A. Martin, and L. Trlifaj for useful discussions.

[1] M. Froissart, Phys. Rev. **123**, 1053 (1961).
[2] A. Martin, Phys. Rev. **129**, 1432 (1963).
[3] T. Kinoshita, J.J. Loeffel, and A. Martin, Phys. Rev. Lett. **10**, 460 (1963).
[4] T. Kinoshita, J.J. Loeffel, and A. Martin, Phys. Rev. B **135**, 1464 (1964).
[5] E.W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Chelsea, New York, 1965).
[6] R. Nevanlinna, *Eindeutige analytische Funktionen*, 2nd ed. (Springer, Berlin, 1953).
[7] M. Jacob and G.C. Wick, Ann. Phys. **7**, 404 (1959).
[8] F. Mahoux and A. Martin, Phys. Rev. **174**, 2140 (1968).
[9] V.V. Ezhela and M.A. Mestvirishvili, Theor. Math. Phys. (USA) [Translation of Teor. Mat. Fiz. (USSR)] **7**, 467 (1971).
[10] N. Vilenkin, *Special Functions and the Theory of Group Representations* (Translations of Mathematical Monographs, Vol. **22** American Mathematical Society, Providence, R. I., 1968).
[11] K. Yamamoto, Nuovo Cimento **27**, 1277 (1963).
[12] Y. Hara, Phys. Rev. B **136**, 507 (1964).
[13] H. Cornille, Nuovo Cimento, **31**, 1101 (1964).
[14] H. Cornille and G. Philippe, Nucl. Phys. B **5**, 627 (1968).

Solution of the unitarity equation with overlapping left and right cuts: A tool for study of the S^* and similar systems [a)]

Porter W. Johnson
Illinois Institute of Technology, Chicago, Illinois 60616

Robert Lee Warnock [b)]
Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720

(Received 1 August 1979; accepted for publication 24 January 1980)

The partial-wave unitarity condition is complicated by the presence of overlapping left and right branch cuts when the lowest exchanged mass is small in comparison to the direct-channel mass. A coupled-channel ND^{-1} method for constructing unitary amplitudes with overlapping cuts is described. The study is motivated in part by the problem of analyzing the $\pi\pi - K\bar{K}$ system near the S^* resonance.

PACS numbers: 11.20.Fm, 11.50.Ec, 11.80.Gw

I. INTRODUCTION

We discuss the unitarity condition for coupled two-body channels in a definite angular-momentum state. For simplicity in notation we take two channels, but our methods apply as well for any finite number. The square of the energy in the center-of-mass frame is denoted by s, and s_i is the threshold of the ith channel, $s_1 \leqslant s_2$. As is appropriate in analytic S-matrix theory, we study a generalization of ordinary unitarity. If the masses of the particles in one channel are not too dissimilar to those in the other, generalized unitarity has the form

$$[T(s_+) - T(s_-)]/2i = T(s_+)\rho(s)T(s_-), \qquad (1.1)$$
$$s > s_1, \quad T(s_\pm) = \lim_{\epsilon \to 0+} T(s \pm i\epsilon),$$

where the 2×2 scattering matrix $T(s)$ is analytic in regions above and below the half-line $s > s_1$. The diagonal matrix of phase-space factors $\rho(s)$ includes unit step functions θ that vanish below channel thresholds:

$$\rho(s) = \{\rho_i(s)\delta_{ij}\}, \quad \rho_i(s) = \theta(s - s_i)q_i(s). \qquad (1.2)$$

For the case of spinless, equal-mass particles in channel i one has

$$q_i(s) = [(s - s_i)/s]^{1/2}. \qquad (1.3)$$

Generalized unitarity (1.1) restricts the amplitudes $T_{12}(s)$ and $T_{22}(s)$ in the region $s_1 < s < s_2$ where channel 2 is closed, whereas ordinary unitarity refers only to open channels.

A complication arises if mass differences are large. Namely, the left cuts of some of the amplitudes overlap the half-line $s > s_1$. This occurs when the lowest mass in a cross channel is sufficiently small in comparison with the mass of the direct channel. The unitarity condition then becomes

$$[T(s_+) - T(s_-)]/2i = T(s_+)\rho(s)T(s_-) + \Delta_L T(s), \qquad (1.4)$$

where $\Delta_L T(s)$ is the matrix of discontinuities of $T(s)$ over the left cuts (denoted collectively by L).

An example is the two-channel problem with $\pi\pi$ and $K\bar{K}$ channels in a definite isospin state, considered near the $K\bar{K}$ threshold where the 4π state has only a small production cross section. Under the assumption of Mandelstam analyticity, the partial-wave amplitude for $K\bar{K}\to K\bar{K}$ has a left cut beginning at the branch point $s = 4(m_K^2 - m_\pi^2)$. According to (1.4) the right cut begins at $s = 4m_\pi^2$, so that the two cuts overlap. The amplitudes for $\pi\pi\to\pi\pi$ and $\pi\pi\to K\bar{K}$ do not have overlapping cuts; their nearest left branch points are at $s = 0$. The possible importance of treating correctly the overlapping cuts in the phenomenology of the $\pi\pi\to K\bar{K}$ system, especially near the S^* resonance, has been emphasized by Yndurain,[1-3] González-Arroyo,[3,4] and co-workers.[3] Although the $\pi\pi\to K\bar{K}$ system has been discussed extensively,[5] it appears that a full explication of the unitarity effects remains to be made. A similar situation of overlapping cuts occurs in the $N\bar{N}$ system, which is of high current interest in connection with baryonium states.[6]

In studying systems with overlapping cuts, from either a dynamical or a phenomenological viewpoint, one encounters a generalization of the standard problem of partial-wave dispersion relations.[7] That is, given the left cut part of the T matrix,

$$B(s) = T(s) - \frac{1}{\pi}\int_{s_1}^{\infty} \frac{T(s'_+)\rho(s')T(s'_-)ds'}{s' - s}, \qquad (1.5)$$

determine the most general $T(s)$ having that left cut part and satisfying the augmented unitarity equation (1.4) as well as appropriate conditions of analyticity and asymptotic behavior. We shall provide a straightforward solution of this problem, based on the matrix ND^{-1} method.[8-11] As in the usual ND^{-1} method, the problem is reduced to solving a linear integral equation for $N(s)$. It is gratifying to find that the equation is identical in form to the usual one. Only the derivation of the equation is altered. Being of Fredholm type under weak conditions on $B(s)$, the equation is amenable to numerical solution.

Our results are applicable in phenomenology as well as in dynamical schemes. In phenomenology the traditional approach to determination of $B(s)$ is to use crossing symmetry and experimental information on scattering in the cross channel. Such an approach determines the nearby singulari-

[a)]Work supported in part by the National Science Foundation and the United States Department of Energy under contract No. W-7405-ENG-48.
[b)]Participating guest.

 0022-2488/81/020385-15$1.00

ties of $B(s)$ to a certain extent, but leaves the distant singularities to be represented by empirical parameters. A potentially more informative approach now under development is to use a new definition of Reggeon exchange, valid at low as well as high energy.[12] The Reggeon exchanges involve all partial waves in the cross channel and form an important (possibly dominant) part of the analytically continued cross-channel absorptive part. It is hoped that a model of $B(s)$ based primarily on Reggeon exchanges will be realistic.

An ambitious scheme for construction of a crossing-symmetric unitary Regge theory, proposed in Refs. 12 and 13 and extended in a forthcoming paper to allow coupled channels, requires solution of a generalization of the problem treated here. In a crossing-symmetric treatment of coupled $\pi\pi$ and $K\bar{K}$ channels, for instance, one must account for the 4π threshold at $s = 16m_\pi^2$ in the $K\bar{K}$ amplitude, which lies to the left of the beginning of the left cut at $s = 4(m_K^2 - m_\pi^2)$. As we shall show in a later paper, this complicated situation of overlapping cuts can be handled in a rather simple way by extending the present ND^{-1} method to allow a matrix of externally prescribed absorption parameters, in analogy to the work of Ref. 14. In the crossing-symmetric theory the absorption parameters for the 4π state, etc., are obtained dynamically through crossing. The extended ND^{-1} method with absorption should also be useful in phenomenology, especially for study of absorption in the $N\bar{N}$ system. A correct treatment of overlapping cuts is conceivably important in assessing the effects of absorption on baryonium states predicted from crossed NN potentials.[6]

Section II contains the general solution of the two-channel problem under rather weak conditions on $B(s)$. It will be evident that the method works as well for n channels. The Castillejo–Dalitz–Dyson (CDD) ambiguity[7] is treated in detail, since a complete treatment for the coupled channel case has not been available in the literature. Recently Nenciu, Rasche, Stihi, and Woolcock[15] criticized the ND^{-1} method and suggested a method based on a pole approximation to $B(s)$ as a replacement. We feel that the discussion of Secs. II and III answers their criticisms and shows that the method is both general and practical. In our experience the pole approximation has not been very useful, since in realistic models $B(s)$ is not given in terms of poles and to approximate it by poles with sufficient accuracy is rather awkward. We note, however, that the pole approximation can be used in the ND^{-1} scheme with overlapping cuts and that it leads as usual to explicit analytic forms for the solution of the integral equation.

In Sec. IV we give an ND^{-1} method for a single-channel problem with absorption present at threshold; for example, $K\bar{K} \to K\bar{K}$. The absorption parameters are regarded as given and left cuts may or may not overlap the absorption cut below threshold.

In Sec. V we discuss a special case of our problem solved recently by González-Arroyo[4]; namely, a two-channel problem in which only the element $B_{22}(s)$ of $B(s)$ is nonzero. We reveal two new aspects of the González–Arroyo solution by deriving it from our formalism: (a) It necessarily entails CDD poles as defined in the two-channel formalism; if there is not at least one CDD pole, only the trivial solution in which $T_{11}(s) = T_{12}(s) = 0$ is obtained; (b) even though the González-Arroyo solution entails arbitrary rational functions, it is not the general solution of the problem with $B_{11}(s) = B_{12}(s) = 0$; rather, it corresponds to putting some elements of the CDD pole residue matrices equal to zero.

In Sec. VI we comment on a proposal of Yndurain for an explicit unitary parametrization of the T matrix with overlapping cuts.

Appendix A is concerned with asymptotic estimates of principal-value integrals under conditions of logarithmic decrease of the density function. Appendix B contains the proof that the integral equation of Sec. II is of Fredholm type under conditions of logarithmic decrease of $B(s)$.

We hope to reexamine in a later paper the phenomenology of the $\pi\pi \to K\bar{K}$ system near the S^* resonance using the methods described.

II. GENERAL SOLUTION FOR TWO-CHANNEL CASE

In this section we solve the two-channel problem, with two pseudoscalar mesons of mass m_i in the ith channel. The phase-space factors are as given in (1.3), with $s_i = 4m_i^2$. We make analyticity assumptions weaker than those implied by the Mandelstam representation, since the extra generality involves little effort.

Let us first recall the implications of the Mandelstam representation. The partial-wave amplitudes $T_{11}(s)$ and $T_{12}(s) = T_{21}(s)$ are analytic in the s plane, each with cuts $(-\infty, 0]$, $[s_1, \infty)$, where $s_1 = 4m_1^2$. If $m_2^2 < 2m_1^2$, $T_{22}(s)$ is analytic in the plane with cuts $(-\infty, 4(m_2^2 - m_1^2)]$, $[s_1, \infty)$.

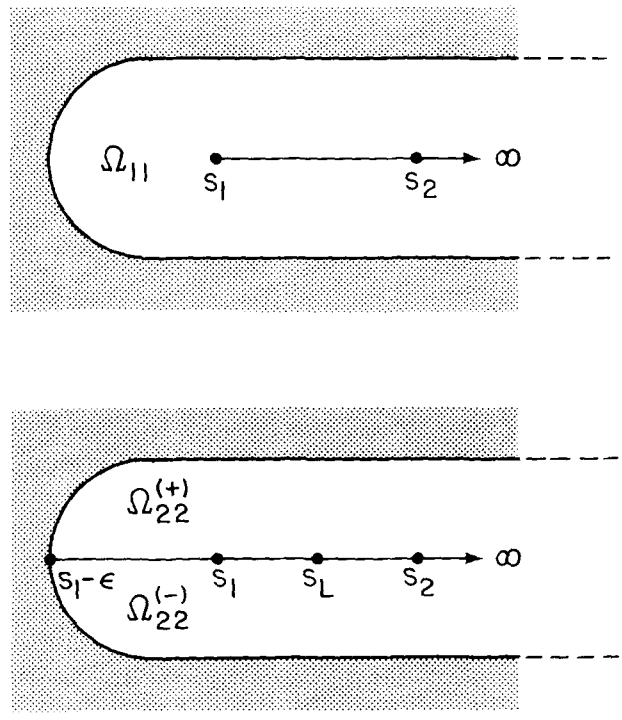


FIG. 1. Possible analyticity domains Ω_{11}, $\Omega_{22}^{(\pm)}$ of $T_{11}(s)$, $T_{22}(s)$, respectively.

If $m_2^2 \geqslant 2m_1^2$, we must regard $T_{22}(s)$ as sectionally analytic, since the cuts overlap and divide the plane in two:

$$T_{22}(s) = \begin{cases} T_{22}^{(+)}(s), & \mathrm{Im}s > 0, \\ T_{22}^{(-)}(s), & \mathrm{Im}s < 0, \end{cases} \tag{2.1}$$

where $T_{22}^{(+)}(s)$ and $T_{22}^{(-)}(s)$ are analytic in their respective half-planes. One has $T_{ij}(s) = T_{ji}(s^*)^*$, which for $i = j = 2$ means that $T_{22}^{(+)}(s) = T_{22}^{(-)}(s^*)^*$.

Our requirements on the T matrix, weaker with respect to analyticity, will be as follows:

(a) $T_{ij}(s) = T_{ji}(s)$; (2.2a)

(b) $T_{11}(s)$ and $T_{12}(s)$ are analytic in open neighborhoods Ω_{11}, Ω_{12} of the half-line $[s_1, \infty)$, as illustrated in Fig. 1; (2.2b)

(c) $$T_{22}(s) = \begin{cases} T_{22}^{(+)}(s), & s \in \Omega_{22}^{(+)}, \\ T_{22}^{(-)}(s), & s \in \Omega_{22}^{(-)}, \end{cases} \tag{2.2c}$$

where $T_{22}^{(\pm)}(s)$ is analytic in $\Omega_{22}^{(\pm)}$. Here $\Omega_{22}^{(\pm)}$ is an open region of the upper half plane with $[s_1 - \epsilon, \infty)$ as part of its boundary, and $\Omega_{22}^{(-)}$ is the complex conjugate of that region (see Fig. 1);

(d) $T(s) = T(s^*)^*$; (2.2d)

(e) $$\Delta T(s) = [T(s_+) - T(s_-)]/2i$$
$$= T(s_+)\rho(s)T(s_-) + \Delta_L T(s), \quad s \geqslant s_1, \tag{2.2e}$$
$$\Delta_L T(s) = \begin{pmatrix} 0 & 0 \\ 0 & \theta(s_L - s)\phi(s) \end{pmatrix}, \quad s_1 < s < s_2,$$

where $\theta(s)$ is the unit step function and $\phi(s) = \Delta T_{22}(s)$, $s_1 \leqslant s \leqslant s_L$;

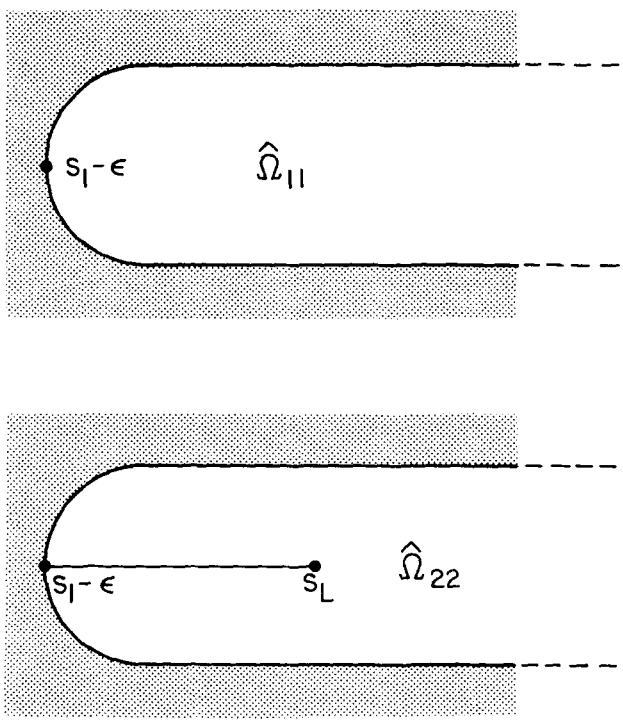


FIG. 2. Possible analyticity domains $\hat{\Omega}_{11}$, $\hat{\Omega}_{22}$ of $B_{11}(s)$, $B_{22}(s)$, respectively.

(f) $$|T(s_+)| \leqslant \kappa(\ln s)^{-\alpha}, \quad s \geqslant s_1,$$
$$|T(s_+) - T(s'_+)| \leqslant \kappa(\ln s)^{-\alpha} \left| \frac{s - s'}{s} \right|^{\mu}, \quad s' \geqslant s \geqslant s_1, \tag{2.2f}$$
$$\alpha > 1, \quad 0 < \mu \leqslant \tfrac{1}{2}.$$

Here and in the following, κ represents a generic positive constant, which is understood to have different values in different equations. The inequalities (2.2f) apply to each element of the matrix $T(s)$ separately. The second of these inequalities follows from the stronger but more comprehensible requirement that $T(s_+)$ be Hölder-continuous for $s < r$ and continuously differentiable for $s > r$ with $|T'(s_+)| < \kappa s^{-1}\ln^{-\alpha} s$, the point r being arbitrary.

We shall determine the entire class of T matrices satisfying conditions (2.2a), (2.2b), (2.2c), (2.2d), (2.2e) and (2.2f) and having the same given left-hand cut term,

$$B(s) = T(s) - \frac{1}{\pi}\int_{s_1}^{\infty} \frac{T(s'_+)\rho(s')T(s'_-)ds'}{s' - s}. \tag{2.3}$$

Note that property (2.2f) ensures convergence of the integral in (2.3). The following conditions on $B(s)$ are a consequence of the conditions on $T(s)$ and the definition (2.3):

(a) $B_{ij}(s) = B_{ji}(s)$; (2.4a)

(b) $B_{11}(s)$, $B_{12}(s)$, and $B_{22}(s)$ are analytic in $\hat{\Omega}_{11} = \Omega_{11} \cup [s_1, \infty)$, $\hat{\Omega}_{12} = \Omega_{12} \cup [s_1, \infty)$, and $\hat{\Omega}_{22} = \Omega_{22}^{(+)} \cup \Omega_{22}^{(-)} \cup [s_L, \infty)$, resp. (see Fig. 2); (2.4b)

(c) $B(s) = B(s^*)^*$; (2.4c)

(d) $$|B(s)| \leqslant \kappa(\ln s)^{-\alpha}, \quad s \geqslant s_1,$$
$$|B(s) - B(s')| \leqslant \kappa(\ln s)^{-\alpha} \left| \frac{s - s'}{s} \right|^{\mu}, \quad s' \geqslant s \geqslant s_1. \tag{2.4d}$$

The property (2.4d) is obtained from (2.3) with the help of Lemma 2 on asymptotic behavior of principal-value integrals which is proved in Appendix A. The other properties of B follow immediately from (2.2).

Henceforth we suppose that a function $B(s)$, satisfying (2.4a), (2.4b), (2.4c), and (2.4d), is given. We seek the most general $T(s)$ that gives that $B(s)$ through (2.3), and which satisfies (2.2a), (2.2b), (2.2c), (2.2d), (2.2e), and (2.2f). Our analysis is based on the nontrivial theorem that any $T(s)$ satisfying conditions (2.2d), (2.2e), and (2.2f) has an ND^{-1} representation with appropriate properties. To be more exact, under those conditions there exists a 2×2 matrix $\mathscr{D}(s)$ such that[16,17]

(a) $\mathscr{D}_{ij}(s)$ is analytic in the plane with cut $[s_i, \infty)$ and is defined by continuity on the cut. The function on the cut, $\mathscr{D}_{ij}(s_\pm)$, is Hölder-continuous on any finite interval. (2.5a)

(b) $\mathscr{D}(s_-) = [1 + 2i\rho(s)T(s_+)]\mathscr{D}(s_+)$; (2.5b)

(c) $\mathscr{D}(s) = \mathscr{D}(s^*)^*$; (2.5c)

(d) $\mathscr{D}(s)$ is nonsingular (has an inverse) at every finite point of the cut plane, including points $s_\pm$ on the cut; (2.5d)

(e) There are integers n_i such that the modified matrix

$$\tilde{\mathscr{D}}(s) = [s^{-n_1}\mathscr{D}_{\cdot 1}(s),\ s^{-n_2}\mathscr{D}_{\cdot 2}(s)]$$

tends to a finite, real, nonsingular limit as $|s| \to \infty$:

$$\tilde{\mathscr{D}}(s) \to \tilde{\mathscr{D}}(\infty) = \tilde{\mathscr{D}}^*(\infty), \quad \det\tilde{\mathscr{D}}(\infty) \neq 0.$$

Here $\mathscr{D}_{\cdot j}(s)$ denotes the jth column of $\mathscr{D}(s)$. (2.5e)

The properties (2.5) clearly do not determine $\mathscr{D}(s)$ uniquely; at the least, one may interchange the columns of a given $\mathscr{D}(s)$ and multiply them by nonzero constants, thereby obtaining a new matrix that satisfies (2.5). Nevertheless, the (nonordered) pair of integers n_1, n_2 is uniquely determined by the asymptotic behavior of $T(s)$, and $n_1 + n_2$ sets the degree of ambiguity in the determination of $T(s)$ from a given $B(s)$, as we shall explain presently.

In the single-channel case $\mathscr{D}(s)$ is determined up to a constant multiplier and has the familiar form

$$\mathscr{D}(s) = A\exp\left(-\frac{s}{\pi}\int_{s_1}^{\infty}\frac{\delta(s')ds'}{s'(s'-s)}\right), \tag{2.6}$$

where A is an arbitrary real constant and $\delta(s)$ is the phase shift, normalized so that $\delta(s_1) = 0$. In the many-channel case there is, in general, no closed expression for $\mathscr{D}(s)$. Rather, $\mathscr{D}(s)$ is obtained through solution of a certain Fredholm integral equation with a kernel constructed from $T(s_+)$. If $\delta(s)$ in (2.6) tends to a limit $\delta(\infty)$ and obeys the bounds

$$\begin{aligned}&|\delta(s) - \delta(\infty)| \leqslant \kappa(\ln s)^{-\alpha},\\ &|\delta(s) - \delta(s')| \leqslant \kappa(\ln s)^{-\alpha}\left|\frac{s-s'}{s}\right|^{\mu}.\\ &s<s', \quad 0<\mu<1, \quad \alpha>1,\end{aligned} \tag{2.7}$$

then

$$\mathscr{D}(s_+) \sim s^{\delta(\infty)/\pi}, \quad s\to+\infty. \tag{2.8}$$

If $\delta(\infty) \geqslant \pi$, one has a CDD ambiguity in the determination of $T(s)$ from a given $B(s)$; cf. Ref. 7. We shall find a similar ambiguity in the two-channel case if $n_1 + n_2 \geqslant 1$. Let us write

$$T(s) = [T(s)\mathscr{D}(s)]\mathscr{D}^{-1}(s) = \mathscr{N}(s)\mathscr{D}^{-1}(s), \tag{2.9}$$

and compute the discontinuity of $\mathscr{N}(s)$ from (2.5b). If the unitarity equation (2.2e) holds, we have

$$\begin{aligned}\Delta\mathscr{N}(s) &= [T(s_+)\mathscr{D}(s_+) - T(s_-)\mathscr{D}(s_-)]/2i\\ &= \{T(s_+) - T(s_-)[1 + 2i\rho(s)T(s_+)]\}\mathscr{D}(s_+)/2i\\ &= \Delta_L T(s)\mathscr{D}(s_+) = \Delta_L T(s)\mathscr{D}(s), \quad s\geqslant s_1.\end{aligned} \tag{2.10}$$

In the final step of this calculation we are able to replace $\mathscr{D}(s_+)$ by $\mathscr{D}(s)$ because of the form of $\Delta_L T(s)$ and the fact that the cut of $\mathscr{D}_{2j}(s)$ begins at $s = s_2$. We have $\Delta\mathscr{N}(s) = 0$, $s\geqslant s_1$, in the simpler case in which left- and right-hand cuts do not overlap. With overlapping cuts,

$$\begin{aligned}&\Delta\mathscr{N}_{1j}(s) = 0,\\ &\Delta\mathscr{N}_{2j}(s) = \theta(s_L - s)\phi(s)\mathscr{D}_{2j}(s),\\ &j = 1,2, \quad s\geqslant s_1.\end{aligned} \tag{2.11}$$

The left-hand cut of each matrix element $\mathscr{N}_{ij}(s)$ does not overlap the right-hand cut of the corresponding $\mathscr{D}_{ij}(s)$.

We next consider the possible asymptotic behaviors of $\mathscr{D}(s)$ allowed by (2.5e) and for each type of behavior write a Cauchy representation for a matrix $D(s)$ closely related to $\mathscr{D}(s)$. The matrices $D(s)$ will subsequently be used to derive integral equations for $N(s) = T(s)D(s)$. We suppose initially that there is no bound-state pole of $T(s)$, and also that neither column of $\mathscr{D}(s)$ tends to the null vector as $|s|\to\infty$; thus, $n_i \geqslant 0$. Presently we shall show that the vanishing of a column of $\mathscr{D}(s)$ at infinity is an exceptional case, not expected to occur in realistic models, provided that there are no bound states.

First take *Case* 1: $n_1 = n_2 = 0$ in (2.5e). We define

$$D(s) = \mathscr{D}(s)\mathscr{D}^{-1}(\infty), \quad N(s) = \mathscr{N}(s)\mathscr{D}^{-1}(\infty). \tag{2.12}$$

By properties (2.5b), (2.5c), and (2.2f),

$$\begin{aligned}\mathrm{Im}D(s_+) &= [D(s_+) - D(s_-)]/2i\\ &= -\rho(s)T(s_+)\mathscr{D}(s_+)\mathscr{D}^{-1}(\infty)\\ &= -\rho(s)\mathscr{N}(s)\mathscr{D}^{-1}(\infty)\\ &= -\rho(s)N(s) = O(\ln^{-\alpha}s).\end{aligned} \tag{2.13}$$

If follows that $D(s)$ has the Cauchy representation (Case 1):

$$D(s) = 1 - \frac{1}{\pi}\int_{s_1}^{\infty}\frac{\rho(s')N(s')ds'}{s'-s}. \tag{2.14}$$

We next consider *Case* 2: $n_1 + n_2 > 0$ in (2.5e). To define $D(s)$ in this case we first choose any polynomials of the form

$$P_j(s) = \begin{cases}\prod_{k=1}^{n_j}(s - s_{kj}), & s_{kj} < s_1, \quad \text{if } n_j > 0,\\ 1, & \text{if } n_j = 0,\end{cases} \tag{2.15}$$

where the real points s_{kj}, $k = 1, 2, \ldots, n_j$, $j = 1, 2$, are all distinct. Then $D(s)$ is defined by

$$D(s) = \mathscr{D}(s)\begin{pmatrix}P_1^{-1}(s) & 0\\ 0 & P_2^{-1}(s)\end{pmatrix}\widetilde{\mathscr{D}}^{-1}(\infty), \tag{2.16}$$

and it has the Cauchy representation

$$\begin{aligned}D(s) = 1 &- \sum_{j=1}^{2}\sum_{k=1}^{n_j}\frac{C^{(kj)}}{s_{kj}-s}\\ &- \frac{1}{\pi}\int_{s_1}^{\infty}\frac{\rho(s')N(s')ds'}{s'-s},\end{aligned} \tag{2.17}$$

since $D(\infty) = 1$ and $N(s) = O(\ln^{-\alpha}s)$. The residue matrices $C^{(kj)}$ have components

$$C^{(kj)}_{lm} = \mathscr{D}_{lj}(s_{kj})\left(\prod_{p=1,p\neq k}^{n_j}(s_{kj} - s_{pj})\right)^{-1}\mathscr{D}^{-1}_{jm}(\infty). \tag{2.18}$$

Henceforth we shall consolidate the indices k and j and write (2.17) as (Case 2):

$$D(s) = 1 - \sum_{i=1}^{n_c}\frac{C_i}{\sigma_i - s} - \frac{1}{\pi}\int_{s_1}^{\infty}\frac{\rho(s')N(s')ds'}{s'-s}. \tag{2.19}$$

Except for possible poles, the matrices $D(s)$ have the same properties (2.5) as $\mathscr{D}(s)$. Also, (2.10) implies that

$$\Delta N(s) = \Delta_L T(s)D(s), \quad s\geqslant s_1. \tag{2.20}$$

The poles in (2.19) are analogous to the familiar CDD poles of the single-channel case: we shall again call them CDD poles. Because the pole positions σ_i are all distinct, the residue matrices C_i are singular:

$$\det C_i = 0. \tag{2.21}$$

That is seen from (2.18): the matrix $C^{(kj)}$ has rank 1, being a dyadic constructed from the vectors $\mathscr{D}_{\cdot j}(s_{kj})$ and $\mathscr{D}^{-1}_{j\cdot}(\infty)$.

We now turn to the derivation of the integral equation obeyed by $N(s) = T(s)D(s)$ for each of the two cases. The integral equation has a dual status. First, it is a necessary

condition on the $N(s)$ associated with any given $T(s)$. Second, it is a means of constructing a properly analytic, unitary, and symmetric $T(s)$ from a given $B(s)$. In applications one usually thinks only of the second aspect, but for the general theory, especially for demonstrating the generality of the solution of the construction problem, it is necessary to consider both aspects. We begin by deriving the equation as a necessary condition on $N(s)$ for a given $T(s)$ and later treat the construction problem.

For a given $T(s)$ satisfying conditions (2.2) and such that Case 1 holds, we examine the matrix function

$$\Lambda(s) = [T(s) - B(s)]D(s) - \frac{1}{\pi}\int_{s_1}^{\infty} \frac{B(s')\rho(s')N(s')ds'}{s' - s}. \tag{2.22}$$

Since $B(s)\rho(s)N(s) = O(\ln^{-2\alpha} s)$, the integral converges. Notice that by (2.3) the difference $T(s) - B(s)$ is defined in the whole cut plane, even though $T(s)$ and $B(s)$ separately may not be, in view of our weak assumptions on the region of definition and analyticity of $T(s)$. Clearly (2.22) defines a function $\Lambda_1(s) = \Lambda(s)$, analytic in the half plane $\mathrm{Im}s > 0$, and another function $\Lambda_2(s) = \Lambda(s)$, analytic in $\mathrm{Im}s < 0$. We show that $\Lambda_2(s)$ is the analytic continuation of $\Lambda_1(s)$ and that in fact $\Lambda_1(s) = \Lambda_2(s) \equiv 0$. For $s < s_1$

$$\Delta\Lambda(s) = [\Delta T(s) - \Delta B(s)]D(s) = 0. \tag{2.23}$$

For $s \geqslant s_1$

$$\begin{aligned}\Delta\Lambda(s) &= \Delta N(s) - \Delta B(s)D(s_-) - B(s_+)\Delta D(s) \\ &\quad - B(s)\rho(s)N(s) \\ &= \Delta_L T(s)D(s) - \Delta_L T(s)D(s) + B(s)\rho(s)N(s) \\ &\quad - B(s)\rho(s)N(s) \\ &= 0.\end{aligned} \tag{2.24}$$

The structure of $\Delta_L T(s)$, assumed in (2.2e), and the θ functions in $\Delta D(s) = -\rho(s)N(s)$ allowed us to replace $D(s_-)$ by $D(s)$ and $B(s_+)$ by $B(s)$ in (2.24). In the case without overlapping cuts the terms $\Delta N(s)$ and $-\Delta B(s)D(s_-)$ are separately zero; here they are nonzero but fortunately cancel. We see that $\Lambda(s)$ is analytic in the entire plane. Also it vanishes at infinity, as is clear from (2.3), (2.14), and (2.22). Thus $\Lambda(s) \equiv 0$ and (2.14) may be substituted into (2.22) to yield the integral equation

$$N(s) = B(s) + \frac{1}{\pi}\int_{s_1}^{\infty} \frac{B(s) - B(s')}{s - s'}\rho(s')N(s')ds'. \tag{2.25}$$

Thanks to the θ function in $\rho(s')$ the domains of the first and second rows of $N(s')$ in the integral are $[s_1, \infty)$ and $[s_2, \infty)$, respectively. Consequently, each $N_{ij}(s)$ is in a region of analyticity over the domain in which Eq. (2.25) is to be solved, as is seen from (2.11).

The derivation of the integral equation proceeds similarly in Case 2. The only change required is to account for the poles of $D(s)$. Referring to (2.22) and (2.19), we see that $\Lambda(s)$ inherits the poles and in fact

$$\Lambda(s) = \sum_{i=1}^{n_c} \frac{1}{s - \sigma_i}[T(\sigma_i) - B(\sigma_i)]C_i. \tag{2.26}$$

This equation yields the result

$$N(s) = B(s) + \sum_{i=1}^{n_c} \frac{1}{s - \sigma_i}[R_i + B(s)]C_i + \frac{1}{\pi}\int_{s_1}^{\infty} \frac{B(s) - B(s')}{s - s'}\rho(s')N(s')ds', \tag{2.27}$$

$$R_i = T(\sigma_i) - B(\sigma_i). \tag{2.28}$$

Henceforth we refer to the general equation (2.27), which includes (2.25) as the special case with $C_i = 0$.

The integral equation (2.27) is amenable to the Fredholm theory in an appropriate Banach space, as is shown in Appendix B. By the Fredholm Alternative Theorem,[20] the integral equation has a unique solution in the space considered, provided that the corresponding homogeneous equation has no nontrivial solution in that space. We shall suppose that the homogeneous equation in fact does not have a nontrivial solution, since the contrary case has not arisen, as far as we know, in realistic physical models. It does arise in the anomalous event of an "extinct bound state" as discussed by Atkinson and Halpern.[18] The assumption that there is no solution of the homogeneous equation allows us to rule out the possibility that a column of $\mathscr{D}(s)$ vanishes at infinity, as promised above. If $\mathscr{D}_{\cdot j}(s)$, the jth column of $\mathscr{D}(s)$, tends to the null vector as $|s| \to \infty$, then it has a Cauchy representation

$$\mathscr{D}_{\cdot j}(s) = -\frac{1}{\pi}\int_{s_1}^{\infty} \frac{\rho(s')\mathscr{N}_{\cdot j}(s')ds'}{s' - s}. \tag{2.29}$$

Owing to the lack of the usual unit matrix term on the right-hand side of (2.29), the corresponding integral equation for $\mathscr{N}_{\cdot j}(s)$ is homogeneous:

$$\mathscr{N}_{\cdot j}(s) = \frac{1}{\pi}\int_{s_1}^{\infty} \frac{B(s) - B(s')}{s - s'}\rho(s')\mathscr{N}_{\cdot j}(s')ds'. \tag{2.30}$$

Thus $\mathscr{N}_{\cdot j}(s) = 0$ and $\mathscr{D}_{\cdot j}(s) = 0$, contrary to the fact that $\mathscr{D}(s)$ in nonsingular. The derivation of (2.30), carried out as before by showing that $\Lambda(s) = 0$, fails if $T(s)$ has a bound-state pole. We defer the discussion of bound states.

Since we have ruled out the possibility that $\mathscr{D}_{\cdot j}(s)$ vanishes at infinity, we may conclude that the matrix $N(s)$ for any $T(s)$ obeying (2.2) satisfies (2.27). Furthermore, the properties (2.2) and (2.5) guarantee that $N(s) = T(s)D(s)$ lies in the Banach space used in the Fredholm theory of Appendix B. Thus, for a given $T(s)$, the matrix $N(s) = T(s)D(s)$ coincides with the unique solution of the integral equation provided by Fredholm theory.

When $B(s)$ rather than $T(s)$ is given, there is no *a priori* certainty that a corresponding satisfactory $T(s)$ exists. By the preceeding remarks we do know that if such a $T(s)$ exists, it must be obtainable in the form $N(s)D(s)^{-1}$, where $N(s)$ is a Fredholm solution of (2.27) for some choice of the parameters C_i and R_i, with an arbitrary choice of the σ_i; here $D(s)$ is given in terms of $N(s)$ by (2.19). We now show that the Fredholm solution $N(s)$ of (2.27) gives a $T(s)$ satisfying (2.2), provided that $\det D(s) \neq 0$ in the cut plane and that, when there are CDD poles, another minor condition holds (condition (2.34) below). This assertion holds for any choice of the parameters consistent with restrictions already laid down. Those restrictions, we recall, are that all parameters be real, that $\sigma_i < s_1$, $\sigma_i \neq \sigma_j$, $\det C_i = 0$, and that the R_i be positive-definite, symmetric matrices. The positive-definite character of R_i follows from its definition and (2.3), since

$T(s_+)\rho(s)T(s_-) = T(s_+)\rho(s)T(s_+)^\dagger$, where $\dagger$ denotes Hermitian adjoint. The matrices C_i and R_i entail only three real parameters each, since C_i is singular and R_i is symmetric.

If $N(s)$ is a solution of (2.27) and $D(s)$ is given by (2.19), we may write $T(s) = N(s)D(s)^{-1}$, the proposed solution of our problem, in the form

$$T(s) = B(s) + \left(\sum_{i=1}^{n_0} \frac{1}{s-\sigma_i} R_i C_i + \frac{1}{\pi}\int_{s_1}^{\infty} \frac{B(s')\rho(s')N(s')ds'}{s'-s}\right) D^{-1}(s). \tag{2.31}$$

This expression is derived by recognizing a term $B(s)D(s)$ on the right-hand side of (2.27). Since $\det D(s)$ is nonvanishing, it is clear that $T(s)$ has analyticity in accord with conditions (2.2), provided that it has no pole at $s = \sigma_i$. To demonstrate absence of a pole we write $D^{-1}(s) = \mathrm{cof}D(s)/\det D(s)$, and show by calculation, using $\det C_i = 0$, that

$$C_i \mathrm{cof} D(s) \simeq O(1), \quad s \to \sigma_i, \tag{2.32}$$

$$\det D(s) \sim \frac{a_i}{s-\sigma_i} + O(1), \quad s \to \sigma_i,$$

$$a_i = \sum_{j\neq 1} (C_{i11}C_{j22} + C_{i22}C_{j11} - C_{i12}C_{j21} - C_{j12}C_{i21}) + C_{i11}[1 + I_{22}(\sigma_i)] + C_{i22}[1 + I_{11}(\sigma_i)] - I_{12}(\sigma_i)C_{i21} - I_{21}(\sigma_i)C_{i12}, \tag{2.33}$$

where $I_{ij}(s)$ denotes the integral that appears in $D_{ij}(s)$. Thus formula (2.31) contains no pole provided that

$$a_i \neq 0. \tag{2.34}$$

Condition (2.34) is the extra requirement for existence of a solution in the presence of CDD poles, mentioned above.

Having proved analyticity, we have yet to show that (2.31) is properly unitary and symmetric. To check unitarity, we first calculate $\Delta N(s) = \Delta(T(s)D(s))$ from (2.31):

$$\begin{aligned}\Delta N(s) &= B(s_+)\Delta D(s) + \Delta B(s) D(s_-) + B(s)\rho(s)N(s) \\ &= -B(s)\rho(s)N(s) + \Delta_L T(s)D(s) + B(s)\rho(s)N(s) \\ &= \Delta_L T(s)D(s).\end{aligned} \tag{2.35}$$

The unitarity condition (2.2e) is now verified as follows:

$$\begin{aligned}T(s_+) - T(s_-) &= [N(s_+)D(s_+)^{-1}D(s_-) - N(s_-)]D(s_-)^{-1} \\ &= [N(s_+)D(s_+)^{-1}D(s_-) - N(s_+) \\ &\quad + 2i\Delta_L T(s)D(s)]D(s_-)^{-1} \\ &= N(s_+)D(s_+)^{-1}[D(s_-) - D(s_+)]D(s_-)^{-1} \\ &\quad + 2i\Delta_L T(s) \\ &= N(s_+)D(s_+)^{-1}\rho(s)N(s)D(s_-)^{-1} + 2i\Delta_L T(s) \\ &= T(s_+)\rho(s)T(s_-) + 2i\Delta_L T(s).\end{aligned} \tag{2.36}$$

As before, the prefactors $\rho(s)$ and $\Delta_L T(s)$ allowed us, on occasion, to replace $s_\pm$ by s. This calculation reveals a situation not present in the case with nonoverlapping cuts. Namely, $T(s)$ satisfies unitarity only if $N(s)$ satisfies the integral equation (2.27). In the nonoverlapping case $N(s)D(s)^{-1}$ is unitary, but in general not symmetric, for an arbitrary real matrix $N(s)$ such that the integral in $D(s)$ is well defined.

Symmetry of $T(s)$ in (2.31) is proved by the method of Bjorken and Nauenberg.[10] We examine the function

$$\Phi(s) = D^T(s)[T(s) - T^T(s)]D(s), \tag{2.37}$$

where the superscript T denotes transposition. Because of the definition (2.3) and the assumed symmetry (2.4a) of $B(s)$, it is clear that $\Phi(s)$ is analytic in the upper and lower half-planes, even though we have not assumed that $T(s)$ is analytic in a whole cut plane. We shall show that the discontinuity of $\Phi(s)$ over the real axis is zero and that $\Phi(s)$ has no pole at $s = \sigma_i$. Since $\Phi(s)$ vanishes at infinity, it must then be identically zero. The symmetry of T will follow, since we have assumed that $D(s)$ is nonsingular. For $s \leqslant s_1$, $\Delta\Phi(s)$ is obviously zero, since $\Delta D(s) = 0$ and $B(s) = B^T(s)$:

$$\begin{aligned}\Delta\Phi(s) &= D^T(s)\Delta[T(s) - T^T(s)]D(s) \\ &= D^T(s)\Delta[B(s) - B^T(s)]D(s) \\ &= 0, \quad s \leqslant s_1.\end{aligned} \tag{2.38}$$

For $s > s_1$ we apply (2.35) and find

$$\begin{aligned}\Delta\Phi(s) &= \Delta[D^T(s)N(s) - N^T(s)D(s)] \\ &= D^T(s_+)\Delta N(s) + \Delta D^T(s)N(s_-) \\ &\quad - N^T(s)\Delta D(s) - \Delta N^T(s)D(s_-) \\ &= D^T(s)\Delta_L T(s)D(s) - N^T(s)\rho(s)N(s) \\ &\quad + N^T(s)\rho(s)N(s) - D^T(s)\Delta_L T(s)D(s) \\ &= 0.\end{aligned} \tag{2.39}$$

The CDD poles in the factors $D^T(s)$ and $D(s)$ of (2.37) cancel. That is seen by introducing (2.31) and invoking the symmetry of R_i; the sum of the pole terms is

$$\sum_i (s-\sigma_i)^{-2} C_i^T (R_i - R_i^T) C_i = 0. \tag{2.40}$$

To show that $T(s)$ of (2.31) satisfies the bounds (2.2f) we refer to the Fredholm theory of Appendix B, which shows that the solution $N(s)$ of the integral equation obeys bounds the same as those of $T(s)$. If $I(s)$ denotes the integral appearing in $D(s)$, then Lemma 2 of Appendix A shows that $I(s_+)$ also obeys bounds like (2.2f). It follows that $|T(s_+)| \leqslant \kappa(\ln s)^{-\alpha}$. To verify the second inequality of (2.2f) we write, with $s \leqslant s'$,

$$\begin{aligned}N(s)D(s_+)^{-1} - N(s')D(s'_+)^{-1} &= [N(s) - N(s')]D(s_+)^{-1} + N(s')D(s_+)^{-1} \\ &\quad \times [D(s'_+) - D(s_+)]D(s'_+)^{-1}.\end{aligned} \tag{2.41}$$

When there are no CDD poles it is obvious that the required bound is satisfied for each of the terms on the right side. With poles, the only additional task is to demonstrate local Hölder continuity near the poles. That is easily done with the help of (2.32) and assumption (2.34).

We have finished the proof that $T(s)$ constructed from a solution of the integral equation (2.27) satisfies all of the conditions (2.2), provided only that $\det D(s) \neq 0$ in the cut plane and $a_i \neq 0$.

The question of how to verify in practice the condition $\det D(s) \neq 0$ arises. In the following section we describe a simple and practical method of verifying the condition, which involves computation in the physical region only. Fortunately, it is not necessary to search the complex plane for zeros of $\det D(s)$.

Suppose that we solve (2.27) with an arbitrary choice of the real, symmetric, positive-definite matrices R_i; let us de-

note these input parameters as R_i^{in}. If $T^{\text{out}}(s)$ denotes the amplitude constructed from (2.31) and the solution of (2.27), will it necessarily happen that $R_i^{\text{out}} = T^{\text{out}}(\sigma_i) - B(\sigma_i) = R_i^{\text{in}}$?

In general the answer is no, because it is always possible to change R_i^{in} without changing R_i^{out}. Since C_i is singular, it has a real left eigenvector v_i with eigenvalue zero:

$$v_i^T C_i = 0. \tag{2.42}$$

If we change R_i^{in} by adding to it the positive-definite symmetric dyadic $\lambda v_i v_i^T$, $\lambda > 0$, there is no resulting change in R_i^{out}, since R_i^{in} enters the equations for $N(s)$ and $T^{\text{out}}(s)$ only in the product $R_i^{\text{in}} C_i$. Furthermore, we may argue that in general

$$R_i^{\text{out}} = R_i^{\text{in}} + \mu v_i v_i^T, \tag{2.43}$$

where $\mu = \mu(R_i^{\text{in}})$ is a real scalar function of R_i^{in}. If we multiply (2.31) on the right by $D(s)$ and equate residues of the poles on either side of the equation, we find that

$$(R_i^{\text{out}} - R_i^{\text{in}}) C_i = 0. \tag{2.44}$$

Both columns of C_i are proportional to the same vector u_i, and $R_i^{\text{out}} - R_i^{\text{in}}$, being real and symmetric, has the representation

$$R_i^{\text{out}} - R_i^{\text{in}} = \sum_{j=1}^{2} \lambda_j w_j w_j^T. \tag{2.45}$$

By (2.42) and the orthogonality of the w_j we see that either $w_j^T u_i = 0$ or $\lambda_j = 0$ for each j, and that at most one of the λ_j is nonzero. If λ_1, say, is nonzero, then w_1 has the same direction as v_i (being orthogonal to u_i) and (2.43) follows. Since R_i^{out} is a nonlinear function of R_i^{in} (in the domain where it is not a constant function) the function $\mu(R_i^{\text{in}})$ is not a simple one.

How many arbitrary parameters are associated with each CDD pole? To answer this question we first note that the pole positions σ_i are not to be counted as free parameters. Suppose that we have constructed an amplitude $T(s)$ from (2.27) with input parameters σ_i, C_i, R_i. Recalling the derivation of (2.27), we see that the same $T(s)$ has a representation $T(s) = \hat{N}(s)\hat{D}(s)^{-1}$, where $\hat{D}(s)$ has new pole positions $\hat{\sigma}_i$ and new residues $\hat{C}_i$, and $\hat{N}(s)$ satisfies (2.27) with $(\sigma_i, C_i, R_i) \to (\hat{\sigma}_i, \hat{C}_i, \hat{R}_i)$. Thus a change in pole position σ_i may always be compensated by a change in C_i and R_i so as to yield the same amplitude $T(s)$. The essential parameters are three in C_i and three in R_i, but it must be remembered that there is a subspace in the space of R_i parameters on which $T(s)$ is constant; i.e., we may add any term of the form $\lambda v_i v_i^T$, $\lambda > 0$, to R_i without changing $T(s)$.

III. BOUND STATES, LEVINSON'S THEOREM, AND A TEST FOR THE PRESENCE OF GHOST POLES

Bound states seem not to occur in meson systems, but there is nevertheless a good technical reason to discuss them. The many-channel version of Levinson's theorem states that

$$(1/2i)\ln \det S(\infty) = -n_b + n_c, \tag{3.1}$$

where n_b is the number of bound-state poles, n_c the number of CDD poles, and S the scattering matrix

$$S(s) = 1 + 2i\rho^{1/2}(s)T(s_+)\rho^{1/2}(s). \tag{3.2}$$

The quantity $\ln \det S(\infty)$ is defined by considering $\ln \det S(s)$ as a continuous function of s, with $\ln \det S(s_1) = 0$, and taking the increment between $s = s_1$ and $s = \infty$. Our interest in bound states and the Levinson relation stems from the circumstance that "ghost" poles (spurious poles of the T matrix lacking a physical interpretation) are counted in Levinson's relation just as though they were bound-state poles. In a system with ghosts (3.1) is replaced by

$$(1/2i)\ln \det S(\infty) = -n_b - n_g + n_c, \tag{3.3}$$

where n_g is the number of ghost poles. In a calculation with the ND^{-1} method based on a specific model of $B(s)$, the number $\ln \det S(\infty)$ is computed easily in conjuction with the solution of the integral equation, n_c is an input parameter, and $n_b = 0$ is usually demanded by the physics of the situation. Thus we can determine the number of ghosts from (3.3), rather than by searching the complex plane for zeros of $\det D(s)$. Should bound states be allowed in the problem, their location and number are easily determined by looking for zeros of $\det D(s)$ on a small interval of the real axis.

Levinson's relation is true for any amplitude $T(s)$ satisfying conditions (2.2), provided that the homogeneous form of Eq. (2.27) has no nontrivial solution (in the space considered in Appendix B). Of course the latter condition is a restriction on $B(s)$ alone and it seems invariably to be met in realistic models. It is understood that the poles of $T(s)$, n_b in number, are all simple poles with factorized residues (i.e., each residue matrix is of rank 1). A proof of Levinson's relation, valid under the conditions stated here, is given in Sec. 5 of Ref. 11. The proof as written applies when the poles of $T(s)$ are at real points $\hat{s} < s_1$. One may also have ghost poles at complex points in conjugate pairs $(\hat{s}, \hat{s}^*)$. An extension of the argument of Ref. 11 is required in that case.

IV. SINGLE-CHANNEL PROBLEM WITH OVERLAPPING CUTS AND ABSORPTION AT THRESHOLD

In some phenomenological studies it may be more practical to treat only one channel explicitly, accounting for coupled channels by empirical absorption parameters. A simple extension of the single-channel N/D method with absorption[7] allows one to handle processes such as $K\bar{K} \to K\bar{K}$ and $N\bar{N} \to N\bar{N}$, which have absorption at threshold and overlapping cuts. The object is to construct unitary single-channel amplitudes of the form

$$T(s) = B_L(s) + \frac{1}{\pi}\int_{s_0}^{\infty} \frac{T(s'_+)q(s')T(s'_-)ds'}{s' - s} + \frac{1}{\pi}\int_{s_I}^{\infty} \frac{F(s')ds'}{s' - s}, \tag{4.1}$$

where the left cut part $B_L(s)$ and the absorption function $F(s)$ are given. We suppose that $B_L(s)$ has the properties of the function $B_{22}(s)$ of Sec. II; [s_0 in (4.1) is to be identified with s_2 in Sec. II]. The inelastic threshold s_I is assumed to be lower than the physical threshold s_0 and may be either to the left or to the right of the end of the left cut at s_L. With the channel considered labeled as the zeroth one, $F(s)$ is the inelastic part of the unitarity sum,

$$F(s) = \sum_{n \neq 0} T_{on}(s_+)\rho_n(s)T_{n0}(s_-), \tag{4.2}$$

where the functions $\rho_n(s)$ contain step functions to account for the closing of channels. The sum over n may actually include integrals if states with more than two particles are involved. We have $F(s)\geqslant 0$ even for $s<s_0$; since the T matrix is real-analytic and symmetric, $T_{n0}(s_-)=T_{0n}(s_+)^*$ even below the threshold of channel 0. For $s\geqslant s_0$, $F(s)$ is expressed in terms of the usual elasticity function $\eta(s)$:

$$F(s)=[1-\eta^2(s)]/4q(s), \quad s\geqslant s_0, \tag{4.3}$$

$$T(s)=[\eta(s)e^{2i\delta(s)}-1]/2iq(s), \quad s\geqslant s_0. \tag{4.4}$$

We suppose that $F(s)$ satisfies bounds like those on $B(s)$ in (2.4d). It then follows from (4.3) that $[1-\eta(s)]/q(s)$ satisfies such bounds as well, and in particular that $\eta(s)\to 1, s\to s_0+$.

In the N/D method with absorption,[7] the function $\mathscr{D}(s)$ is defined in terms of the real phase shift $\delta(s)$ of (4.4) by the expression (2.6). In the present extended method we use the same $\mathscr{D}(s)$, but use a $B(s)$ different from the usual one; namely,

$$B(s)=B_L(s)+\frac{1}{\pi}\int_{s_I}^{s_0}\frac{F(s')ds'}{s'-s}+\frac{1}{2\pi}\int_{s_0}^{\infty}\frac{1-\eta(s')}{q(s')(s'-s)}ds'. \tag{4.5}$$

In other words, we treat the part of the absorption cut between s_I and s_0 just as though it were a left cut contribution. The derivation of the integral equation then proceeds in the same way as in Ref. 7. In the case without CDD poles the equation reads

$$\eta(s)n(s)=\mathrm{Re}B(s)+\frac{1}{\pi}\int_{s_0}^{\infty}\frac{\mathrm{Re}B(s)-\mathrm{Re}B(s')}{s-s'}q(s')n(s')ds', \tag{4.6}$$

where $n(s)=-\mathrm{Im}D(s_+)/q(s)$. The amplitude is obtained in terms of $n(s)$ [which is not the same as the numerator function $N(s)$] by the formula

$$T(s)=B(s)+\frac{1}{\pi D(s)}\int_{s_0}^{\infty}\frac{\mathrm{Re}B(s')q(s')n(s')ds'}{s'-s}, \tag{4.7}$$

$$D(s)=1-\frac{1}{\pi}\int_{s_0}^{\infty}\frac{q(s')n(s')ds'}{s'-s}. \tag{4.8}$$

Each of the last two terms in (4.5), contributing to $\mathrm{Re}B(s)$, has a logarithmic singularity at $s=s_0$. The singularities of the two terms cancel, however, because $F(s)$ is Hölder-continuous and

$$F(s)=\frac{1-\eta^2(s)}{4q(s)}\sim\frac{1-\eta(s)}{2q(s)}, \quad s\to s_0+. \tag{4.9}$$

As a result $\mathrm{Re}B(s)$ is Hölder-continuous for $s\geqslant s_0$ and the integral equation (4.6) is of Fredholm type on the space of Appendix B, provided that $\eta(s)$ has no zero. A solution of the integral equation gives an amplitude (4.7) that is properly analytic and satisfies unitarity in the form

$$\mathrm{Im}T(s_+)=T(s_+)\theta(s-s_0)q(s)T(s_-)+F(s)+\Delta_L T(s), \quad s>s_I. \tag{4.10}$$

provided that $D(s)$ has no zero in the cut plane. As in Sec. III, a practical test for the presence of ghost zeros of $D(s)$ may be based on Levinson's relation, which in the present case holds in the form

$$\delta(\infty)=-\pi(n_b-n_c). \tag{4.11}$$

V. A SPECIAL CASE SOLVED BY GONZÁLEZ-ARROYO

We return to the two-channel problem of Sec. II and discuss a special case treated by González-Arroyo[4,3]; namely, the case in which the left cut parts of T_{11} and T_{12} vanish and nonrelativistic kinematics hold:

$$B_{11}(s)=B_{12}(s)=B_{21}(s)=0, \tag{5.1}$$

$$\rho_i(s)=\theta(s-s_i)(s-s_i)^{1/2}, \quad i=1,2. \tag{5.2}$$

Because $\rho_i(s)$ grows at infinity, we must assume that $B_{22}(s)$ vanishes more rapidly than does $B(s)$ of Sec. II. Instead of (2.4d), we take

$$|B_{22}(s)|\leqslant\kappa s^{-\alpha}, \quad s\geqslant s_2,$$

$$|B_{22}(s)-B_{22}(s')|\leqslant\kappa s^{-\alpha}\left|\frac{s-s'}{s}\right|^{\mu}, \quad s'\geqslant s\geqslant s_2,$$

$$\tfrac{1}{2}<\alpha<1, \quad 0<\mu<\tfrac{1}{2}. \tag{5.3}$$

For a given $B(s)$ satisfying (5.1) and (5.3), we seek the general $T(s)$ satisfying (2.2a), (2.2b), (2.2c), (2.2d), (2.2e), and bounds such as (5.3) instead of (2.2f). For such a $T(s)$ there is a $\mathscr{D}(s)$ satisfying (2.5a), (2.5b), (2.5c), (2.5d), and (2.5e) and a corresponding $D(s)$, as defined in (2.16), having the representation (2.19). Consequently, the integral equation (2.27) holds. The first row of the matrix equation is trivial, giving $N_{1j}(s)$ explicitly as a function of CDD parameters:

$$N_{1j}(s)=\sum_{i=1}^{n_c}\frac{(R_iC_i)_{1j}}{s-\sigma_i}, \quad j=1,2. \tag{5.4}$$

From the second row of the integral equation we have

$$N_{21}(s)=\sum_{i=1}^{n_c}\{[R_i+B(s)]C_i\}_{21}\frac{1}{s-\sigma_i}+\frac{1}{\pi}\int_{s_2}^{\infty}\frac{B_{22}(s)-B_{22}(s')}{s-s'}q_2(s')N_{21}(s')ds', \tag{5.5}$$

and

$$N_{22}(s)=B_{22}(s)+\sum_{i=1}^{n_c}\{[R_i+B(s)]C_i\}_{22}\frac{1}{s-\sigma_i}+\frac{1}{\pi}\int_{s_2}^{\infty}\frac{B_{22}(s)-B_{22}(s')}{s-s'}q_2(s')N_{22}(s')ds'. \tag{5.6}$$

The integral in the D matrix elements (2.16) corresponding to (5.4) may be evaluated to obtain

$$D_{1j}(s)=\delta_{1j}+\sum_{i=1}^{n_c}\{(C_i)_{1j}+(R_iC_i)_{1j}\times[(s_1-\sigma_i)^{1/2}+iq_1(s)]\}\frac{1}{s-\sigma_i}. \tag{5.7}$$

We suppose as in Sec. II that the homogeneous version of the matrix integral equation for $N(s)$, equivalent in the present case to the homogeneous version of the scalar equation (5.5), has no nontrivial solution. Then if there are no CDD poles, the integral equation (5.5) for $N_{21}(s)$ is homogeneous and has only the trivial solution $N_{21}(s)=0$. Without CDD poles we obtain only the trivial solution in which channel 2 is completely decoupled, and $T_{11}(s)=T_{12}(s)=T_{21}(s)=0$. Thus

the solution of González-Arroyo necessarily entails CDD poles as defined in the two-channel formalism. Since González-Arroyo reduced his problem to a one-channel case by a special device, this fact was not previously apparent.

To reduce the problem to a one-channel case through our formalism we circumvent solution of Eq. (5.5) and require solution of (5.6) alone. Accordingly, we suppose that the solution of (5.6), in a Banach space appropriate to conditions (5.3), is given; see the remark at the end of Appendix B. The solution $N_{22}(s)$ obeys conditions like (5.3); of course, the same is true of the solution $N_{21}(s)$ of (5.5). Because the inhomogeneous term in (5.5) is $O(s^{-1})$, it is possible to show that $N_{21}(s) = O(s^{-1})$ and $D_{21}(s_+) = O(s^{-1/2})$. The proof is done by showing that the integral operator "improves" the asymptotic behavior of $N_{21}(s)$. That is, if $N_{21}(s) = O(s^{-\alpha})$, then the integral is $O(s^{-2\alpha+1/2})$. By iteration of this argument one eventually gets $N_{21}(s) = O(s^{-1})$.

We exploit the symmetry of the T matrix, writing $T(s) = \hat{T}(s)/\det D(s)$ and

$$\begin{aligned}\hat{T}_{12}(s) &= -N_{11}(s)D_{12}(s) + N_{12}(s)D_{11}(s) = \hat{T}_{21}(s) \\ &= -N_{22}(s)D_{21}(s) + N_{21}(s)D_{22}(s). \end{aligned} \tag{5.8}$$

From (5.4) and (5.7) we may compute $\hat{T}_{12}(s)$; it is just a rational function, since the terms from the imaginary parts of $D_{12}(s)$ and $D_{11}(s)$ cancel. With that observation and a knowledge of $N_{22}(s)$ and $D_{22}(s)$ we can use (5.8) to find the general form of $N_{21}(s)$ and $D_{21}(s)$. The rational function $\hat{T}_{12}(s)$ is

$$\begin{aligned}\hat{T}_{12}(s) &= -N_{11}(s)\mathrm{Re}D_{12}(s) + N_{12}(s)\mathrm{Re}D_{11}(s) \\ &= -\sum_i \frac{(R_iC_i)_{11}}{s-\sigma_i} \sum_j [(C_j)_{12} + (s_1-\sigma_j)^{1/2}(R_jC_j)_{12}] \\ &\quad \times \frac{1}{s-\sigma_j} + \sum_i \frac{(R_iC_i)_{12}}{s-\sigma_i} \\ &\quad \times \left(1 + \sum_j [(C_j)_{11} + (s_1-\sigma_i)^{1/2}(R_jC_j)_{11}] \frac{1}{s-\sigma_j}\right) \\ &= \frac{\Phi(s)}{P(s)}, \quad s \geqslant s_1, \end{aligned} \tag{5.9}$$

where

$$P(s) = \prod_{i=1}^{n_c} (s-\sigma_i) \tag{5.10}$$

and $\Phi(s)$ is a polynomial of degree not greater than $n_c - 1$. The second-order poles, corresponding to $i = j$ in the sums of (5.9), cancel because of the condition $\det C_i = 0$.

Equation (5.8) may be construed as a Riemann–Hilbert boundary-value problem for determination of $D_{21}(s)$. Since $N_{2j}(s) = -\mathrm{Im}D_{2j}(s)/q_2(s)$, the real part of (5.8) reads

$$\begin{aligned} q_2(s)\frac{\Phi(s)}{P(s)} &= -\mathrm{Im}D_{21}(s_+)\mathrm{Re}D_{22}(s_+) + \mathrm{Re}D_{21}(s_+)\mathrm{Im}D_{22}(s_+) \\ &= [D_{22}(s_+)D_{21}(s_-) - D_{22}(s_-)D_{21}(s_+)]/2i. \end{aligned} \tag{5.11}$$

By the rearrangement displayed in the second line of (5.11), the Riemann–Hilbert problem[17] is transformed to an inhomogeneous Hilbert problem[17]:

$$D_{21}(s_-) = \frac{D_{22}(s_-)}{D_{22}(s_+)} D_{21}(s_+) + \frac{2iq_2(s)\Phi(s)}{P(s)D_{22}(s_+)}. \tag{5.12}$$

To solve the Hilbert problem we invoke the ubiquitous phase integral

$$d(s) = \exp\left(-\frac{s}{\pi}\int_{s_2}^{\infty} \frac{\delta(s')ds'}{s'(s'-s)}\right), \tag{5.13}$$

$$e^{-2i\delta(s)} = D_{22}(s_+)/D_{22}(s_-). \tag{5.14}$$

Notice that $\delta(s)$ is the phase shift for the amplitude $N_{22}(s)/D_{22}(s)$, which obeys elastic unitarity and is not to be confused with the channel-2 scattering amplitude $T_{22}(s)$. It is easy to see that $D_{22}(s)/d(s)$, being real for $s > s_2$, is a rational function with poles only at $s = \sigma_i$; we may write

$$D_{22}(s) = \frac{\Psi(s)}{P(s)} d(s), \tag{5.15}$$

where $\Psi(s)$ is a polynomial of degree n_z equal to the number of zeros of $D_{22}(s)$. Nothing prevents $D_{22}(s)$ from having zeros, in general, since poles of $N_{22}(s)/D_{22}(s)$ are not poles of the full T matrix. We have $\delta(\infty) = \pi(n_c - n_z)$, since $D_{22}(s_+) \sim 1$ and $d(s_+) \sim \delta(\infty)/\pi$, $s \to +\infty$. Now substitute $d(s_-)/d(s_+)$ for $D_{22}(s_-)/D_{22}(s_+)$ in (5.12) and use (5.15) to obtain

$$\begin{aligned}&[D_{21}(s_+)/d(s_+) - D_{21}(s_-)/d(s_-)]/2i \\ &\quad = -\frac{q_2(s)}{|d(s_+)|^2}\frac{\Phi(s)}{\Psi(s)}, \quad s \geqslant s_2. \end{aligned} \tag{5.16}$$

Thus we have the discontinuity of $D_{21}(s)/d(s)$ over the cut $[s_2, \infty)$ and it is real as required. In addition we know that $D_{21}(s)$ is analytic in the plane with cut $[s_2, \infty)$, except for simple poles at $s = \sigma_i$, and that it vanishes at infinity: $D_{21}(s_+) = O(s^{-1/2})$. Since $d(s) \sim s^{n_c - n_z}$ at infinity, $D_{21}(s)/d(s)$ obeys an unsubtracted dispersion relation if $n_c \geqslant n_z$. The right-hand side of (5.16) is $O(s^{-1/2+n_z-n_c})$ since degree $(\Phi) \leqslant n_c - 1$. For $n_c \geqslant n_z$ we have the representation

$$\begin{aligned} D_{21}(s) = d(s)\Bigg(&-\frac{1}{\pi}\int_{s_2}^{\infty} \frac{q_2(s')}{|d(s'_+)|^2}\frac{\Phi(s')}{\Psi(s')}\frac{ds'}{s'-s} \\ &+ \sum_{i=1}^{n_c} \frac{(C_i)_{21}}{d(\sigma_i)}\frac{1}{s-\sigma_i}\Bigg). \end{aligned} \tag{5.17}$$

Note that $d(s)$ may be redefined through multiplication by a constant, but that (5.17) is invariant to such a change $(d \to \lambda d, \Psi \to \lambda^{-1}\Psi)$. For $n_z - n_c = n_s \geqslant 1$ we must introduce n_s subtractions and replace (5.17) by the formula

$$\begin{aligned} D_{21}(s) = Q(s)d(s)\Bigg(&\sum_{i=1}^{n_s} \frac{D_{21}(t_i)}{d(t_i)}\frac{1}{Q'(t_i)(s-t_i)} \\ &- \frac{1}{\pi}\int_{s_2}^{\infty} \frac{q_2(s')}{|d(s'_+)|^2}\frac{\Phi(s')}{\Psi(s')}\frac{ds'}{Q(s')(s'-s)} \\ &+ \sum_{i=1}^{n_c} \frac{(C_i)_{21}}{Q(\sigma_i)d(\sigma_i)}\frac{1}{s-\sigma_i}\Bigg), \end{aligned} \tag{5.18}$$

where $Q(s)$ is a polynomial with distinct roots $t_i < s_2$, none of the t_i coinciding with a σ_j. The function $N_{21}(s) = -\mathrm{Im}D_{21}(s_+)/q_2(s)$ may be computed from (5.8), (5.9), or by taking the discontinuity of (5.17) or (5.18). By either method we find

$$N_{21}(s) = N_{22}(s)\frac{D_{21}(s)}{D_{22}(s)} + \frac{\Phi(s)}{P(s)D_{22}(s)}. \tag{5.19}$$

The representation (5.17) of $D_{21}(s)$ is determined by $(C_i)_{21}$ and the functions $\Phi(s)$, $D_{22}(s)$. The latter are in turn determined by the matrices C_i and R_i, through (5.9), (5.6), and (2.19). Thus we have determined $D_{21}(s)$ in terms of the input parameters C_i and R_i without having solved the integral equation (5.5), provided that $n_c - n_z \geqslant 0$. On the other hand, we can assert that $N_{21}(s)$ as determined by (5.19) and (5.17) in fact solves the integral equation, since the matrix $D(s)$ constructed from a solution of (2.27) satisfies all the requirements that led to the unique function (5.17).

In $n_c - n_z < 0$ (which implies that $n_z \geqslant 2$), then $D_{21}(z)$ is not determined uniquely by the above considerations because of the unknown subtraction constants $D_{21}(t_i)/d(t_i)$ in (5.18). Consequently, we cannot be sure that the corresponding $N_{21}(s)$ satisfies (5.5). Nevertheless, we can demonstrate that $T(s)$ constructed from (5.18), (5.19), and the other previously determined elements of $D(s)$ and $N(s)$ actually is a solution of our problem for arbitrary subtraction constants (provided, as usual, that $\det D(s) \neq 0$ in the cut plane). It then follows that $N_{21}(s)$ constructed from (5.19) and (5.18) with arbitrary subtraction constants satisfies (5.5), but with a value of the parameter $(R_i C_i)_{21}$ that may only be computed *a posteriori* as $\{[T(\sigma_i) - B(\sigma_i)]C_i\}_{21}$ from the $T(s)$ constructed. To show that $T(s)$ [constructed with (5.18), (5.19), and arbitrary subtraction constants] is a solution of our problem we have only to verify unitarity, since proper analyticity is evident and symmetry of $T(s)$ was ensured through the use of (5.8). Unitarity follows from the calculation (2.36) if (2.35) holds. The first row of (2.35) is trivial because of (5.4), and we have $\Delta N_{22}(s) = [\Delta_L T(s) D(s)]_{22} = \Delta_L T_{22}(s) D_{22}(s)$ as is usual for a single-channel N/D problem. To finish the proof of (2.35), one has only to show that $\Delta N_{21}(s) = \Delta_L T_{22}(s) D_{21}(s)$, and that is easily done by (5.19) and (5.18). For $s > s_2$, $\Delta N_{21}(s) = 0$ because $N_{21}(s)$ is real, being the discontinuity of the product of two real-analytic functions displayed in (5.18). For $s < s_2$, (5.19) gives

$$\Delta N_{21}(s) = \Delta N_{22}(s) \frac{D_{21}(s)}{D_{22}(s)} = \Delta_L T_{22}(s) D_{21}(s). \tag{5.20}$$

To make contact with the solution of González-Arroyo, we look at the K matrix[19]

$$K(s) = N(s)[\mathrm{Re}D(s)]^{-1}, \quad s \geqslant s_2. \tag{5.21}$$

The solution in question is such that the element $K_{22}(s)$ is equal to the K matrix for the "decoupled" channel-2 problem, namely, $N_{22}(s)/\mathrm{Re}D_{22}(s)$:

$$\begin{aligned} K_{22}(s) &= \frac{(N_{22}/\mathrm{Re}D_{22})\mathrm{Re}D_{11} - (N_{21}/\mathrm{Re}D_{22})\mathrm{Re}D_{12}}{\mathrm{Re}D_{11} - (\mathrm{Re}D_{21}/\mathrm{Re}D_{22})\mathrm{Re}D_{12}} \\ &= N_{22}(s)/\mathrm{Re}D_{22}(s). \end{aligned} \tag{5.22}$$

Condition (5.22) can be met in only two ways: either $\mathrm{Re}D_{12}(s) = 0$ or $N_{21}(s)\mathrm{Re}D_{22}(s) - N_{22}(s)\,\mathrm{Re}D_{21}(s) = 0$. The latter equation implies that $K_{12}(s) = K_{21}(s) = 0$, however, from which it follows that $T_{12}(s) = T_{21}(s) = 0$; i.e., that the solution is trivial. We must take $\mathrm{Re}D_{12}(s) = 0$ and by (5.7) we see that the González–Arroyo solution corresponds to a particular choice of CDD parameters such that

$$(C_i)_{12} = 0, \quad (R_i C_i)_{12} = 0. \tag{5.23}$$

With $\mathrm{Re}D_{12}(s) = 0$ one has

$$K_{12}(s) = K_{21}(s) = N_{12}(s)/\mathrm{Re}D_{22}(s), \tag{5.24}$$

and

$$K_{11}(s) = \frac{1}{\mathrm{Re}D_{11}(s)} \left(N_{11}(s) - N_{12}(s) \frac{\mathrm{Re}D_{21}(s)}{\mathrm{Re}D_{22}(s)} \right). \tag{5.25}$$

According to (5.4) and (5.15), we may write (5.24) in the form

$$K_{12}(s) = \frac{\Psi_{12}(s)}{\Psi(s)} \frac{1}{\mathrm{Re}d(s)}, \tag{5.26}$$

where $\Psi_{12}(s) = N_{12}(s)P(s)$ is a polynomial of degree not greater that $n_c - 1$. González-Arroyo has

$$\begin{aligned} K_{12}(s) &= [1 - iq_2(s)K_{22}(s)] \frac{\chi^{(0)}(s)}{d(s_+)} \\ &= \frac{D_{22}(s_+)}{\mathrm{Re}D_{22}(s)} \frac{\chi^{(0)}(s)}{d(s_+)} = \frac{\chi^{(0)}(s)}{\mathrm{Re}d(s)}, \end{aligned} \tag{5.27}$$

where $\chi^{(0)}(s)$ is a rational function that is $O(s^{n_c - 1 - n_z})$ at infinity and has poles at the zeros of $D_{22}(s)$ [i.e., of $\Psi(s)$], in agreement with our function $\Psi_{12}(s)/\Psi(s)$ of (5.26). The argument of Ref. 4 seems to allow poles of $\chi^{(0)}(s)$ at other points as well, but our generally valid expression (5.26) shows that additional poles are not possible: we have $\chi^{(0)}(s) = \Psi_{12}(s)/\Psi(s)$ with poles only at the zeros of $D_{22}(s)$.

Next let us evaluate (5.25) using expression (5.17). With the help of (5.15) and (5.14) we get

$$\begin{aligned} K_{11}(s) = \Bigg[& N_{11}(s) + \frac{\Psi_{12}(s)}{\Psi(s)} \\ & \times \left(\frac{P}{\pi} \int_{s_2}^{\infty} \frac{q_2(s')}{|d(s'_+)|^2} \frac{\Phi(s')}{\Psi(s')} \frac{ds'}{s' - s} \right. \\ & \left. - \sum_{i=1}^{n_c} \frac{(C_i)_{21}}{d(\sigma_i)} \frac{1}{s - \sigma_i} + \frac{q_2(s)\tan\delta(s)\Phi(s)}{|d(s_+)|^2 \Psi(s)} \right) \Bigg] \\ & \times \frac{1}{\mathrm{Re}D_{11}(s)}, \quad s \geqslant s_2. \end{aligned} \tag{5.28}$$

The corresponding formula in Ref. 4 is

$$\begin{aligned} K_{11}(s) = {} & h^{(0)}(s) + \frac{P}{\pi} \int_{s_2}^{\infty} \frac{q_2(s')}{|d(s'_+)|^2} \left(\frac{\Psi_{12}(s')}{\Psi(s')} \right)^2 \frac{ds'}{s' - s} \\ & + \frac{q_2(s)\tan\delta(s)}{|d(s_+)|^2} \left(\frac{\Psi_{12}(s)}{\Psi(s)} \right)^2, \quad s \geqslant s_2, \end{aligned} \tag{5.29}$$

where $h^{(0)}(s)$ is a rational function that has poles at the zeros of $\Psi(s)$ and is $O(s^{-1})$ at infinity. In order that the terms proportional to $\tan\delta(s)$ in (5.28) and (5.29) agree, it is necessary that $\Phi(s) = \Psi_{12}(s)$ and $\mathrm{Re}D_{11}(s) = 1$. According to (5.9) and the condition $\mathrm{Re}D_{12}(s) = 0$ already imposed, $\Phi(s) = \Psi_{12}(s)$ follows from $\mathrm{Re}D_{11}(s) = 1$. By (5.7) the latter is true if and only if

$$(C_i)_{11} = 0, \quad (R_i C_i)_{11} = 0. \tag{5.30}$$

With $\Phi(s) = \Psi_{12}(s)$ we still have to resolve the discrepancy between the integrals that appear in (5.28) and (5.29). Consider the function

$$f(s) = \frac{\Psi_{12}(s)}{\Psi(s)} \frac{1}{\pi} \int_{s_2}^{\infty} \left[\frac{q_2(t)}{|d(t_+)|^2} \frac{\Psi_{12}(t)}{\Psi(t)} \right] \frac{dt}{t - s}. \tag{5.31}$$

The bracketed expression is $O(t^{-n_c+n_z-1/2})$ at infinity and we are assuming that $n_c \geqslant n_z$; hence the integral is $O(s^{-1/2})$. The factor in front of the integral is $O(s^{n_c-n_z-1})$. We may write a dispersion relation for $f(s)$ with n subtractions, where

$$n = \begin{cases} 0, & n_c = n_z, \quad n_z + 1, \\ n_c - n_z - 1, & n_c - n_z > 1. \end{cases} \tag{5.32}$$

Namely,

$$f(s) = \sum_{m=0}^{n-1} f^{(m)}(0)s^m + s^n\left[\sum_{i=1}^{n_c} \frac{a_i}{s-t_i} + \frac{1}{\pi}\int_{s_z}^{\infty}\left[\frac{q_2(t)}{|d(t_+)|^2}\left(\frac{\Psi_{12}(t)}{\Psi(t)}\right)^2\right]\frac{dt}{t^n(t-s)}\right], \tag{5.33}$$

where the sum over i is due to the poles of $1/\Psi(s)$ at points t_i, assumed distinct. The bracketed factor in the integrand is $O(t^{-3/2})$ so that we can remove all n subtractions in the integral by iteration of the identity

$$\frac{s^n}{(s-t)t^n} = \frac{s^{n-1}}{t^{n-1}}\left(\frac{1}{s-t} + \frac{1}{t}\right). \tag{5.34}$$

Thus

$$\begin{aligned} \mathrm{Re} f(s) &= \frac{\Psi_{12}(s)}{\Psi(s)}\frac{P}{\pi}\int_{s_z}^{\infty}\frac{q_2(t)}{|d(t_+)|^2}\frac{\Psi_{12}(t)}{\Psi(t)}\frac{dt}{t-s} \\ &= \frac{P}{\pi}\int_{s_z}^{\infty}\frac{q_2(t)}{|d(t_+)|^2}\left(\frac{\Psi_{12}(t)}{\Psi(t)}\right)^2\frac{dt}{t-s} + R(s), \end{aligned} \tag{5.35}$$

where

$$R(s) = \sum_{m=0}^{n-1} f^{(m)}(0)s^m + s^n\sum_{i=1}^{n_c}\frac{a_i}{s-t_i} + \sum_{m=0}^{n-1} s^m\frac{1}{\pi}\int_{s_z}^{\infty}\frac{q_2(t)}{|d(t_+)|^2}\left(\frac{\Psi_{12}(t)}{\Psi(t)}\right)^2\frac{dt}{t^{m+1}}. \tag{5.36}$$

We see that the integrals in (5.28) and (5.29) differ by a rational function that has poles at the t_i and which in general is $O(s^{n-1})$ at infinity.

Finally, in order that (5.28) and (5.29) be compatible, the rational function

$$N_{11}(s) + R(s) - \frac{\Psi_{12}(s)}{\Psi(s)}\sum_{i=1}^{n_c}\frac{(C_i)_{21}}{d(\sigma_i)}\frac{1}{s-\sigma_i} \tag{5.37}$$

must have the properties required of $h^{(0)}(s)$. If $n_c = n_z$, $n_z + 1$, then $R(s) = O(s^{-1})$ and all terms in (5.37) are $O(s^{-1})$ at infinity as required. Otherwise the second and third terms of (5.37) must cancel appropriately at infinity. González-Arroyo tacitly assumed, in fact, that $n_z = 0$. With that assumption we get a solution of his form when $n_c = 1$ and the CDD residues satisfy conditions (5.23) and (5.30). Even though the solution of González–Arroyo contains arbitrary rational functions, it is far from being the general solution of the problem posed.

VI. REMARK ON A UNITARY PARAMETRIZATION SUGGESTED BY YNDURÀIN

We have shown how to construct properly analytic amplitudes satisfying the unitarity equation (2.2e), but the construction has the disadvantage of requiring the solution of an integral equation. For phenomenology it would be useful to have a parametrization of $T(s)$, analogous to the usual K matrix parametrization, that would automatically satisfy (2.2e). Yndurain[1] has proposed a parametrization which has the required property in the region $s_1 \leqslant s \leqslant s_2$. Define $\bar{T}(s)$ such that $\bar{T}_{ij}(s) = T_{ij}(s)$, except for $i = j = 2$, and

$$\bar{T}_{22}(s) = T_{22}(s) - \frac{1}{\pi}\int_0^{s_1}\frac{\phi(s')ds'}{s'-s}. \tag{6.1}$$

There is nothing special about the lower limit 0 in the integral; any lower limit less than s_1 will do. Define a matrix $\bar{M}(s)$, which is related to $\bar{T}(s)$ in the way that $M(s) = K^{-1}(s)$ is related to $T(s)$:

$$\bar{T}(s)^{-1} = \bar{M}(s) - i\rho(s). \tag{6.2}$$

Now we may show that the unitarity condition (2.2e) is equivalent to the reality condition $\bar{M}(s) = \bar{M}(s)^*$ in the region $s_1 \leqslant s \leqslant s_2$. Let us consider the region $s \geqslant s_1$, supposing that $\bar{M}(s)$ is real in that region. We write

$$\begin{aligned} T(s) &= \bar{T}(s) + \hat{T}(s), \\ \hat{T}(s) &= \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{\pi}\int_0^{s_1}\frac{\phi(s')ds'}{s'-s} \end{bmatrix}. \end{aligned} \tag{6.3}$$

Since reality of $\bar{M}(s)$ implies that

$$\Delta\bar{T}(s) = \bar{T}(s_+)\rho(s)\bar{T}(s_-), \tag{6.4}$$

we have

$$\Delta T(s) = \bar{T}(s_+)\rho(s)\bar{T}(s_-) + \Delta_L T(s). \tag{6.5}$$

Also,

$$T(s_+)\rho(s)T(s_-) = \bar{T}(s_+)\rho(s)\bar{T}(s_-) + U(s), \tag{6.6}$$

where

$$\begin{aligned} U_{ij} &= \bar{T}_{-i1}\rho_1\hat{T}_{+1j} + \bar{T}_{-i2}\rho_2\hat{T}_{+2j} + \hat{T}_{-i1}\rho_1\bar{T}_{+1j} \\ &\quad + \hat{T}_{-i2}\rho_2\bar{T}_{+2j} + \hat{T}_{-i1}\rho_1\hat{T}_{+1j} + \hat{T}_{-i2}\rho_2\hat{T}_{+2j} \\ &= \rho_2(\bar{T}_{-i2}\hat{T}_{+2j} + \hat{T}_{-i2}\bar{T}_{+2j} + \hat{T}_{-i2}\hat{T}_{+2j})\delta_{i2}\delta_{j2}. \end{aligned} \tag{6.7}$$

Since $U(s) = 0$ for $s \leqslant s_2$, Eq. (6.5) is indeed the unitarity equation for $s_1 \leqslant s \leqslant s_2$.

For $s > s_2$, however, unitarity is not equivalent to $\bar{M}(s)$ being real, since $U_{22}(s) \neq 0$ in that region. Indeed, unitarity for $s > s_2$ is equivalent to $M(s) = K^{-1}(s)$ being real, where

$$T(s)^{-1} = M(s) - i\rho(s). \tag{6.8}$$

The relation between M and $\bar{M}$ is as follows:

$$[1 + (\bar{M}_{22} - i\rho_2)\hat{T}_{22}]\begin{pmatrix} M_{11} - i\rho_1 & M_{12} \\ M_{21} & M_{22} - i\rho_2 \end{pmatrix} = \begin{pmatrix} [1 + (\bar{M}_{22} - i\rho_2)\hat{T}_{22}](\bar{M}_{11} - i\rho_1) - \bar{M}_{12}^2\hat{T}_{22} & \bar{M}_{12} \\ \bar{M}_{12} & \bar{M}_{22} - i\rho_2 \end{pmatrix}. \tag{6.9}$$

Clearly $\bar{M}(s)$ does not change continuously into $M(s)$ at $s = s_2$; rather, at $s = s_2$ we have

$$M = \bar{M} - \frac{\hat{T}_{22}}{1 + \bar{M}_{22}\hat{T}_{22}} \begin{pmatrix} \bar{M}_{12}^2 & \bar{M}_{12}\bar{M}_{22} \\ \bar{M}_{12}\bar{M}_{22} & \bar{M}_{22}^2 \end{pmatrix}. \quad (6.10)$$

The matrices $M(s)$ and $\bar{M}(s)$ are two different analytic functions that one would try to represent in terms of a few empirical parameters so as to meet the following conditions:

(i) $\bar{M}(s) = \bar{M}(s)^*$, $s_1 < s < s_2$;

(ii) $M(s) = M(s)^*$, $s > s_2$;

(iii) $M(s)$ and $\bar{M}(s)$ are related by Eq. (6.9), $s \geqslant s_1$.

(iv) The analyticity properties of $M(s)$ and $\bar{M}(s)$ should reflect to a reasonable extent the correct analyticity properties of $T(s)$, especially the nearby singularities corresponding to the principal particle exchanges.

It seems rather difficult to satisfy all of these requirements simultaneously; in particular it seems hard to satisfy (iii) in such a way that (i) and (ii) would also hold. We would expect Yndurain's proposal to be rather limited in usefulness. The only alternative that we can think of, short of solving the integral equation (2.27), is to make a pole approximation for $B(s)$. As is well known, the kernel of the equation is then separable and solution of the equation is reduced to quadratures and solution of algebraic equations. Unfortunately, for a realistic representation of $B(s)$ one usually needs so many poles that the resulting formulas are not very illuminating.

ACKNOWLEDGMENTS

We thank Dr. González-Arroyo for copies of unpublished work. R. Warnock thanks the Theoretical Physics Group of Lawrence Berkeley Laboratory for kind hospitality.

APPENDIX A: ESTIMATES OF PRINCIPAL VALUE INTEGRALS

We are concerned with the asymptotic behavior and continuity properties of principal value integrals of the form

$$g(s) = P\int_{s_0}^{\infty} \frac{f(t)dt}{t - s}. \quad (A1)$$

In the following a, θ, and δ are fixed positive constants, and κ is "some positive constant" which is understood to have different values in different inequalities.

Lemma 1: Suppose that $f(t)$ obeys the conditions

$$|f(t)| \leqslant \frac{\kappa}{t^a}\left(\frac{t - s_0}{t}\right)^{\theta}, \quad (A2a)$$

$$|f(t_1) - f(t_2)| \leqslant \frac{\kappa}{t_1^a}\left(\frac{t_2 - s_0}{t_2}\right)^{\theta - \delta}\left|\frac{t_1 - t_2}{t_1}\right|^{\delta}, \quad t_2 \geqslant t_1, \quad a + \delta < 1, \quad \theta > \delta. \quad (A2b)$$

Then the integral $g(s)$ of (A1) is such that

$$|g(s)| \leqslant \kappa/s^a; \quad (A3a)$$

$$|g(s_1) - g(s_2)| \leqslant \frac{\kappa}{s_1^a}\left|\frac{s_1 - s_2}{s_1}\right|^{\delta}, \quad s_2 \geqslant s_1. \quad (A3b)$$

Lemma 2: Suppose that $f(t)$ obeys the conditions

$$|f(t)| < \frac{\kappa}{(\ln t)^a}\left(\frac{t - s_0}{t}\right)^{\theta}; \quad (A4a)$$

$$|f(t_1) - f(t_2)| \leqslant \frac{\kappa}{(\ln t_1)^a}\left(\frac{t_2 - s_0}{t_2}\right)^{\theta - \delta}\left|\frac{t_1 - t_2}{t_1}\right|^{\delta}, \quad t_2 \geqslant t_1, \quad a > 1, \quad \theta > \delta. \quad (A4b)$$

Then the integral $g(s)$ of (A1) is such that

$$|g(s)| \leqslant \kappa/(\ln s)^{a-1}; \quad (A5a)$$

$$|g(s_1) - g(s_2)| \leqslant \frac{\kappa}{(\ln s_1)^{a-1}}\left|\frac{s_1 - s_2}{s_1}\right|^{\delta}, \quad s_2 \geqslant s_1. \quad (A5b)$$

We give a proof of Lemma 2; A proof of Lemma 1 follows the same lines, but is somewhat easier. To verify (A5a), we write

$$g = P\int_4^{\mu s} + \int_{\mu s}^{\infty} = g_1 + g_2, \quad \mu > 1, \quad (A6)$$

and majorize g_2 immediately:

$$|g_2| \leqslant \kappa\int_{\mu s}^{\infty} \frac{dt}{\ln^a t\,(t - s)} \leqslant \kappa\int_{\mu s}^{\infty} \frac{dt}{t \ln^a t} \leqslant \frac{\kappa}{\ln^{a-1} s}. \quad (A7)$$

For g_1 we use the identity

$$g_1(s) = \int_{s_0}^{\mu s} \frac{f(t) - f(s)}{t - s}\,dt + f(s)\frac{(\mu - 1)s}{s - s_0}. \quad (A8)$$

By introducing (A4b) and (A4a) in the first and second terms of (A8), respectively, we see that $g_1(s)$ is bounded at small s, say $s < 2s_0$, and consequently (A5a) holds at small s. For $s > 2s_0$ the logarithmic term in (A8) clearly satisfies (A5a). The other term is decomposed and bounded as follows, with $\frac{1}{2} < \lambda < 1$:

$$\begin{aligned}\left(\int_{s_0}^{\lambda s} + \int_{\lambda s}^{\mu s}\right)\left|\frac{f(t) - f(s)}{t - s}\right| dt &\leqslant \kappa\int_{s_0}^{\lambda s} \frac{dt}{t^{\delta}\ln^a t}\frac{1}{|\lambda s - s|^{1-\delta}} + \kappa\int_{\lambda s}^{\mu s} \frac{dt}{s^{\delta}\ln^a s}\frac{1}{|t - s|^{1-\delta}} \\ &\leqslant \frac{\kappa}{s^{1-\delta}}\int_{s_0}^{\lambda s} \frac{dt}{t^{\delta}\ln^a t} + \frac{\kappa}{\ln^a s}\int_{\lambda}^{\mu} \frac{du}{|u - 1|^{1-\delta}} \\ &\leqslant \frac{\kappa}{\ln^a s}. \end{aligned} \quad (A9)$$

Thus (A5a) is proved and we see that the dominant part of $g(s)$ at large s is from the tail of the integral, $g_2(s)$.

To establish (A5b) we split the integral as follows:

$$g = P\int_{s_0}^{2s_2} + \int_{2s_2}^{\infty} = g_1 + g_2. \quad (A10)$$

The bound of g_2 is easily obtained:

$$\begin{aligned}|g_2(s_1) - g_2(s_2)| &\leqslant \kappa|s_1 - s_2|\int_{2s_2}^{\infty} \frac{dt}{\ln^a t\,(t - s_1)(t - s_2)} \\ &\leqslant \kappa|s_1 - s_2|\int_{2s_2}^{\infty} \frac{dt}{t^2\ln^a t} \leqslant \frac{\kappa|s_1 - s_2|}{s_2\ln^a s_2} \\ &\leqslant \frac{\kappa}{\ln^a s_2}\left|\frac{s_1 - s_2}{s_1}\right| \leqslant \frac{\kappa}{\ln^a s_1}\left|\frac{s_1 - s_2}{s_1}\right|^{\delta}. \end{aligned} \quad (A11)$$

Now put $s_2 = s_1(1 + b)$ and note that we may restrict atten-

tion to small b, say $b\leqslant\frac{1}{8}$. For $b>\frac{1}{8}$ the required bound is a direct consequence of (A5a):

$$|g(s_1)-g(s_2)|\leqslant|g(s_1)|+|g(s_2)|\leqslant\frac{\kappa}{\ln^{a-1}s_1}$$
$$\leqslant\frac{\kappa b^{\delta}}{\ln^{a-1}s_1}=\frac{\kappa}{\ln^{a-1}s_1}\left|\frac{s_1-s_2}{s_1}\right|^{\delta}. \quad \text{(A12)}$$

Let us extend the domain of $f(t)$ to include the interval $\frac{1}{2}s_0\leqslant t\leqslant s_0$, putting $f(t)=0$ on that interval. The property (A4b) holds for the extended function; for $t_1<s_0$ and $t_2\geqslant s_0$

$$|f(t_1)-f(t_2)|=|f(t_2)|\leqslant\frac{\kappa}{\ln^a t_2}\left(1-\frac{s_0}{t_2}\right)^{\theta}$$
$$\leqslant\frac{\kappa}{\ln^a t_2}\left(1-\frac{s_0}{t_2}\right)^{\theta-\delta}\left(1-\frac{t_1}{t_2}\right)^{\delta}$$
$$\leqslant\frac{\kappa}{\ln^a t_1}\left(\frac{t_2-s_0}{t_2}\right)^{\theta-\delta}\left(\frac{t_1-t_2}{t_1}\right)^{\delta}. \quad \text{(A13)}$$

By (A8) we have

$$g_1(s)=\int_{s_0/2}^{2s_2}\frac{f(t)-f(s)}{t-s}\,dt+f(s)\ln\left(\frac{2s_2-s}{s-s_0/2}\right)$$
$$=h_1(s)+h_2(s), \quad \text{(A14)}$$

and

$$|h_2(s_1)-h_2(s_2)|\leqslant|f(s_1)-f(s_2)|\left|\ln\left(\frac{2s_2-s_1}{s_1-s_0/2}\right)\right|+$$
$$|f(s_2)|\left|\ln\left(\frac{2s_2-s_1}{s_1-s_0/2}\right)-\ln\left(\frac{2s_2-s_2}{s_2-s_0/2}\right)\right|. \quad \text{(A15)}$$

The logarithmic factor in the first term of (A15) is clearly bounded by a constant for $b\leqslant\frac{1}{8}$ and $s_1\geqslant s_0$. For the second term we use the mean-value theorem, noting that

$$\sup_{s_1\leqslant s\leqslant s_2}\left|\frac{d}{ds}\ln\left(\frac{2s_2-s}{s-s_0/2}\right)\right|$$
$$\leqslant\sup\left|\frac{1}{2s_2-s}\right|+\sup\left|\frac{1}{s-s_0/2}\right|$$
$$=1/s_2+1/(s_1-s_0/2)\leqslant\kappa/s_1. \quad \text{(A16)}$$

The difference of logarithms in (A15) is then less than $\kappa b\leqslant\kappa b^{\delta}$ and the required bound of the increment of h_2 is obtained from (A15). To estimate the increment of h_2 we break the integral into three parts,

$$h_1=\int_{s_0/2}^{s_1(1-2b)}+\int_{s_1(1-2b)}^{s_1(1+2b)}+\int_{s_1(1+2b)}^{2s_2}=j_1+j_2+j_3. \quad \text{(A17)}$$

The separate terms in the increment of j_2 are so small that we need not consider their difference:

$$|j_2(s_1)-j_2(s_2)|$$
$$\leqslant|j_2(s_1)|+|j_2(s_2)|$$
$$\leqslant\frac{\kappa}{s_1^{\delta}\ln^a s_1}\int_{s_1(1-2b)}^{s_1(1+2b)}dt\left(\frac{1}{|t-s_1|^{\delta}}+\frac{1}{|t-s_2|^{\delta}}\right)$$
$$=\frac{\kappa}{\ln^a s_1}\int_{1-2b}^{1+2b}du\left(\frac{1}{|u-1|^{1-\delta}}+\frac{1}{|u-1-b|^{1-\delta}}\right)$$
$$\leqslant\frac{\kappa b^{\delta}}{\ln^a s_1}. \quad \text{(A18)}$$

Next we estimate

$$j_1(s_2)+j_3(s_2)-j_1(s_1)-j_3(s_1)$$
$$=[f(s_1)-f(s_2)]\left(\int_{s_0/2}^{s_1(1-2b)}+\int_{s_1(1+2b)}^{2s_2}\right)\frac{dt}{t-s}$$
$$+\left(\int_{s_0/2}^{s_1(1-2b)}+\int_{s_1(1+2b)}^{2s_2}\right)dt\,[f(t)-f(s_2)]$$
$$\times\left(\frac{1}{t-s_2}-\frac{1}{t-s_1}\right)=k_1+k_2. \quad \text{(A19)}$$

The part k_1 is easily disposed of:

$$|k_1|=|f(s_1)-f(s_2)|\left|\ln\left(\frac{2s_2-s_1}{s_1-s_0/2}\right)\right|\leqslant\frac{\kappa}{\ln^a s_1}b^{\delta}. \quad \text{(A20)}$$

For the first integral in k_2 we need a further decomposition to handle the combination of two poles and a logarithm:

$$|k_{21}|\leqslant\kappa|s_1-s_2|\int_{s_0/2}^{s_1(1-2b)}\frac{dt}{t^{\delta}\ln^a t}\frac{1}{|t-s_2|^{1-\delta}}\frac{1}{|t-s_1|}$$
$$=\kappa b\int_{s_0/2s_1}^{1-2b}\frac{du}{u^{\delta}\ln^a s_1 u}\frac{1}{|u-1-b|^{1-\delta}}\frac{1}{|u-1|}$$
$$\leqslant\kappa b\left(\int_{s_0/2s_1}^{1/2}\frac{du}{u^{\delta}\ln^a s_1 u}\right.$$
$$\left.+\frac{1}{\ln^a s_1}\int_{1/2}^{1-2b}\frac{du}{|u-1-b|^{1-\delta}|u-1|}\right)$$
$$\leqslant\kappa b\left(\frac{1}{s_1^{1-\delta}}\int_{s_0/2}^{s_1/2}\frac{dt}{t^{\delta}\ln^a t}+\frac{1}{\ln^a s_1}\int_{1/2}^{1-2b}\frac{du}{|u-1|^{2-\delta}}\right)$$
$$\leqslant\frac{\kappa b}{\ln^a s_1}(1+b^{\delta-1})\leqslant\frac{\kappa b^{\delta}}{\ln^a s_1}. \quad \text{(A21)}$$

To complete the proof of Lemma 2 we treat the second integral in k_2:

$$|k_{22}|\leqslant\frac{\kappa|s_1-s_2|}{s_2^{\delta}\ln^a s_2}\int_{s_1(1+2b)}^{2s_2}\frac{dt}{|t-s_2|^{1-\delta}}\frac{1}{|t-s_1|}$$
$$=\frac{\kappa|s_1-s_2|}{s_2^{\delta}s_1^{1-\delta}\ln^a s_2}\int_{1+2b}^{2(1+b)}\frac{du}{|u-1-b|^{1-\delta}}\frac{1}{|u-1|}$$
$$\leqslant\frac{\kappa b}{\ln^a s_2}\int_{1+2b}^{2(1+b)}\frac{du}{|u-1-b|^{2-\delta}}<\frac{\kappa b^{\delta}}{\ln^a s_1}. \quad \text{(A22)}$$

Notice that if two functions $f_1(t)$, $f_2(t)$ satisfy (A4), then the product $f_1(t)f_2(t)$ satisfies (A4) with the exponent a replaced by $2a$. Consequently, when we estimate the integral in (2.3) using (2.2f) and the definition (1.2) of $\rho(t)$, we find that it obeys conditions like (A5) with $a=2\alpha$, $\delta=\mu$. Since $2\alpha-1>\alpha$, we thereby establish conditions (2.4d) on $B(s)$.

APPENDIX B: FREDHOLM THEORY OF THE INTEGRAL EQUATION

We show that the integral equation (2.27) may be treated by Fredholm theory[20] under conditions (2.4d) on $B(s)$. We map the interval $[s_1,\infty)$ onto $(0,1]$. The choice of the mapping is not crucial; we take $t=s_1/s$ for convenience. [In a numerical calculation of the Fredholm solution it is usually best to choose the mapping $t(s)$ so as to make the integrand finite and nonzero at the point corresponding to $s=\infty$.] We multiply the equation by $(\ln s)^{\alpha}$, $\alpha>1$, and seek a solution $\phi(t)=(\ln s)^{\alpha}N(s)$ in a Banach space U consisting of real matrix functions $\phi(t)$ continuous on the closed interval $[0,1]$ with norm

$$\|\phi\|=\sup_{t,i,j}|\phi_{ij}(t)|. \quad \text{(B1)}$$

Let us define the operator K by the formula

$$K\phi(t) = (\ln s)^{\alpha}\int_{s_1}^{\infty} \frac{B(s)-B(s')}{s-s'}\rho(s')\frac{\phi(t')}{(\ln s')^{\alpha}}\,ds'. \tag{B2}$$

As we shall see presently, K maps U into itself if $K\phi(0)$ is defined to be zero. According to the Ascoli–Arzelà criterion,[20] K is compact (completely continuous) if the sequence $\{K\phi_n(t)\}$ is bounded and equicontinuous, where $\{\phi_n(t)\}$ is any bounded sequence of functions in U.

Let $\{\phi_n(t)\}$ be a bounded sequence in U, $\|\phi_n\| < \kappa$, and check boundedness of $\|K\phi_n\|$ as follows:

$$\|K\phi_n\| \leqslant \sup_s \sum_{i,j} (\ln s)^{\alpha} \times \int_{s_1}^{\infty} \left| \frac{B_{ij}(s)-B_{ij}(s')}{s-s'}\rho_j(s') \right| \frac{\|\phi_n\|}{(\ln s')^{\alpha}}\,ds'. \tag{B3}$$

An analysis like that in (A6), (A7), and (A9) (but not requiring subtraction of a logarithim) shows that the integral in (B3) is $O(\ln^{-2\alpha+1} s)$, thus $\|K\phi_n\| \leqslant \kappa$, since $\alpha > 1$. Incidentally we have shown that $K\phi(t) \to 0$, $t \to 0$. With the definition $K\phi(0) = 0$ the function $K\phi(t)$ is continuous on the closed interval $[0,1]$; K maps U into itself.

The requirement of equicontinuity of the functions $K\phi_n(t)$ is that for any $\epsilon > 0$,

$$\max_{i,j} |[K\phi_n(t_1) - K\phi_n(t_2)]_{ij}| < \epsilon \tag{B4}$$

when $|t_1 - t_2| < \delta(\epsilon)$, where δ is independent of n. With $f_n(t_1,t_2)$ defined as the left side of (B4), $\{\phi_n\}$ any bounded sequence, and $s_1 \leqslant s_2$, we have

$$\begin{aligned} f_n(t_1,t_1) \leqslant{}& \kappa|(\ln s_1)^{\alpha} - (\ln s_2)^{\alpha}| \\ &\times \sum_{i,j}\int_{s_1}^{\infty} \left| \frac{B_{ij}(s_2)-B_{ij}(s')}{s_2-s'}\rho_j(s') \right| \frac{ds'}{(\ln s')^{\alpha}} + \kappa(\ln s_1)^{\alpha} \\ &\times \sum_{ij}\int_{s_1}^{\infty} \left| \frac{B_{ij}(s_1)-B_{ij}(s')}{s_1-s'} - \frac{B_{ij}(s_2)-B_{ij}(s')}{s_2-s'} \right| \\ &\times \rho_j(s')\frac{ds'}{(\ln s')^{\alpha}} = g + h. \end{aligned} \tag{B5}$$

The right side of (B5) is independent of n and we have only to show that it vanishes with $|t_1 - t_2|$. The analysis of (A10)–(A22), simpler now because we needn't bother with subtraction of logarithmic terms, shows that the second term h in (B5) has the bound

$$h(t_1,t_2) \leqslant \frac{\kappa}{(\ln s_1)^{\alpha-1}} \left| \frac{s_1-s_2}{s_1} \right|^{\mu} = \frac{\kappa}{(\ln s_1)^{\alpha-1}} \left| \frac{t_1-t_2}{t_2} \right|^{\mu}. \tag{B6}$$

Also, we may bound the two terms in h separately to get

$$h(t_1,t_2) \leqslant \frac{\kappa}{(\ln s_1)^{\alpha-1}} + \kappa\frac{(\ln s_1)^{\alpha}}{(\ln s_2)^{2\alpha-1}} \leqslant \frac{M}{(\ln s_1)^{\alpha-1}}. \tag{B7}$$

For any $\epsilon > 0$ let us divide the interval of t_2 into two parts, $t_2 \leqslant \eta(\epsilon)$ and $t_2 > \eta(\epsilon)$, where $\eta(\epsilon)$ is chosen to make

$$M/[\ln(s_0/2\eta)]^{\alpha-1} < \tfrac{1}{2}\epsilon, \tag{B8}$$

with M as in (B7). Then if $t_2 \leqslant \eta(\epsilon)$ and $|t_1 - t_2| < \eta(\epsilon)$ we have by (B7) that $h(t_1,t_2) < \frac{1}{2}\epsilon$. On the other hand, if $t_2 > \eta(\epsilon)$ we have by (B6) that

$$h(t_1,t_2) \leqslant \frac{\kappa}{(\ln s_0)^{\alpha-1}} \left| \frac{t_1-t_2}{\eta(\epsilon)} \right|^{\mu} < \tfrac{1}{2}\epsilon \tag{B9}$$

for $|t_1 - t_2|$ less than some $\zeta(\epsilon)$. Hence for $|t_1 - t_2| < \min[\eta(\epsilon),\zeta(\epsilon)] = \delta_1(\epsilon)$ we have $h(t_1,t_2) < \frac{1}{2}\epsilon$. To majorize the first term g in (B5) we apply the mean-value theorem to the difference of logarithms and bound the integral as usual to obtain

$$g(t_1,t_2) \leqslant \frac{\kappa}{(\ln s_1)^{\alpha}} \left| \frac{s_1-s_2}{s_1} \right|. \tag{B10}$$

Alternatively, we may bound the two logarithmic terms separately and find

$$g(t_1,t_2) \leqslant \kappa/(\ln s_1)^{\alpha-1}. \tag{B11}$$

The argument used above then shows that $g(t_1,t_2) < \frac{1}{2}\epsilon$ for $|t_1 - t_2|$ less than some $\delta_2(\epsilon)$. For $|t_1 - t_2| < \min[\delta_1(\epsilon),\delta_2(\epsilon)] = \delta(\epsilon)$ we have $f_n(t_1,t_2) < \epsilon$, and the proof of equicontinuity and compactness of K is complete.

Our hypothesis $B(s) = O([\ln s]^{-\alpha})$, $\alpha > 1$, is close in some sense to being the weakest asymptotic condition on $B(s)$ that leads to a Fredholm equation in a classical Banach space of continuous functions. With $B(s) \sim (\ln s)^{-1}$ the operator K is noncompact in a space analogous to U, but may be regularized by extraction of a noncompact part in such a way that the problem is reduced to a regular Fredholm problem.[21] Under still weaker conditions on $B(s)$ a regularization is possible, but only at the expense of new arbitrary constants entering the equations.[22]

Since Eq. (2.27) entails a compact operator, it may be solved numerically by various well-developed methods; see, for instance, the review of Atkinson[23] and the book of Anselone;[24] the latter deals with the rigorous justification of discretization.

The operator of Eq. (5.6), multiplied by s^{α}, is compact on a Banach space V under conditions (5.2) and (5.3) on $\rho(s)$ and $B_{22}(s)$. Here V consists of real continuous functions $\phi(t) = s^{\alpha}N_{22}(s)$ with

$$\|\phi\| = \sup|\phi(t)|. \tag{B12}$$

The proof of compactness is the same as that above, but with the estimates of Lemma 1 of Appendix A replacing those of Lemma 2.

[1] F.J. Yndurain, Nucl. Phys. B **89**, 317 (1975).
[2] F.J. Yndurain, Nucl. Phys. B **88**, 318 (1975).
[3] M. Aguilar-Benitez *et al.*, "Study of a $K^0_S K^0_S$ Threshold Enhancement Observed in the Final States $\bar{p}p \to K^0_S K^0_S +$ pions at 700–760 MeV/c," CERN preprint CERN/EP/PHYS 78-5, 1978).
[4] A. González-Arroyo, "Extension of the K-Matrix Formalism to Amplitudes with Overlapping Left-hand Cuts," Preprint FTUAM/78-2, Universidad Autónoma de Madrid, 1978; "Analisis Multicanal de las Resonancias Escalares e Isoescalares," Ph.D. Thesis, Universidad Autónoma de Madrid, 1978.
[5] H. Nielsen and G.C. Oades, Nucl. Phys. B **55**, 301 (1973); N. Hedegaard-Jensen, *ibid* B **68**, 397 (1974); T. Shimada, *ibid.* B **100**, 225 (1975); J.J. Brehm, *ibid.* B **55**, 167 (1973); S.M. Flatté, Phys. Lett. B **63**, 224 (1976); N. Kimura, T. Miyakoshi, and A. Kanazawa, "Bias-Free Pole Search on the S^* Resonance," preprint EPHOU-78 IV 01, Hokkaido University, 1978.

[6]I.S. Shapiro, Phys. Rep. C **35**, 129 (1978); B.O. Kerbikov and I.S. Shapiro, "Quasinuclear Baryonium and Annihilation." preprint ITEP-159, Institute of Theoretical and Experimental Physics, Moscow, 1978; L.N. Bogdanova, V. E. Markushin, and I. S. Shapiro, Yad. Fiz. **30**, 480 (1979); C.B. Dover and J.M. Richard, Brookhaven preprint BNL-24981, 1978; C.B. Dover and M. Goldhaber, Phys. Rev. D **15**, 1997 (1977); Yu. A. Simonov and J.A. Tjon, "Effect of Annihilation on Resonances in the $N\bar{N}$ System", Utrecht preprint, 1978.
[7]G. Frye and R.L. Warnock, Phys. Rev. **130**, 478 (1963); R.L. Warnock, in *Lectures in Theoretical Physics*, edited by H.H. Aly (Wiley, New York, 1968).
[8]R. Jost and R. Newton, Nuovo Cimento **1**, 590 (1955).
[9]J.D. Bjorken, Phys. Rev. Lett. **4**, 473 (1960).
[10]J.D. Bjorken and M. Nauenberg, Phys. Rev. **121**, 1250 (1961).
[11]R.L. Warnock, Nuovo Cimento **50**, 894 (1967); A **52**, 637(E) (1967).
[12]R.L. Warnock, Physica A **96**, 321 (1979); *Themes in Contemporary Physics*, edited by S. Deser *et al.* (North–Holland, Amsterdam, 1979).
[13]P.W. Johnson and R.L. Warnock, Phys. Rev. D **15**, 2354 (1977); D **15**, 2366 (1977); P.W. Johnson, R.L. Warnock, and M. Kaekebeke, *ibid.* D **15**, 482 (1977).
[14]R.L. Warnock, Phys. Rev. **146**, 1109 (1966).
[15]G. Nenciu, G. Rasche, M. Stihi, and W.S. Woolcock, Helv. Phys. Acta **51**, 608 (1978).
[16]The existence of such a function $\mathscr{D}(s)$ was proved in Ref. 11 by reducing the question to the Hilbert problem in n unknown functions.[17] The conditions on $T(s)$ assumed in Ref. 11 follow, as far as essentials are concerned, from (2.2). Property (2.5e) was obtained only under stronger conditions on $T(s)$ however. As we shall show in a forthcoming paper, it is possible to simplify greatly the argument of Ref. 11, and at the same time prove the existence of a $\mathscr{D}(s)$ with properties (2.5) for any $T(s)$ satisfying conditions (2.2).
[17]N.I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, 1953).
[18]D. Atkinson and M. Halpern, Phys. Rev. **149**, 1133 (1966).
[19]The definition (5.21) of $K(s)$ agrees with definition of Ref. 4 only for $s > s_2$. For $s < s_2$, the $K(s)$ of Ref. 4 is the analytic continuation of (5.21). Notice that (5.21) is the boundary value of one analytic function for $s < s_2$ and that of another analytic function for $s > s_2$.
[20]L.A. Liusternik and V.J. Sobolev, *Elements of Functional Analysis* (Unger, New York, 1961).
[21]P.W. Johnson and R.L. Warnock, J. Math. Phys. **16**, 1880 (1975).
[22]D. Atkinson and R.L. Warnock, Phys. Rev. D**16**, 1948 (1977).
[23]K.E. Atkinson, *A Survey of Numerical Methods for the Solution of Fredholm Integral Equations of the Second Kind* (Society for Industrial and Applied Mathematics, Philadelphia, 1976).
[24]P.M. Anselone, *Collectively Compact Operator Approximation Theory and Applications to Integral Equations*, (Prentice-Hall, Englewood Cliffs, 1971).

Representation matching and the gauge invariant energy-momentum tensor

Zhu Dongpei [a)]
Physics Department, University of Wisconsin–Madison, Madison Wisconsin 53706

(Received 15 August 1980; accepted for publication 19 September 1980)

We atttack the problem of finding appropriate representations of symmetry groups so that a conserved quantity derived by the Noether theorem from one symmetry of a theory preserves the other symmetry. This representation matching problem is well solved in two-dimensional gauge theory where we find a new representation of the translation group which leads straightforwardly to the symmetric, gauge-invariant energy-momentum tensor.

PACS numbers: 11.30. – j, 02.20.Qs

I. INTRODUCTION

The success of gauge theory has revived interest in some old problems. The connection between constants of the motion and symmetry is one of them. It is well known that a conserved physical quantity is deeply related with symmetry. The Noether theorem[1,2] provides the connection between a physical observable and its symmetry source. But when a theory simultaneously possesses several kinds of symmetry, a conserved quantity derived from one symmetry by the Noether theorem may not preserve the other symmetry; this leaves some ambiguities in interpreting this quantity as a physical observable. A gauge theory is one such case. The energy-momentum tensor and the angular momentum tensor derived from space-time symmetry are not gauge-invariant. The problem exists in the gauge field sector which is described by a Lagrangian of the form

$$\mathscr{L}_{eff} = -\tfrac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} = -\tfrac{1}{2}\mathrm{tr}F_{\mu\nu}F^{\mu\nu}. \tag{1}$$

Here

$$F_{\mu\nu} = T^a F^a_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] \tag{2}$$

and

$$A_\mu = T^a A^a_\mu. \tag{3}$$

The generators of the gauge group are normalized by

$$\mathrm{tr}T^a T^b = \tfrac{1}{2}\delta^{ab}. \tag{4}$$

Under the translation

$$x_\mu \to x'_\mu = x_\mu + b_\mu \tag{5}$$

the field A_μ is usually assumed to transform as a derivative or a "scalar"

$$A'_\mu(x') = A_\mu(x). \tag{6}$$

With the continuity of A_μ, one obtains an infinitesimal form of "local variation"[3]

$$\begin{aligned}\delta^* A_\mu &= A'_\mu(x) - A_\mu(x) \\ &= -b^\nu \partial_\nu A_\mu(x).\end{aligned} \tag{7}$$

Then the conserved current is

$$J_\mu \sim -(-2\mathrm{tr}F^\nu_\mu \delta^* A_\nu + \mathscr{L} b_\mu) \tag{8}$$

$$= -b^\rho(2\mathrm{tr}F^\nu_\mu \partial_\rho A_\nu + \mathscr{L} g_{\rho\mu}). \tag{8'}$$

Thereby we derive out an energy-momentum tensor

$$T_{\mu\nu} = -2\mathrm{tr}F^\sigma_\mu \partial_\nu A_\sigma - \mathscr{L} g_{\mu\nu}. \tag{9}$$

Obviously it is not invariant under the gauge transformation

$$A'_\mu = U(A_\mu + U^{-1}\partial_\mu U)U^{-1}, F'_{\mu\nu} = UF_{\mu\nu}U^{-1}. \tag{10}$$

(Furthermore, this energy-momentum tensor does not couple to gravity[4].) The same thing happens when one deals with the Lorentz rotation and angular momentum tensor.

A conventional way to remedy these undesirable features is to add a divergence free term (superpotential[4]) to $T_{\mu\nu}$, making it symmetric and gauge invariant. For (9), one defines another quantity

$$\theta_{\mu\nu} = T_{\mu\nu} + 2\mathrm{tr}\partial^\sigma(F_{\mu\sigma}A_\nu) = 2\mathrm{tr}F_{\mu\rho}F_{\rho\nu} - \mathscr{L} g_{\mu\nu}. \tag{11}$$

Here we have used the equations of motion

$$[D_\mu, F^{\mu\nu}] = 0, \tag{12}$$

$$D_\mu \equiv \partial_\mu + A_\mu. \tag{13}$$

Because this method is a bit artificial and obscures the connection between the conserved quantity and symmetry, it is interesting to find a way showing (11) as a natural inference of the Noether theorem. From the above process we can see that the form of a conserved tensor is related to what we choose for the representation of a symmetry group. The non-gauge-invariant energy-momentum tensor is a consequence of choosing the field A_μ to generate a trivial representation of the translation group. Obviously, it is not necessary to require an unphysical vector field transform in this way.

Thus the problem is to find a suitable representation of the translation group which makes the energy-momentum tensor gauge-invariant. This is a representation matching problem. It can be more precisely stated as follows: When a theory possesses two kinds of symmetry, can we find appropriate representations of the symmetry groups so that the conserved quantity derived from one symmetry by the Noether theorem preserves the other symmetry?

Recently there has been some work along this line.[4–9] The authors of Ref. 5 found that the discussion of the closure of the supersymmetry leads to the concept of a gauge-invariant translation from which one can obtain the gauge-invariant energy-momentum tensor. Using a gauge transformation to eliminate the asymmetric, non-gauge-invariant part while making coordinate transformation, the author of Ref. 6 modifies the translation law (7) into

[a)] On leave from the Department of Modern Physics, China University of Science and Technology, Hefei, The People's Republic of China.

 0022-2488/81/020400-03$1.00

$$\delta^* A = b^{\nu} F_{\nu\mu}. \tag{14}$$

Then the desirable tensor follows. Unfortunately, as the author has pointed out, the transformation law (14) cannot preserve the integrability condition. This transformation does not give a representation of the group.

This paper attempts to deal with this particular representation matching problem. We find that this problem can be well solved in two-dimensional gauge theory. In Sec. II we provide a new representation of the translation group in the Hilbert space of a $U(1)$ gauge field. We obtain the gauge-invariant tensor straightforwardly with the Noether theorem after using the Lorentz condition. To discuss the algebra, we also derive a new representation of rotations. In the light cone system the new transformation law gives an eigenform of the translation. Then we generalize this method to the nonabelian two-dimensional theory in Sec. III. Finally, Sec. IV is a discussion of the problems encountered in the four-dimensional case.

II. REPRESENTATION OF 2D TRANSLATION

Two-dimensional (2D) gauge theory is a good testing field where many ideas can be examined.[10] Now we provide another example to show the special advantage of 2D theory. In this section we only discuss U(1) gauge theory.

It is well known that the translation group only has one trivial finite order irreducible representation[3] induced from Eq. (6). But one can nontrivially represent this group in the Hilbert space. In the U(1) case, (5) and (7) can be written as

$$x'_{\mu} = x_{\mu} - ib^{\alpha} P_{\alpha,\mu}{}^{\nu} x_{\nu}, \tag{15}$$

$$\delta^* A_{\mu} = ib^{\alpha} P_{\alpha,\mu}{}^{\nu} A_{\nu}. \tag{16}$$

The infinitesimal operators P_{α} are

$$P_{\alpha,\mu}{}^{\nu} = i\partial_{\alpha} g_{\mu}{}^{\nu}; \tag{17}$$

they satisfy the abelian algebra

$$[P_{\alpha}, P_{\beta}] = 0. \tag{18}$$

Nevertheless, it is not necessary that every component of a vector field transform uniformly. Actually (5) also can be written in the form

$$x'_{\mu} = x_{\mu} - ib^{\alpha} Q_{\alpha,\mu}{}^{\nu} x_{\nu}, \tag{19}$$

$$Q_{\alpha,\mu}{}^{\nu} = -i(g_{\alpha}{}^{\nu}\partial_{\mu} - g_{\alpha\mu}\partial^{\nu}), \tag{20}$$

or using a matrix notation

$$x' = B(b)x, \tag{21}$$

$$x = \begin{pmatrix} x_0 \\ x_1 \end{pmatrix}, \tag{22}$$

$$B(b) = e^{-ib^{\alpha} Q_{\alpha}}, \tag{23}$$

$$Q_{\alpha} = -i\epsilon\epsilon_{\alpha}{}^{\beta}\partial_{\beta}, \tag{24}$$

$$\epsilon = (\epsilon_{\mu}{}^{\nu}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{25}$$

It is very easy to verify that the Q_{α}'s satisfy the same algebra as the P_{α}'s,

$$[Q_{\alpha}, Q_{\beta}] = 0. \tag{26}$$

Now we suppose that under translation the vector field A_{μ} transforms in the same way as x_{μ}

$$A'_{\mu}(x') = B(b)_{\mu}{}^{\nu} A_{\nu}. \tag{27}$$

We obtain a complete variation in the infinitesimal form

$$\delta A_{\mu} = ib^{\alpha}[Q_{\alpha,\mu}{}^{\nu}, A_{\nu}] = -b^{\alpha}(\partial_{\mu} A_{\alpha} - g_{\mu\alpha}\partial^{\nu} A_{\nu}) \tag{28}$$

and a local variation

$$\delta^* A_{\mu} = ib^{\alpha}[Q_{\alpha,\mu}{}^{\nu} + P_{\alpha,\mu}{}^{\nu}, A_{\nu}] \tag{29}$$

$$= b^{\alpha}(F_{\mu\alpha} - g_{\mu\alpha}\partial^{\nu} A_{\nu}). \tag{30}$$

Substituting this back into (8), we obtain the right form of the energy-momentum tensor

$$\theta_{\mu\nu} = F_{\mu}{}^{\rho} F_{\rho\nu} - \mathcal{L} g_{\mu\nu} \tag{31}$$

providing the Lorentz condition

$$\chi = \partial^{\alpha} A_{\alpha} = 0. \tag{32}$$

Under transformation (29), the field strength and the Lorentz condition are not changed

$$\delta F_{\mu\nu} = F'_{\mu\nu}(x') - F_{\mu\nu}(x) = b^{\alpha}(g_{\mu\alpha} - g_{\nu\alpha})\partial^{\rho} A_{\rho} = 0, \tag{33}$$

$$\delta\partial^{\rho} A_{\rho} = b^{\alpha}\partial^{\rho} F_{\rho\alpha} = 0. \tag{34}$$

So the theory is translation-invariant.

Equation (29) generates a representation of the translation group. The infinitesimal operators are Q_{α}'s. To see their position in the whole Poincaŕe algebra, we need to study them together with rotations. Using the same method one can obtain three forms of 2D rotation

$$x'_{\mu} = x_{\mu} + \omega_{\mu}{}^{\nu} x_{\nu} \tag{35}$$

$$= x_{\mu} + (-i)(\omega^{\alpha\beta}/2)\Sigma_{\alpha\beta,\mu}{}^{\nu} x_{\nu} \tag{36}$$

$$= x_{\mu} + (-i/2)\omega^{\alpha\beta} L_{\alpha\beta,\mu}{}^{\nu} x_{\nu} \tag{37}$$

$$= x_{\mu} + (-i/2)\omega^{\alpha\beta} N_{\alpha\beta,\mu}{}^{\nu} x_{\nu}. \tag{38}$$

Here

$$\Sigma_{\alpha\beta,\mu}{}^{\nu} = i(g_{\alpha\mu} g_{\beta}{}^{\nu} - g_{\alpha}{}^{\nu} g_{\beta\mu}), \tag{39}$$

$$L_{\alpha\beta,\mu}{}^{\nu} = -i(x_{\alpha}\partial_{\beta} - x_{\beta}\partial_{\alpha}) g_{\mu}{}^{\nu} = x_{\beta} P_{\alpha,\mu}{}^{\nu} - x_{\alpha} P_{\beta,\mu}{}^{\nu}, \tag{40}$$

$$N_{\alpha\beta,\mu}{}^{\nu} = x_{\beta} Q_{\alpha,\mu}{}^{\nu} - x_{\alpha} Q_{\beta,\mu}{}^{\nu}. \tag{41}$$

In the matrix notation these operators have the forms

$$\Sigma_{\alpha\beta} = -i\epsilon\epsilon_{\alpha\beta}, \tag{42}$$

$$L_{\alpha\beta} = x_{\beta} P_{\alpha} - x_{\alpha} P_{\beta}, \tag{43}$$

$$N_{\alpha\beta} = i\epsilon(x_{\alpha}\epsilon_{\beta}{}^{\gamma} - x_{\beta}\epsilon_{\alpha}{}^{\gamma})\partial_{\gamma}. \tag{44}$$

Then we have the following commutation relations:

	$B = P_0$	P_1	L_{01}	Q_0	Q_1	N_{01}	Σ_{01}
$A = P_0$	$[A,B] = 0$	0	$-iP_1$	0	0	$-iQ_1$	0
P_1		0	$-iP_0$	0	0	$-iQ_0$	0
L_{01}			0	iQ_1	iQ_0	0	0
Q_0				0	0	$-iP_1$	0
Q_1					0	$-iP_0$	0
N_{01}						0	0
Σ_{01}							0

This algebra is larger than the Poincaŕe algebra.[11] But we can define some appropriate operators

$$T_{\mu} = \tfrac{1}{2}(P_{\mu} + Q_{\mu}), \tag{46}$$

$$M_{\alpha\beta} = \tfrac{1}{2}(L_{\alpha\beta} + N_{\alpha\beta} + \Sigma_{\alpha\beta})$$
$$= \tfrac{1}{2}(x_\beta T_\alpha - x_\alpha T_\beta) + \tfrac{1}{2}\Sigma_{\alpha\beta} \quad (47)$$

which satisfy the usual Poincaré algebra

$$[T_\mu, T_\nu] = 0$$
$$[T_\mu, M_{\alpha\beta}] = i g_{\mu\beta} T_\alpha - i g_{\mu\alpha} T_\beta \quad (48)$$

Using T_μ as translation generator we achieve the right energy-momentum tensor; likewise, the generator $L_{\alpha\beta} + N_{\alpha\beta}$ provides the correct gauge-invariant angular momentum tensor.

This infinite-dimensional representation (46) of the translation generator is "reducible". Actually, when we turn to the light cone system[12]

$$x_\pm = (1/\sqrt{2})(x_0 \pm x_1), \quad (49)$$
$$\partial^\pm = (1/\sqrt{2})(\partial^0 \pm \partial^1), \quad (50)$$
$$A_\pm = (1/\sqrt{2})(A_0 \pm A_1), \quad (51)$$

the transformation law (29) becomes

$$\delta^* A_+ = -2(b^- \partial_-) A_+$$
$$\delta^* A_- = -2(b^+ \partial_+) A_- \quad (52)$$

These equations are more likely showing the characteristic of the translation and we notice that in the light-cone system the different degrees of freedom under translation are separated.

III. NONABELIAN CASE

Turing to 2D nonabelian gauge theory, we first rewrite the curvature matrix in the form

$$F_{\mu\nu} = [\tilde{D}_\mu, A_\nu] - [\tilde{D}_\nu, A_\mu], \quad (53)$$
$$\tilde{D}_\mu \equiv \partial_\mu + \tfrac{1}{2} A_\mu. \quad (54)$$

Comparing to the U(1) curvature

$$F_{\mu\nu} = [\partial_\mu, A_\nu] - [\partial_\nu, A_\mu], \quad (55)$$

we find that we only need to make a replacement of $\partial_\mu \to \tilde{D}_\mu$ when dealing with the nonabelian case. So we change the transformation law (29) into

$$\delta^* A_\mu = i b^\rho [\tilde{T}_{\rho,\mu}{}^\nu, A_\nu]$$
$$= -b^\rho F_{\rho\mu} - b_\mu \partial^\nu A_\nu. \quad (56)$$

Here

$$\tilde{T}_{\rho,\mu}{}^\nu = i(g_\mu{}^\nu \tilde{D}_\rho - g_\rho{}^\nu \tilde{D}_\mu + g_{\rho\mu} \tilde{D}_\rho{}^\nu). \quad (57)$$

In the matrix notation, the $\tilde{T}_\rho$ have the forms

$$\tilde{T}_0 = i\begin{pmatrix} \tilde{D}_0 & -\tilde{D}_1 \\ -\tilde{D}_1 & \tilde{D}_0 \end{pmatrix},$$
$$\tilde{T}_1 = i\begin{pmatrix} \tilde{D}_1 & -\tilde{D}_0 \\ -\tilde{D}_0 & \tilde{D}_1 \end{pmatrix}. \quad (58)$$

They also satisfy the Lie–Cartan integrability condition

$$[\tilde{T}_0, \tilde{T}_1] = 0 \quad (59)$$

and so we can get a finite form of the representation after exponentiation.

Using this representation we soon obtain the correct energy-momentum tensor

$$\theta_{\mu\nu} = -\mathrm{tr}(F_\mu{}^\rho F_{\nu\rho}) - \mathcal{L} g_{\mu\nu} \quad (60)$$

with the condition

$$[\tilde{D}^\mu, A_\mu] = 0, \quad (61)$$

which implies the conventional Landau gauge

$$\partial^\mu A^a_\mu = 0. \quad (62)$$

IV. DISCUSSION

We have shown that in the 2D case the representation matching problem can be well solved. This is because in this case the antisymmetric tensor $\epsilon_\mu{}^\nu$ is just an exchange operator. In the 4D case the above trick does not work. We may write the translation in the form

$$x'_\mu = x_\mu + b_\mu$$
$$= x_\mu + (b^\alpha/(4c + a))(a g_\alpha{}^\nu \partial_\nu + c g_{\alpha\mu} \partial^\nu + d \epsilon_{\alpha\rho\mu}{}^\nu \partial^\rho) x_\nu.$$

We define a set of operators

$$T'_{\alpha,\mu}{}^\nu = i(g_\mu{}^\nu \partial_\alpha + a g_\alpha{}^\nu \partial_\mu + c g_{\alpha\mu} \partial^\nu + d \epsilon_{\alpha\rho\mu}{}^\nu \partial^\rho). \quad (63)$$

The commutators among them are

$$[T'_\alpha, T'_\beta]^\nu_\mu = -[(ac + d^2)(g_{\alpha\mu} g_\beta{}^\nu - g_\alpha{}^\nu g_{\beta\mu}) \partial^\rho \partial_\rho$$
$$+ (a^2 - d^2)(g_\beta{}^\nu \partial_\mu \partial_\alpha - g_\alpha{}^\nu \partial_\mu \partial_\beta)$$
$$+ (c^2 - d^2)(g_{\alpha\mu} \partial_\beta \partial^\nu - g_{\beta\mu} \partial_\alpha \partial^\nu)$$
$$- 2d(c \epsilon_{\rho\alpha\mu\beta} \partial^\nu + a \epsilon_{\rho\beta\alpha}{}^\nu \partial_\mu) \partial^\rho]. \quad (64)$$

Whatever one chooses for a,c, and d, except all zeros, will not make the T_α satisfy the abelian algebra. So it does not provide any representation of the translation group other than the usual one. But there is a possiblity that they may form a representation of some sort of larger group. So for the 4D case the representation-matching problem is still open.

ACKNOWLEDGMENTS

I would like to thank Professor B. Durand and Dr. C. Zachos for discussions and help in the preparation of the manuscript.

[1] E. L. Hill, Rev. Mod. Phys. **23**, 253 (1951).
[2] T. H. Boyer, Am J. Phys., **34**, 475 (1966).
[3] P. Roman, *Theory of Elementary Particles* (North–Holland, Amsterdam 1960).
[4] For a summary see S. Treiman, R. Jackiw, and D. Gross, *Lectures on Current Algebra and Its Applications* (Princeton U. P., Princeton, N.J., 1972), p. 97.
[5] B. deWit and D. Z. Freeman, Phys. Rev. D **12**, 4286 (1975).
[6] R. Jackiw, Phys. Rev. Lett. **41**, 1635 (1978).
[7] R. Jackiw, N. S. Manton, MIT preprint, CTP 825 (1979).
[8] R. Jackiw, Univ. of California preprint NSF ITP 80-15 (1980).
[9] E. Eriksen and J. M. Leinaas, "Gauge invariance and the transformation properties of the electromanetic four-potential", University of Oslo preprint, report 80-03 (1980).
[10] C. Sorensen and G. H. Thomas, Phys. Rev. D **21**, 1625 (1980), where many references concerning 2D theory can be found.
[11] L. S. Brown, Nuovo Cimento **29**, 617 (1963).
[12] G. 't Hooft, Nucl. Phys. B **75**, 461 (1974).

Relativistic quantum dynamical group for hadrons

P. Roman[a] and J. Haavisto[b]
Max Planck Institut, Starnberg, West Germany

(Received 8 May 1979; accepted for publication 1 February 1980)

Motivated by previous work on relativistic quantum dynamics expressed in algebraic terms, we introduce a fully relativistic generalization of the Hooke group. The mathematical properties, relation to other proposed quantum dynamical groups, and the unitary ray representations of this group are studied. Hadrons are viewed as de Sitter type microuniverses, where the quantum dynamics is then determined by the relativistic Hooke group. Wave equations are studied, a mass formula is derived, the emergence of a Regge type formula is deduced, and correspondence with other extended hadron models is noted.

PACS numbers: 11.30. − j, 11.70. + q, 03.65.Fd

1. INTRODUCTION

During the past few years, we attempted, in a series of papers, to establish and study relativistic quantum dynamical groups suitable for the description of subnuclear phenomena. The prototype of such an enterprise is, naturally, the central extension of the nonrelativistic Galilei group,[1] $\mathscr{G}_4$, which contains a complete account not only of quantum kinematics but also of the inertial dynamics of a free particle and which also supplies the Heisenberg rules of quantization. If was shown[2] that $\mathscr{G}_4$ can be rigorously motivated and deduced from a locality postulate (gauge principle). It is easy to show[3] that an analogous line of arguments, employing the (flat) Minkowski space (rather than the Euclidean 3-space) as the event-space, leads to a relativistic generalization $\mathscr{G}_5$ of the Galilei group, which has been established earlier on intuitive grounds[4] and studied in several papers.[5] The group $\mathscr{G}_5$ contains the Poincaré group as a subgroup. Its generators correspond to the (Lorentz) rotations and boosts, energy-momentum, event localization, and a relativistic development operator (Hamiltonian). One of the Lie brackets provides a relativistic generalization of the Heisenberg commutation relations and two other Lie brackets correctly render the inertial free motion of a relativistic quantal particle. The irreducible unitary representations lead to spin-towers. The Newton–Wigner position operator emerges in a natural way.[3]

Since $\mathscr{G}_5$ describes only free motions, we must search for a generalization which would account naturally for the emergence of force-effects. The hint in this direction can be found in the interesting work of Bacry and Lévy-Leblond,[6] who introduced, by the method of group contraction from the de Sitter group, a nonrelativistic but "cosmologic" quantum dynamical group $\mathscr{H}_4$ which nowadays is usually called the Hooke group. This algebraic system is a simple generalization of the nonrelativistic Galilean structure with the essential difference that the "inertial motion" now corresponds to the behavior of a quantal particle under the influence of a harmonic oscillator force.[7] Some time ago we showed[8] that this Hooke group arises in a natural way (similarly as did the Galilei group) if one applies the locality principle, this time not to a flat Euclidean event space but rather to a *uniformly curved* three-dimensional event space.

Before engaging in a project aimed at the relativistic generalization of the Hooke group, we found it necessary to study whether this structure is capable of giving a low-energy (low excitation) approximative description of hadrons. In order to make the application of $\mathscr{H}_4$ to hadrons sensible, we first must ask: why should the event space for hadron-physics be curved? In a recent paper[9] we demonstrated that the required large curvature of space in a small region may arise from a vacuum contribution to the hadronic energy-momentum tensor within the framework of completely unified spontaneously broken gauge theories of the Yang–Mills–Einstein–Higgs type.[10] In a sense, in such a theory a hadron may be characterized as an oscillating, topologically open de Sitter type "microuniverse," a "bubble" embedded in the external overall flat Minkowski macroscopic world. Since the low-speed, small spatial distance approximation[6] (i.e., the space-speed contraction) of SO(3,2) is $\mathscr{H}_4$, it follows that the low-lying, nonrelativistic collective excitations of a hadron will be indeed described by the Hooke quantum dynamics. We pursued this line of argument[9] and established a relation between our (nonrelativistic) microuniverse model and the nonrelativistic SU(3) quark model with harmonic forces.[11]

In this paper we address ourselves to the question: what might be the exact, fully relativistic quantum dynamics inside the de Sitter microuniverse that corresponds to hadrons? Mathematically, this requires the establishing, via the locality (gauge) principle, of an algebraic structure that is based on an event space that has the geometry of the (homogeneous and isotropic) de Sitter space. Once we, in this manner, are motivated to deduce the relativistic generalization, $\mathscr{H}_5$, of the Hooke group, we shall study its properties, its representations, the wave equations that follow, and their basic applications to hadron physics.

2. DERIVATION OF THE RELATIVISTIC HOOKE GROUP FROM LOCALITY

The procedure followed in this section is analogous to our method used for deriving the nonrelativistic Hooke

[a]Permanent address from September 1978: Office of the Dean of Graduate Studies and Research, State University of New York, Plattsburgh, New York 12901.
[b]Mailing address: c/o Department of Physics, Boston University, 111 Cummington Street, Boston, Massachusetts 02215.

 0022-2488/81/020403-09$1.00

group[8] and the reader is referred to that paper for some details and for motivations of certain assumptions and steps.

We first wish to construct relativistic quantum kinematics in a curved 4-space of maximal symmetry. Accordingly, we adopt

Assumption 1: The space of events is the (uniformly curved) topologically time-open[12] four dimensional de Sitter space $\mathscr{S}$.

This space can be embedded in a five-dimensional flat space $E_{3,2}$ as the surface $S_{3,1}$ of a pseudosphere with radius r which is described by

$$x_4^2 - \mathbf{x}^2 + x_0^2 = r^2. \tag{2.1}$$

The group of symmetries of $\mathscr{S}$ is then equivalent to "rotations" of $S_{3,1}$; hence it is isomorphic to SO(3,2) with the Lie algebra

$$[M_{ab},M_{cd}] = i(g_{ac}M_{bd} + g_{bd}M_{ac} - g_{ad}M_{bc} - g_{bc}M_{ad}), \tag{2.2}$$

$$(a,b = 0,1,2,3,4),$$

where $g_{00} = -g_{11} = -g_{22} = -g_{33} = g_{44} = 1$. It is convenient to introduce the notations

$$J_{\mu\nu} \equiv M_{\mu\nu}, \quad \Pi_\mu \equiv r^{-1}M_{4\mu} \quad (\mu,\nu = 0,1,2,3), \tag{2.3}$$

and then (2.2) becomes

$$[J_{\mu\nu},J_{\rho\sigma}] = i(g_{\mu\rho}J_{\nu\sigma} + g_{\nu\sigma}J_{\mu\rho} - g_{\mu\sigma}J_{\nu\rho} - g_{\nu\rho}J_{\mu\sigma}), \tag{2.4a}$$

$$[J_{\mu\nu},\Pi_\rho] = -i(g_{\mu\rho}\Pi_\nu - g_{\nu\rho}\Pi_\mu), \tag{2.4b}$$

$$[\Pi_\mu,\Pi_\nu] = ig_{44}r^{-2}J_{\mu\nu}. \tag{2.4c}$$

This algebra can be realized in the Hilbert space of square integrable functions $\psi(x)$ on $S_{3,1}$ as follows[13]:

$$\Pi_\mu \sim -ir^{-1}(r^2 - g_{\mu\nu}x^\mu x^\nu)^{-1/2}(-x_\mu g_{\nu\rho}x^\nu\partial^\rho + g_{\nu\rho}x^\nu x^\rho\partial_\mu - r^2\partial_\mu), \tag{2.5a}$$

$$J_{\mu\nu} \sim -i(x_\mu\partial_\nu - x_\nu\partial_\mu). \tag{2.5b}$$

As we did in the derivation of $\mathscr{G}_4$, $\mathscr{G}_5$, and $\mathscr{H}_4$, we now formulate the crucial *locality postulate*: We demand that a local phase transformation be a globally, unitarily implementable automorphism of the Hilbert space. More formally, we introduce

Assumption 2: To every transformation

$$\psi(x) \to \exp[i\omega(x)]\psi(x) \tag{2.6}$$

with a differentiable $\omega(x)$ there corresponds in the Hilbert space a unitary symmetry operator U such that

$$(U\psi)(x) = \exp[i\omega(x)]\psi(x). \tag{2.7}$$

Using the realizations (2.5), we calculate

$$(U\Pi_\mu U^{-1}\psi)(x) = \exp[i\omega(x)](-ir^{-1}(r^2 - g_{\mu\nu}x^\mu x^\nu)^{-1/2}) \times(-x_\mu g_{\nu\rho}x^\nu\partial^\rho + g_{\nu\rho}x^\nu x^\rho\partial_\mu - r^2\partial_\mu)\exp[-i\omega(x)],$$

i.e.,

$$\Pi_\mu \to \Pi_\mu + (x_\mu x_\rho\partial^\rho\omega - x_\rho x^\rho\partial_\mu\omega + r^2\partial_\mu\omega) \times[r^{-1}(r^2 - x_\nu x^\nu)^{-1/2}]. \tag{2.8}$$

Similarly we find

$$J_{\mu\nu} \to J_{\mu\nu} - x_\mu\partial_\nu\omega + x_\nu\partial_\mu\omega. \tag{2.9}$$

As in Refs. 2, 3, and 8, we insist that the local phase transformation (2.6) be unitarily implementable, i.e., setting $U \equiv \exp(iF)$ with a self-adjoint F, transformations (2.8) and (2.9) be implementable as

$$\Pi_\mu \to \exp(iF)\Pi_\mu\exp(-iF), \tag{2.10}$$

$$J_{\mu\nu} \to \exp(iF)J_{\mu\nu}\exp(-iF), \tag{2.11}$$

where F should be constructed from the algebra of observables. In other words, we adopt

Assumption 3: The algebra of observables is large enough to guarantee that local phase transformations (2.6), giving the changes (2.8) and (2.9) of the event space symmetry generators, can be realized in the form (2.10) and (2.11) with F being a self-adjoint function on the algebra of observables.

Similar to the case in Refs. 2, 3, and 8, it can be seen that F cannot be expressed as a function of Π_μ and $J_{\mu\nu}$ alone and thus *we must enlarge the algebra of observables*. To see how to do this, we note that from (2.5), (2.8), (2.9) it follows (in lowest order) that

$$i[F,\Pi_\mu] = r^{-1}(r^2 - x_\nu x^\nu)^{-1/2}(x_\mu x_\rho\partial^\rho\omega - x_\rho x^\rho\partial_\mu\omega + r^2\partial_\mu\omega), \tag{2.12}$$

$$i[F,J_{\mu\nu}] = -x_\mu\partial_\nu\omega + x_\nu\partial_\mu\omega. \tag{2.13}$$

Equation (2.12) shows that (unless ω = const.) the commutator $[F,\Pi_\mu]$ must contain at least a c-number term and an operator whose realization in Hilbert space is $x_\mu x_\mu$ (no summation). To gain insight, we take the flat space limit $r \to \infty$ and use the notation

$$\lim_{r\to\infty}\Pi_\mu \equiv \check{\Pi}_\mu, \quad \lim_{r\to\infty}J_{\mu\nu} \equiv \check{J}_{\mu\nu},$$

Then (2.8) and (2.9) become simply

$$\check{\Pi}_\mu \to \check{\Pi}_\mu + \partial_\mu\omega, \tag{2.14a}$$

$$\check{J}_{\mu\nu} \to \check{J}_{\mu\nu} - x_\mu\partial_\nu\omega + x_\nu\partial_\mu\omega. \tag{2.14b}$$

Let us now write the generic form of the gauge function as

$$\omega_{(\mu)} \equiv a_0 + a_\mu x^\mu\sum_{n=1}^{\infty}(a_\sigma x^\sigma)^{n-1},$$

(no summation over μ) and denote the corresponding generator by $F_{(\mu)}$. The flat space limit of this we denote by $\check{F}_{(\mu)}$. Then from Eqs. (2.14) it follows that

$$[\check{F}_{(\mu)},\check{\Pi}_\nu] = ig_{\mu\nu},$$
$$[\check{F}_{(\rho)},\check{J}_{\mu\nu}] = i(g_{\rho\nu}\check{F}_{(\rho)} - g_{\rho\mu}\check{F}_{(\nu)}). \tag{2.15}$$

One would expect that similar relations hold for the curved space quantities. However, as in the nonrelativistic case,[8] one can show that this is not possible. But again the argument used in Ref. 8 applies: We are concerned with *local* transformations $\omega(x)$ and thus it is consistent to only consider *local displacements* that are small compared with r. Equivalently: The appropriate generators can be taken to be the limits (2.14a). From now on we shall write P_μ for $\check{\Pi}_\mu$ (and $J_{\mu\nu}$ for $\check{J}_{\mu\nu}$) and also set $Q_\mu \equiv l^{-1}F_{(\mu)}$, where the constant l has dimension length. From (2.14a) the gauge behavior of P_μ is

$$P_\mu \to P_\mu + \partial_\mu\omega.$$

From (2.4c) it follows that $[P_\mu,P_\nu]=0$. Also, Eq. (2.4a) is unchanged and (2.4b) holds with P_ρ instead of Π_ρ. Finally, using (2.15) and the discussion above, one establishes, in a manner completely paralleling the proof of Theorem 1 in Ref. 2, the Lie brackets for Q_μ. The complete Lie algebra reads as follows:

$$[P_\mu,P_\nu]=[Q_\mu,Q_\nu]=0, \tag{2.16a}$$

$$[J_{\mu\nu},P_\rho]=-i(g_{\mu\rho}P_\nu-g_{\nu\rho}P_\mu), \tag{2.16b}$$

$$[J_{\mu\nu},Q_\rho]=-i(g_{\mu\rho}Q_\nu-g_{\nu\rho}Q_\mu), \tag{2.16c}$$

$$[J_{\mu\nu},J_{\rho\sigma}]=-i(g_{\mu\rho}J_{\nu\sigma}+g_{\nu\sigma}J_{\mu\rho}-g_{\mu\sigma}J_{\nu\rho}-g_{\nu\rho}J_{\mu\sigma}), \tag{2.16d}$$

$$[P_\mu,Q_\nu]=-il^{-1}g_{\mu\nu}. \tag{2.16e}$$

For this we see that *the structure of our kinematic group is*

$$K=\mathrm{SL}(2,C)'\otimes[T_4^P\otimes(T_4^Q\times T_1^{I\ '})], \tag{2.17}$$

which is identical to the one for flat Minkowski event-space, cf. Ref. 3. The realization of the algebra K is given by

$$P_\mu\sim i\partial_\mu, \tag{2.18a}$$

$$Q_\mu\sim -l^{-1}x_\mu, \tag{2.18b}$$

$$J_{\mu\nu}\sim i(x_\mu\partial_\mu-x_\nu\partial_\mu)+\Sigma_{\mu\nu}, \tag{2.18c}$$

where $\Sigma_{\mu\nu}$ is an $\mathrm{SL}(2,C)$ spin matrix. At this point, the reader is advised to consult, for interpretation and further comments, p. 2054 in Ref. 3.

We now proceed to introduce dynamics, as was done in Refs. 2, 3, and 8, via

Definition 1: A development transformation of an isolated system is a kinematical symmetry[14] characterized by

$$J_{\mu\nu}\to J_{\mu\nu},\quad P_\mu\to g(Q_\mu,P_\mu,J_{\rho\sigma}),\quad Q_\mu\to f(Q_\mu,P_\mu,J_{\rho\sigma}). \tag{2.19}$$

The motivation is that the geometry of the event space requires that the generator(s) of intrinsic development transformations be invariant under Lorentz transformations but, unlike the case for a flat space, they need not be invariant under arbitrary large translations.

Motivated by obvious arguments fully analogous to those in Refs. 2, 3, and 8 we make the following postulates:

Assumption 4: Development transformations form a one-parameter Lie group T_1^S and thus they are represented by $U_\sigma=\exp(i\sigma S)$.

Assumption 5: S is contained in the algebra generated by $P_\mu,Q_\mu,J_{\rho\sigma}$.

These assumptions, together with the invariance requirements implied by Definition 1, determine the form of the development operator,

$$S=S(P^2,Q^2,QP,TP,TQ,I), \tag{2.20}$$

where, as in Refs. 3 and 5, the $\mathrm{SL}(2,C)$-spin T is defined by[15]

$$T_{\mu\nu}\equiv J_{\mu\nu}-l(Q_\mu P_\nu-Q_\nu P_\mu). \tag{2.21}$$

The development transformations give rise to an equivalence relation on the algebra of observables generated by K, so that, as in our previous work, we can define a *dynamical group* G by

Assumption 6: The kinematical group K is isomorphic to the quotient group modulo T_1^S of some Lie group G.

From $K\approx G/T_1^S$ it then follows that S and the generators of K must form a closed Lie algebra. Therefore the form (2.20) will be restricted to have the following structure:

$$S=AP^2+BQ^2+C(PQ+QP)+D, \tag{2.22}$$

with A, B, C, D being real constant c-numbers. We then see that the Lie algebra of G has [in addition to (2.16a)–(2.16d)] the commutators

$$[S,P_\mu]=i2l^{-1}(BQ_\mu+CP_\mu), \tag{2.23a}$$

$$[S,Q_\mu]=-i2l^{-1}(CQ_\mu+AP_\mu), \tag{2.23b}$$

$$[S,J_{\mu\nu}]=0. \tag{2.23c}$$

In order to fix the as yet undetermined constants in (2.22) we posit, motivated by the same intuitive ideas spelled out in Ref. 8, the following:

Assumption 7: The transformation T corresponding to inversion of dynamical development,

$$T\colon U_\sigma\to U_{-\sigma}, \tag{2.24}$$

is a kinematical symmetry.[14]

Assumption 8: The operator T of development inversion is invariant under local phase transformations,

$$\exp[i\omega(Q)]T\exp[-i\omega(Q)]=T. \tag{2.25}$$

From (2.24) and Assumption 4 it follows that

$$(iS)'\equiv T(iS)T^{-1}=-iS. \tag{2.26}$$

Transforming (2.23) with T and using (2.25) for the case of a linear gauge transformation, we find, by arguments in complete analogy to the work of Ref. 8, that

$$A=-\tfrac{1}{2}l,\quad B=-\tfrac{1}{2}l^3\nu^2,\quad C=0,$$

so that (2.22) becomes

$$S=-\tfrac{1}{2}lP^2-\tfrac{1}{2}l^3\nu^2Q^2+D, \tag{2.27}$$

where ν is a constant (determining a scale of units) and D is an arbitrary constant.[16] Furthermore, the final form of the Lie brackets (2.23) will now be

$$[S,P_\mu]=-i\nu^2l^2Q_\mu, \tag{2.28a}$$

$$[S,Q_\mu]=iP_\mu, \tag{2.28b}$$

$$[S,J_{\mu\nu}]=0. \tag{2.28c}$$

In summary: the Lie algebra of the dynamical group G is given by (2.16) and (2.28). Thus, its structure is

$$G\equiv\tilde{\mathscr{H}}_5$$

$$=T_1^S\otimes K=T_1^S\otimes[(SL(2,C)'\otimes T_4^Q)\otimes(T_4^P\times T_1^{I\ '})]. \tag{2.29}$$

Clearly, this is *a relativistic generalization of the centrally extended Hooke group* and we shall denote it by $\tilde{\mathscr{H}}_5$. We note that the only difference between the algebra of[3,4] $\mathscr{G}_5$ and that of $\tilde{\mathscr{H}}_5$ is the commutator (2.28a).

We call the reader's attention to p. 1667 of Ref. 8 and point out that entirely analogous remarks now hold for the relativistic case as well. In particular, we see that "relativistic cosmologic time" can be interpreted, via our construction, in a purely group theoretical manner. Finally, we note that T has the meaning of cosmologic time reversal. Its action is characterized by

$$Q'_\mu\equiv TQ_\mu T^{-1}=Q_\mu,\quad P'_\mu\equiv TP_\mu T^{-1}=-P_\mu, \tag{2.30}$$

and it is antilinear. Thus, from (2.26) it follows that[17]

$$S' \equiv TST^{-1} = S. \tag{2.31}$$

We conclude this section by presenting a realization of the Lie algebra (2.16), (2.28) of $\mathscr{H}_5$, given on a Hilbert space of square integrable functions defined on $E_{3,1}(p) \times E_1(\tau)$. Here $E_{3,1}(p)$ is the Fourier-dual (momentum space) of the kinematical group K. The additional variable τ labels the sequence ("slices") of Hilbert spaces that are strung together in the big Hilbert space by means of the dynamical development T_1^S. We find

$$P_\mu \sim \cos(\nu\tau)p_\mu \cdot - il^{-1}\nu \sin(\nu\tau)\partial_{p_\mu}, \tag{2.32a}$$

$$Q_\mu \sim il^{-1} \cos(\nu\tau)\partial_{p_\mu} + \nu^{-1} \sin(\nu\tau)p_\mu \cdot, \tag{2.32b}$$

$$J_{\mu\nu} \sim i(\partial_{p_\nu} \cdot p_\mu - \partial_{p_\mu} \cdot p_\nu) + \Sigma_{\mu\nu}, \tag{2.32c}$$

$$S \sim i\partial_\tau. \tag{2.32d}$$

3. DISCUSSION AND REPRESENTATION THEORY OF THE RELATIVISTIC HOOKE GROUP

3.1 Group properties and Casimir invariants

It is convenient to expresses the structure (2.29) of the locally compact 16 parameter group $\mathscr{H}_5$ in the isomorphic form

$$\mathscr{H}_5 = (T_1^S \times SL(2,C)^J) \otimes [T_4^Q \otimes (T_4^P \times T_1^{I\ '})]. \tag{3.1}$$

The generic element of $\mathscr{H}_5$ can be written as

$$\tilde{h} \equiv (\Theta;h) = (\Theta;\sigma,a,b,\Lambda), \tag{3.2}$$

where Θ is a phase parameter [associated with the central extension by $T_1^{I\ '}$, cf. Eq. (2.16)] and σ,a,b,Λ are the parameters associated with the generators S of dynamical development, P of translations, Q of "relativistic Hooke boosts," and J of Lorentz transformations, respectively. The *composition law* is

$$\tilde{h}_1\tilde{h}_2 = (\Theta_1 + \Theta_2 + \xi(h_1,h_2);h_1h_2) \tag{3.3}$$

with

$$h_1h_2 = (\sigma_1 + \sigma_2,\ a_1 \cos\nu\sigma_2 + \nu^{-1}b_1 \sin\nu\sigma_2 + \Lambda_1a_2,\ b_1 \cos\nu\sigma_2 - \nu a_1 \sin\nu\sigma_2 + \Lambda_1b_2,\ \Lambda_2\Lambda_2) \tag{3.4a}$$

and

$$\xi(h_1,h_2) = l^{-1}\{[(2\nu)^{-1}b_1^2 - \tfrac{1}{2}\nu a_1^2]\cos\nu\sigma_2 \sin\nu\sigma_2 + a_1b_1 \cos^2\nu\sigma_2 + \Lambda_1a_2(b_1\cos\nu\sigma_2 - a_1\nu \sin\nu\sigma_2)\}. \tag{3.4b}$$

The Casimir invariants are

$$\mathscr{R}_0 = l^{-1}I, \tag{3.5a}$$

$$\mathscr{R}_1 = \tfrac{1}{2}T_{\mu\nu}T^{\mu\nu}, \tag{3.5b}$$

$$\mathscr{R}_2 = \tfrac{1}{4}\epsilon_{\mu\nu\rho\sigma}T^{\mu\nu}T^{\rho\sigma}, \tag{3.5c}$$

$$\mathscr{R}_3 = P^2 + \nu^2l^2Q^2 + 2l^{-1}S. \tag{3.5d}$$

The first three coincide with the Casimir invariants of $\mathscr{G}_5$ (cf. Refs. 3–5), because only the $[S,P_\mu]$ commutator of $\mathscr{H}_5$ differs from those of $\mathscr{G}_5$. The invariant $\mathscr{R}_3$ can be inferred from contraction arguments (cf. Ref. 18, Appendix C) and it differs from the $\mathscr{D}_3$ invariant of $\mathscr{G}_5$ only by the presence of the Q^2 term. From (3.5d) we see that now S will have a *discrete spectrum*, in contradistinction to the case of $\mathscr{G}_5$. This has crucial physical consequences, to which we shall return in Sec. 4.

3.2 Relation of $\mathscr{H}_5$ to other relativistic groups

It is known[6] that the nonrelativistic Hooke group can be considered as a particular contraction of SO(3,2), with respect to the rotation group SO(3). Similarly, it is expected that the relativistic Hooke group can be viewed as some contraction of a larger homogeneous pseudo-orthogonal group, with respect to its SO(3,1) Lorentz subgroup. As was shown in Ref. 18, p.46, the well-known group extension method of Rosen[19] shows that the parent group must be a pseudo-orthogonal group of dimension six, i.e., SO(p,q) with $p+q=6$. In order to study the relation of $\mathscr{H}_5$ to other quantum mechanical and possible dynamical groups, we consider below the entire family of groups arising from systematically performed contractions of SO(p,q) with $p+q=6$, all done with respect to the Lorentz subalgebra. Let $M_{\alpha\beta}$ ($\alpha,\beta = 0,1,2,3,4,5$) denote the generators of SO(p,q), and write

$$M_{4\mu} \equiv P_\mu, \quad M_{5\mu} \equiv Q_\mu, \quad M_{45} \equiv S. \tag{3.6}$$

Then the SO(p,q) algebra reads[20]

$$[M_{\mu\nu},M_{\rho\sigma}] = i(g_{\mu\rho}M_{\nu\sigma} + g_{\nu\sigma}M_{\mu\rho} - g_{\mu\sigma}M_{\nu\rho} - g_{\rho\nu}M_{\mu\sigma}), \tag{3.7a}$$

$$[P_\mu,P_\nu] = ig_{44}M_{\mu\nu}, \quad [Q_\mu,Q_\nu] = ig_{55}M_{\mu\nu}, \tag{3.7b}$$

$$[P_\mu,Q_\nu] = ig_{\mu\nu}S, \tag{3.7c}$$

$$[S,Q_\mu] = -ig_{55}P_\mu, \quad [S,P_\mu] = ig_{44}Q_\mu, \tag{3.7d}$$

$$[M_{\mu\nu},P_\rho] = i(g_{\mu\rho}P_\nu - g_{\nu\rho}P_\mu), \tag{3.7e}$$

$$[M_{\mu\nu},Q_\rho] = i(g_{\mu\rho}Q_\nu - g_{\nu\rho}Q_\mu), \tag{3.7f}$$

$$[M_{\mu\nu},S] = 0, \tag{3.7g}$$

where $\mu,\nu = 0,1,2,3$.

If we are interested in Lorentz subalgebra-preserving Wigner–Inönü contractions that also preserve the Lorentz transformation character of P,Q,S, then we have three distinct choices[21]:

(a) *Speed-Space contraction*: Replace Q by ϵQ, P by ϵP, and take the limits

$$\lim_{\epsilon \to 0} \epsilon Q = Q', \quad \lim_{\epsilon \to 0} \epsilon P = P'.$$

(b) *Space–Time contraction*: Replace P by ϵP, S by ϵS, and take the limits

$$\lim_{\epsilon \to 0} \epsilon P = P', \quad \lim_{\epsilon \to 0} \epsilon S = S'.$$

(c) *Speed-Time contraction*: Replace Q by ϵQ, S by ϵS, and take the limits

$$\lim_{\epsilon \to 0} \epsilon Q = Q', \quad \lim_{\epsilon \to 0} \epsilon S = S'.$$

Explicit calculation then gives the following P,Q,S subalgebras.[22]

1. *Speed-Space contraction gives*[23]:

$$[P_\mu,P_\nu] = [Q_\mu,Q_\nu] = [P_\mu,Q_\nu] = 0,$$
$$[S,Q_\mu] = -ig_{55}P_\mu, \quad [S,P_\mu] = ig_{44}Q_\mu. \tag{3.8}$$

If we take $g_{44}=g_{55}=g_{00}=1$, corresponding to SO(3,3), or alternatively $-g_{44}=-g_{55}=g_{00}=1$, corresponding to SO(5,1), then we have (two possible isomorphic versions of) our oscillatory *relativistic Hooke group*[24] $\mathscr{H}_5$. Naturally, we now have only the "geometrical" group (with $[P,Q]=0$) and not its central extension $\tilde{\mathscr{H}}_5$. (To obtain $\tilde{\mathscr{H}}_5$, one must use a contraction procedure where the rhs of $[P,Q]$ contracts to a constant. The way to do this has been done in Appendix C or Ref. 18.)

If, on the other hand, we take $g_{44}=-g_{55}=g_{00}=1$, corresponding to SO(4,2), then (3.8) becomes the algebra of *Castell's "preferred algebra II"*, which he calls the macroscopic group.[25] To fully appreciate the difference in interpretation between our $\tilde{\mathscr{H}}_5$ and Castell's group II, one must note that Castell considered the conformal group with the *physical* generators $\hat{P}_\mu,\hat{Q}_\mu,\hat{D},\hat{J}_{\mu\nu}$, and the SO(4,2) generators contracted over are linear combinations of these. The correspondence between our P_μ,Q_μ,S and the conformal (or SO(4,2)] generators is as follows:

$$M_{4\mu}=P_\mu=\tfrac{1}{2}(\hat{P}_\mu+\hat{K}_\mu),\quad M_{5\mu}=Q_\mu=\tfrac{1}{2}(\hat{P}_\mu-\hat{K}_\mu),$$
$$M_{45}=S=\hat{D},\quad M_{\mu\nu}=\hat{J}_{\mu\nu}.$$

2. *Space–Time contraction gives*:

$$[P_\mu,P_\nu]=0,\quad [Q_\mu,Q_\nu]=ig_{55}M_{\mu\nu},\quad [P_\mu,Q_\nu]=g_{\mu\nu}S,$$
$$[S,Q_\mu]=-ig_{55}P_\mu,\quad [S,P_\mu]=0. \tag{3.9}$$

This structure is isomorphic to the algebra of the inhomogeneous para-de Sitter group.[27]

3. *Speed-Time contraction gives*:

$$[P_\mu,P_\nu]=ig_{44}M_{\mu\nu},\quad [Q_\mu,Q_\nu]=0,\quad [P_\mu,Q_\nu]=g_{\mu\nu}S,$$
$$[S,Q_\mu]=0,\quad [S,P_\mu]=ig_{44}Q_\mu. \tag{3.10}$$

This is isomorphic to the inhomogeneous de Sitter algebra.

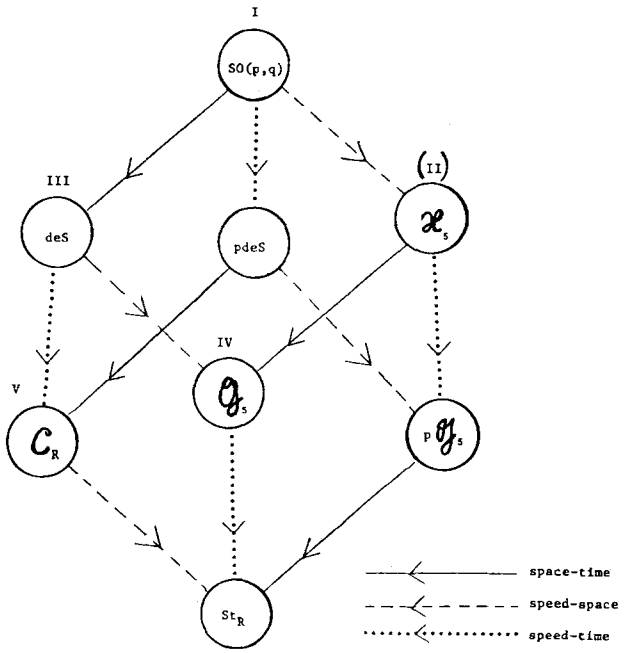


FIG. 1. Results of contractions.

One can now perform further contractions of appropriate type (a), (b), (c). The results are represented in Fig. 1. (Notation: deS = inhomogeneous de Sitter; C_R = relativistic Carroll; St_R = relativistic static. The prefix "p" on deS and $\mathscr{G}_5$ is short for "para."[27] The Roman numerals I–V correspond to Castell's notation.[25]) The corresponding subalgebras are

Relativistic Galilei group $\mathscr{G}_5$:

$$[P_\mu,P_\nu]=[Q_\mu,Q_\nu]=[P_\mu,Q_\nu]=0,$$
$$[S,Q_\mu]=0,\quad [S,P_\mu]=ig_{44}Q_\mu. \tag{3.11}$$

Relativistic para-Galilei group $p\mathscr{G}_5$:

$$[P_\mu,P_\nu]=[Q_\mu,Q_\nu]=[P_\mu,Q_\nu]=0,$$
$$[S,Q_\mu]=-ig_{55}P_\mu,\quad [S,P_\mu]=0. \tag{3.12}$$

Relativistic Carroll group C_R:

$$[P_\mu,P_\nu]=[Q_\mu,Q_\nu]=0,\quad [P_\mu,Q_\nu]=ig_{\mu\nu}S,$$
$$[S,Q_\mu]=[S,P_\mu]=0. \tag{3.13}$$

Relativistic static group St_R:

$$[P_\mu,P_\nu]=[Q_\mu,Q_\nu]=[P_\mu,Q_\nu]=0,$$
$$[S,Q_\mu]=[S,P_\mu]=0. \tag{3.14}$$

3.3 Representations of $\tilde{\mathscr{H}}_5$

The irreducible unitary ray representations of $\tilde{\mathscr{H}}_5$ can be obtained by Mackey's method of induced representations. The quickest specific approach is to find a relativistic generalization of the work by Dubois[28] who constructed the representations of the nonrelativistic $\tilde{\mathscr{H}}_4$. We sketch[29] the method as appropriate for $\tilde{\mathscr{H}}_5$.

We write

$$\tilde{\mathscr{H}}_5=T_1^S\times\tilde{\Gamma}, \tag{3.15}$$

where

$$\tilde{\Gamma}\equiv \mathrm{SL}(2,C)^J\otimes[T_4^Q\otimes(T_4^P\times T_1^{I\ 1})] \tag{3.16}$$

is the maximal invariant subgroup. We can induce the irreps of $\tilde{\mathscr{H}}_5$ from those of $\tilde{\Gamma}$; the irreps of $\tilde{\Gamma}$, in turn, can be induced from those of the Abelian subgroup

$$\tilde{T}_4^P\equiv T_1^{I\ 1}\times T_4^P, \tag{3.17}$$

i.e., we consider $\tilde{\Gamma}$ in the form

$$\tilde{\Gamma}=[\mathrm{SL}(2,C)^J\otimes T_4^Q]\otimes T_4^P. \tag{3.18}$$

The irreps of $\tilde{T}_4^P$ are, obviously, of the form[30]

$$U(\theta,a)=\exp[i(\eta\theta+pa)], \tag{3.19}$$

where η is a real number and p a 4-vector.[31] Let us denote $\mathrm{SL}(2,C)^J\otimes T_4^Q\equiv\Delta$ and label $\delta\epsilon\Delta$ by the pair (Λ,b). If we define the linear functional

$$\langle\eta,p|\theta,a\rangle=\eta\theta+pa,$$

then the homomorphisms of Δ in the group of automorphisms of $\tilde{T}_4^P$ are associated with the action of Δ on $\tilde{T}_4^P$ by the rule

$$\langle\delta(\eta,p)|\theta,a\rangle=\langle\eta,p|\delta^{-1}(\theta,a)\rangle$$

for every $\delta\epsilon\Delta$, $(\eta,p)\epsilon\tilde{T}_4^P$. The transformation induced by δ

(η,p) is given by

$$\delta(\eta,p) = (\eta, \Lambda^{-1}p - \eta l^{-1}\Lambda^{-1}b). \tag{3.20}$$

For each value of η the induced automorphisms of $\tilde{T}_4^P$ yield an orbit. Their collection provides a partition of $\tilde{T}_4^P$. We take a point (η_0,p_0) on the orbit η_0 and wish to construct the irreps of the little group Δ_{p_0} associated to that point. Since the structure of the little group does not depend on the orbit nor on the particular point chosen on it, we may take $\eta = 1$ and chose $p_0 = 0$. We then find that the little group is isomorphic to an SL(2,C) group.[32] Thus, putting things together, the representation of $\tilde{\Gamma}$ can be described, on a suitable function space, as follows[33]:

$$[U(\theta,a,b,\Lambda)\psi^{lpkc}_{ss_3}](p)$$
$$= \exp[i(\theta + pa)]D^{kc}(\Lambda)_{ss_3,s's_3'}$$
$$\times \psi^{lpkc}_{s's_3'}(\Lambda^{-1}(p - l^{-1}b)). \tag{3.21}$$

Here D^{kc} is the well-known matrix representation of SL(2,C)^T.

To induce now the irreps of $\tilde{\mathscr{H}}_5$ from those of $\tilde{\Gamma}$, we must find the orbits in the space $\hat{\tilde{\Gamma}}$ of equivalence classes of irreps of $\tilde{\Gamma}$, which arise under the action of T_1^S. Denoting by $\tilde{\gamma}$ an element of $\tilde{\Gamma}$, it can be shown that there exists an equivalence operator W such that

$$U[\sigma(\tilde{\gamma})] = W(\sigma)U(\tilde{\gamma})W^{-1}(\sigma),$$

with W given by

$$W(\sigma) = \exp(iE'\sigma),$$

where

$$E' \equiv \tfrac{1}{2}lp^2 + \tfrac{1}{2}l^3\nu^2q^2, \tag{3.22}$$

and where q is a 4-vector. Therefore, every point of $\tilde{\Gamma}$ is an orbit in itself, whose little group is (isomorphic to) T_1^S. The irreps of the Abelian T_1^S are, of course, of the form

$$U(\sigma) = \exp(-iu\sigma), \tag{3.23}$$

where u is a real constant. Putting all these results together, we finally obtain the explicit action of $\tilde{\mathscr{H}}_5$ on a suitable function space over the variables p and u as follows:

$$[U(\theta;\sigma,a,b,\Lambda)\psi^{lpkc}_{ss_3}](p;u)$$
$$= \exp\{i[\theta + (\tfrac{1}{2}lp^2 + \tfrac{1}{2}l^3\nu^2q^2 - u)\sigma + pa]\}$$
$$\times D^{kc}(\Lambda)_{ss_3,s's_3'}\psi^{lpkc}_{s's_3'}(\Lambda_p^{-1} - l^{-1}\Lambda^{-1}b;u). \tag{3.24}$$

The "q_μ" in the exponent must be interpreted as the differential operator $il^{-1}(\partial/\partial p_\mu)$ acting on $\psi(p)$.

There exists another, alternative procedure to derive the irreps of $\tilde{\mathscr{H}}_5$. This constuction, based on the generalization of the representation theory of $\tilde{\mathscr{G}}_5$ as presented in a previous work,[34] has the advantage that it gives a deeper insight; a detailed account of it can be found on pp. 53–65 or Ref. 18. In this approach the starting point is the decomposition

$$\tilde{\mathscr{H}}_5 = N \otimes H, \tag{3.25}$$

with

$$N = (T_1^{l\,\prime} \times T_4^P) \otimes T_4^Q, \quad H = \mathrm{SL}(2,C)^J \times T_1^S \tag{3.26}$$

and the irreps of $\tilde{\mathscr{H}}_5$ are induced from those of N. Here we want to use this approach only to discuss the *classes of irreps* that $\tilde{\mathscr{H}}_5$ can have.

The automorphisms in N given by $n \to hnh^{-1}$ (with $h \in H$) define orbits in N characterized by

$$p^2 + \nu^2 l^2 q^2 = \text{const.} \tag{3.27}$$

One then finds that the little group is isomorphic to SL(2,C)$^T \times T_1^S$. Under its action, all points of the orbit can be reached. Examining, in particular, the action of the T_1^S subgroup on points of the orbit, one can show that

$$p \to p\cos\nu\sigma - l^{-1}q\sin\nu\sigma.$$

In particular, taking $\sigma = \pi/2\nu$, we have $p \to l^{-1}q$. Thus, if on a certain orbit $p^2 > 0$ everywhere, then also $q^2 > 0$ everywhere on the orbit, and similar statements hold for $p^2 < 0$, $p^2 = 0$. Because of the disconnectedness of SL(2,C), timelike vectors cannot be transformed into spacelike vectors, etc. Taking these comments into account as well as the fact that, because of ray equivalence, one can always choose $\mathscr{P}_3 = 0$, Eq. (3.27) easily leads to the following classification of the irreps[35]:

Class I: $u > 0$, $p^2 > 0$, $q^2 > 0$,

Class II: $u < 0$, $p^2 < 0$, $q^2 < 0$,

Class III: $u = 0$, $p^2 = q^2 = 0$, $p \neq 0$, $q \neq 0$,

Class IV: $u = 0$, $p^2 = q^2 = 0$, $p = 0$, $q = 0$.

Here u is the eigenvalue of S. For physical applications, we are interested only in Class I.

4. PHYSICAL CONSEQUENCES OF $\tilde{\mathscr{H}}_5$

4.1 Wave equation and mass spectrum

In conformity with Section 1 of this paper and in the spirit of Ref. 9 we consider now the relativistic hadrons as bilocal objects: a microuniverse bubble embedded in the (flat) external Minkowski world. As in most bilocal models, the relationship between the external and internal dynamical entities is not obvious and is expected to be model dependent. There exists, however, an algebraic approach, due to Nambu,[36] which yields a general, plausible, and easily applicable prescription to combine the internal and external symmetries. We shall use this approach to determine the internal symmetry group in terms of the Hooke generators.

Nambu's procedure entails two essential criteria for the internal group:

(a) it contains a subgroup characterizing internal symmetry.

(b) it is large enough to contain within its generators elements that are fourvectors or tensors under the Lorentz group.

In general, one will then have infinite multiplets, and one can construct Lorentz invariant wave equations with the internal generators coupled to the external momentum.

In the spirit of requirement (b) we see that the $\tilde{\mathscr{H}}_5$ generators available for the internal symmetry are the fourvectors P_μ, Q_μ, and the scalar S. To meet condition (a), it is convenient to define

$$A_\mu^\dagger \equiv (\nu/\sqrt{2})(lQ_\mu - i\nu^{-1}P_\mu),$$

$$A_\mu = (\nu/\sqrt{2})(lQ_\mu + i\nu^{-1}P_\mu). \tag{4.1}$$

Then it is easily checked that the sixteen bilinears $A^\dagger_\mu A_\nu$ in the enveloping algebra form a U(3,1) Lie algebra. In order to have a semisimple group, we fix the eigenvalue of

$$N \equiv A^\dagger_\mu A^\mu. \tag{4.2}$$

Actually, the eigenvalues of N are

$$n \equiv n_1 + n_2 + n_3 - n_0, \tag{4.3}$$

where n_μ are nonnegative integers. Fixing N leads (via the elimination of trace) from U(3,1) to SU(3,1). This is our internal symmetry. We note that fixing N is equivalent to fixing S since from (3.5d) we see that (in a given representation)

$$S = -l\nu(N+1) + \tfrac{1}{2}l\mathscr{R}_3. \tag{4.4}$$

The SU(3,1) generators $A^\dagger_\mu A_\nu$ form a tensor, the $A^\dagger_\mu + A_\mu$ behaves as a (self-adjoint) vector, and S (essentially $A^\dagger_\mu A^\mu$) behaves as a scalar under Lorentz transformations. Thus, the most general Nambu-type equation will be

$$[A^\dagger_\mu A_\nu \mathsf{P}^\mu \mathsf{P}^\nu + \alpha(A^\dagger_\mu + A_\mu)\mathsf{P}^\mu + \beta\mathsf{P}^2 - \gamma S]\Psi^{n,b}(\mathsf{P}) = 0. \tag{4.5}$$

Here P^μ denotes the external fourmomentum of the system. The wave function is an (infinite component) multiplet member, with level-label n and degeneracy label b, and external Poincaré label suppressed. Since, by our above argument, the eigenvalue of S is fixed,[37] we can identify [because of (4.4)] the label n with the eigenvalue (4.3) of N.

The action of $A^\dagger_\mu A_\nu$ is simply to transform the degenerate eigenvectors of S among themselves, for a given level. But $A^\dagger_\mu$ and A_μ, being raising–lowering operators, will transform eigenvectors of S with a label n to those with labels $n+1$ and $n-1$, respectively. Thus, (4.5) will be an eigenvalue equation for S eigenstates only if $\alpha = 0$. Then we have, finally, the wave equation

$$(A^\dagger_\mu A_\nu \mathsf{P}^\mu \mathsf{P}^\nu + \beta\mathsf{P}^2 - \gamma S)\Psi^{n,b}(\mathsf{P}) = 0. \tag{4.6}$$

To extract physical information from (4.6), we go to the rest system, $\mathsf{P}_k = 0$, $\mathsf{P}_0 = m$, where m is the hadron mass. We have

$$(A^\dagger_0 A_0 m^2 + \beta m^2 - \gamma S)\Psi^{n,b}(\mathsf{P}_0) = 0. \tag{4.7}$$

Because of ray equivalence, from now on we will take, without restricting generality, $\mathscr{R}_3 = 0$. Then from (4.7), (4.4), and (4.3) we obtain

$$[(n_0+\beta)m^2 + \gamma l\nu(n_1+n_2+n_3-n_0+1)]\Psi^{n,b} = 0,$$

i.e., we have the *mass spectrum*

$$m^2 = -\gamma l\nu(n_1+n_2+n_3-n_0+1)/(n_0+\beta). \tag{4.8}$$

We are interested in Class I representations with $m^2 = \mathsf{P}^2 > 0$, where, according to the discussion at the end of Sec. 3.3 the eigenvalue u of S is positive. From (4.4) (with $\mathscr{R}_3 = 0$) it follows that this is the case if[35]

$$n_0 < n_1 + n_2 + n_3 + 1. \tag{4.9}$$

Therefore, if $\gamma > 0$, we must have $\beta > -n_0$ and, if $\gamma < 0$, we must have $\beta < -n_0$ in order to have $m^2 > 0$.

Clearly, we now have a *discrete mass spectrum*. However, even though the restriction to Class I avoids the emergence of $m^2 \leqslant 0$ solutions, the timelike oscillations still lead to a feature of the spectrum which is contradicting even qualitative experience: taking in (4.8) the maximum possible values $n_0 = n_1 + n_2 + n_3$, the spectrum of m^2 has an accumulation point at $m^2 = 0$. Hence, even though our general analysis does not explain it on theoretical grounds, we will *assume that timelike oscillations are suppressed*, $n_0 = 0$. Equation (4.8) then gives the reasonable mass spectrum

$$\begin{aligned} &m^2 = \text{const}(n+1), \\ &n \equiv n_1 + n_2 + n_3, \\ &n_i = \text{non-negative integer}. \end{aligned} \tag{4.10}$$

4.2 Relations with other hadron models

The (degenerate) eigenfunction $\Psi^{n,b}$ in the Nambu type Eq. (4.6) is, in general, a member of an infinite dimensional multiplet in the internal space, and each member of this multiplet is an irrep of the Poincaré group. We can write

$$\Psi^{n,b}(\mathsf{P}) \equiv \phi^n \psi_a(\mathsf{P}) \tag{4.11}$$

where ϕ^n is an internal state vector and ψ_a is a wave function in external space with Poincaré label "a". Furthermore, we can express ϕ^n as some wave function in the internal (Hooke) momentum space,

$$\phi^n \equiv \phi^n(p).$$

It is essentially a product of four Hermite plynomials. Since n reflects the eigenvalue u of S, we may conveniently write

$$\phi^n = \phi(u;p) \tag{4.12a}$$

where,[38] from (4.4),

$$u = -l\nu(n_1+n_2+n_3-n_0+1). \tag{4.12b}$$

The composite, bilocal wave function of the entire "hadron" = Hooke bubble in Minkowski space can now be written

$$\Phi^n(p,\mathsf{P}) = \phi(u,p)\psi_a(\mathsf{P}). \tag{4.13}$$

For simplicity we assume that the external wave function belongs to the scalar representation of the Poincaré group,

$$(\mathsf{P}^2 - m^2)\psi_a(\mathsf{P}) = 0. \tag{4.14}$$

For the internal space part we obtain, when using the realization[39] $P_\mu \sim p_\mu, Q_\mu \sim il^{-1}\partial_{p_\mu}$ of the Casimir invariant[38] (3.5d)

$$(p^2 - \nu^2\Box_p + 2l^{-1}u)\phi(u;p) = 0. \tag{4.15}$$

The general relation between m^2 and the eigenvalues u and S as given by (4.12b) is expressed by the formula (4.8), i.e.,

$$m^2 = \gamma u/n_0 + \beta. \tag{4.16}$$

The Equations (4.14) and (4.15) are identical in form to those which emerge in the relativistic harmonic oscillator model of Feynman *et al.*[40] and which were further developed (connected to the parton model) by Kim and Noz.[41] In the work of these authors, the starting point is a system of "quarks" which are *assumed* to interact via relativistic harmonic forces. The equations in question are then obtained by introducing center-of-mass and relative coordinates, but to obtain Poincaré invariant c.m. motion equations it is necessary to *forbid* oscillations of the center of mass ("spurious

oscillations"). In addition, timelike oscillations must be suppressed in order to avoid tachyon solutions. In our approach, we do not have c.m. oscillations, so that there is nothing to artificially forbid. Also, the m^2 is positive in our model without suppressing timelike oscillations (but in order to avoid an accumulation point at $m^2 = 0$ we also have to use the device of suppression). It is clear that, because the emergence of Eq. (4.15), all results and conclusions of Refs. 40 and 41 will also hold in our framework.

We now use (4.15) to study the degeneracy of the mass levels. As before, we assume that timelike oscillations are to be suppressed, which means

$$A_0 A^{\dagger}_0 \phi(u;p) = A_0 A^{\dagger}_0 \phi(u;p) = 0.$$

Then (4.15) becomes

$$(p^2 - \nu^2 \nabla_p^2 + 2l^{-1}u)\phi(u;p) = 0, \tag{4,17a}$$

which is, of course, a three-dimensional harmonic oscillator wave equation and which in configuration space reads

$$(-\nabla_x^2 + \nu^2 x^2 + 2l^{-1}u)\phi(u;x) = 0. \tag{4.17b}$$

Introducing spherical polar coordinates and separating, the radial wave equation becomes

$$\left[\frac{d^2}{dr^2} - \frac{j(j+1)}{r^2} - \nu^2 r^2 - 2l^{-1}u\right] R^j_k(u;r) = 0. \tag{4.18}$$

The Laguerre polynomials R^j_k are characterized, for given j, by the integer $k = 0,1,...$ and the value of u. The physically acceptable (normalizable) solutions are those for which

$$-2l^{-1}u = \nu(2k + j + 1).$$

Since, from (4.12b) (with $n_0 = 0$) we have

$$2l^{-1}u = -2\nu(n+1) \quad (n \equiv n_1 + n_2 + n_3),$$

we see that

$$n = k + j/2. \tag{4.19}$$

Substituting this into the mass formula (4.10), we get

$$m^2 = C^{-1}(j + 2k + 2), \text{ i.e.,}$$
$$j = Cm^2 - 2k - 2, \quad k = 0,1,2,..., \tag{4.20}$$

where C is a constant. Thus, interpreting j (which is the "internal orbital angular momentum") as the spin of the physical hadron, we have a family of *Regge trajectories* with recurrences at intervals of $2j$. In particular, trajectories of opposite signature coincide. However, the above formula cannot be directly compared with experiments, because, by its derivation, it refers to a single collective excitation of the de Sitter microworld bubble whereas, as we discussed in detail in Ref. 9, hadrons correspond to multiple excitations. The nonarbitrary value of $\mathscr{R}_3$ for irreducible terms in the direct product of representations will then be reflected in the modification of the intercept of the trajectory, i.e., instead of (4.20) we would have

$$j = Cm^2 - 2k + B, \tag{4.21}$$

where, together with C, the B is also an undetermined constant.[42] It is possible, but perhaps not too profitable, to compare (4.21) with, say, the known meson spectrum.

ACKNOWLEDGMENTS

One of us (P.R.) is greatly indebted to Professors C.F. von Weizsäcker and L. Castell of the Max Planck Institut who extended their hospitability during the time when this paper was written. Very illuminating and stimulating discussions with the second-named colleague are also appreciated.

[1]For a comprehensive review see J.-M. Lévy-Leblond, in *Group Theory and Its Applications*, edited by E.M. Loebl (Academic, New York, 1971), Vol. II.

[2]P. Roman and J.P. Leveille, J. Math. Phys. **15**, 1760 (1974).

[3]P. Roman and J.P. Leveille, J. Math. Phys. **15**, 2053 (1974).

[4]J.J. Aghassi, P. Roman, and R.M. Santilli, Phys. Rev. D **1**, 2753 (1970).

[5]See, for example: J.J. Aghassi, P. Roman, and R.M. Santilli, J. Math. Phys. **11**, 2297 (1970); J.J. Aghassi, P. Roman, and R.M. Santilli, Nuovo Cimento **5A**, 551 (1971); R.M. Santilli, Particles and Nuclei **1**, 81 (1970); P.L. Huddleston, M. Lorente, and P. Roman, Found. Phys. **5**, 75 (1975). See also L. Castell, Nuovo Cimento **46A**, 1 (1966), and **49A**, 285 (1967).

[6]H. Bacry and J.M. Lévy-Leblond, J. Math. Phys. **9**, 1605 (1968).

[7]The presence of the harmonic potential term has been shown to be a manifestation of the long-range effect of curvature.

[8]P. Roman and J. Haavisto, J. Math. Phys. **17**, 1664 (1976).

[9]P. Roman and J. Haavisto, Inter. J. Theor. Phys. **16**, 951 (1978).

[10]See, for example, E. Cremer and J. Sherk, Nucl. Phys. **B118**, 61 (1977); Y.M. Cho and P.G.O. Freund, Phys. Rev. D **12**, 1588 (1975); F.A. Bais and R.J. Russell, Phys. Rev. D **11**, 2642 (1975); P. Nath and R. Arnowitt, Phys. Rev. D **15**, 1033 (1977).

[11]D. Faiman and A.W. Hendry, Phys. Rev. **173**, 1720 (1968).

[12]The reason for this choice is that SO(3,2) [rather than SO(4,1)] is the symmetry of the microuniverse, cf. Ref. 9.

[13]See Appendix B of Ref. 8.

[14]That is, unitarily and globally implementable with generators constructed from the algebra of observables. See J.M. Jauch, Helv. Phys. Acta **37**, 284 (1964).

[15]In our realization, $T_{\mu\nu}$ is given simply by the $\Sigma_{\mu\nu}$ matrices in Eq. (2.18c).

[16]It is proportional to one of the Casimir invariants; see Section 3.1.

[17]One also has $J'_{\mu\nu} \equiv TJ_\mu T^{-1} = J_{\mu\nu}$.

[18]J. Haavisto, "Quantum-Dynamical Symmetry Groups in Curved Spaces," Ph.D. thesis, Boston University, 1978. The reader is warned this manuscript contains many misprints and minor errors.

[19]J. Rosen, Nuovo Cimento **35**, 698 (1965).

[20]In order to facilitate comparison with the literature and for historical reasons we have chosen a sign convention so that the $M_{\mu\nu}$ correspond to our $-J_{\mu\nu}$.

[21]The nomenclature we follow has been introduced by the authors of Ref. 6 in relation to SO(3,2) and SO(4,1) contractions.

[22]The Lorentz subalgebra (3.7a) remains, of course, always the same, and so do the $[M,P]$, $[M,Q]$, and $[M,S]$ brackets (3.7c), (3.7d), (3.7e).

[23]In the following we omit slashes on the generators and write P instead of P', etc.

[24]To compare with our previous formulas, take the circular frequency $\nu = 1$ in the latter. This choice of units could be avoided if we denoted g_{44} of SO(p,q) by $l^2\nu^2 g_{44}$.

[25]See the last two papers in Ref. 5.

[26]The change of signs in the P, Q, S subalgebra leads to replacing cos, sin by cosh, sinh in the realizations by differential operators of P and Q.

[27]In the "para" groups, the roles of P and Q are interchanged relative to the "regular" groups.

[28]J.G. Dubois, Nuovo Cimento **B15**, 1 (1972).

[29]Some further details can be found in Appendix D of Ref. 18.

[30]For the identification of the parameters θ,a etc. as belonging to the generators $l^{-1}I,P$, etc., Sec. 3.1.

[31]The pairs (η,p) label elements of the groups of characters $\hat{T}^{l'}_4$ = dual of $\tilde{T}^{l'}_4$.

[32]This SL(2,C) is generated *not* by the $J_{\mu\nu}$ but rather by the $T_{\mu\nu}$ as defined in (2.21).

[33]The upper labels on ψ are representation identification labels: l refers to $\mathscr{R}_0$, ρ designates the eigenvalue of $\mathscr{R}_3$; k and c together refer to the quantum numbers characterizing the SL(2,C)$^{l'}$ representations associated

with the Casimir invariants $\mathscr{R}_1$ and $\mathscr{R}_2$. The lower labels s,s_3 are state-labels, connected with the spintower level and the spin projection.

[34]See the second paper in Ref. 5.

[35]One also has to note that, according to the results of ref. 3, for the real world $l^{-1} = -\hbar$ and hence $l < 0$.

[36]Y. Nambu, Prog. Theor. Phys. Suppl. **1**, Nos. **37–38**, 368 (1966).

[37]We consider, naturally, a fixed $\mathscr{F}_5$ representation, where l and $\mathscr{R}_3$ are given.

[38]We take, as in the above discussions and as in the sequel, $\mathscr{R}_3 = 0$.

[39]These follow from (2.32) or from (2.18).

[40]R.P. Feynman, M. Kislinger, and F. Ravndal, Phys. Rev. D **3**, 2706 (1971).

[41]Y.S. Kim and M.E. Noz, Phys. Rev. D **15**, 335 (1977) and earlier works quoted therein.

[42]B contains the value of $\mathscr{R}_3$ relevant to the irreducible component of the particlar quasiparticle collective excitation.

Representations of the Poincaré group in a K_3-basis and the infinite momentum limit

M. Quirós
Instituto de Estructura de la Materia, Serrano, 119, Madrid-6, Spain
Departamento de Fisica, Universidad de Alcala, Alcala de Henares, Madrid, Spain

J. Ramirez Mittelbrunn [a)]
Max–Planck-Institut für Physik und Astrophysik, München, Federal Republic of Germany

(Received 22 April 1980; accepted for publication 7 August 1980)

The Poincaré group is represented taking as a complete set of commuting observables $\{\mathbf{P}_T, K_3, \mathscr{T}_3\}$, where K_3 is the boost along the third axis and $\mathscr{T}_3$ the third component of the null plane spin. We name it K_3-representation. There appears in it a parameter K, with dimensions of momentum, from which the infinite momentum limit can be implemented in a natural way as the contraction $k \to \infty$. K_3-states and wavefunctions are well defined in the infinite momentum limit. They are related to null plane states and wavefunctions by a Mellin transformation. The convergence properties of null plane functions translate into analyticity properties of K_3-functions in the complex λ (eigenvalue of K_3)-plane.

PACS numbers: 11.30.Cp, 03.65.Fd, 02.20.Qs, 02.20.Rt

I. INTRODUCTION

The present work is a first step towards a rigorous analysis of the group content of the limit $p_3 \to \infty$ in relativistic quantum mechanics. This limit has usually been handled in the literature through the infinite-momentum frame (IMF) construction,[1] where the Poincaré group is boosted along the third axis as

$$\hat{O} = \lim_{\omega \to \infty} e^{i\omega K_3} O e^{-i\omega K_3}. \tag{1}$$

This framework has been very fruitful in the description of high energy processes[2] but it is not entirely satisfactory from a theoretical point of view. Two main features of the transformation (1) support this assertion:

(i) As regards the Poincaré group, the transformation (1) is an inner automorphism, as has been pointed out by Bacry and Chang,[3] so that no mathematical simplification is obtained from it,

(ii) The IMF states are ill-defined because the transformation $\exp(i\omega K_3)$ is obviously singular in the limit $\omega \to \infty$.

The aim of this work will be to introduce an approach overcoming the two difficulties we have just mentioned.

It is usually thought that a kinematic approximation is described by an Inönü–Wigner–Mickelsson–Niederle[4,5] contraction of the Poincaré group and its unitary irreducible representations. As an example to illustrate this assertion let us note that the nonrelativistic kinematics, $|\mathbf{v}|/c$ small, and the covariance group in this region, the Galileo group, are obtained by the contraction $c \to \infty$ of the Poincaré group. In particular the rigorous link between relativistic and nonrelativistic wave functions, for free and interacting particles of arbitrary spin, has recently been established[6] by the present authors using the theory of contraction of Lie groups and its representations.[5] In this case the contracted group, the Galileo group, is simpler than the original group while the contracted states are the nonrelativistic states and, thus, well defined.

In general an Inönü–Wigner contraction[4] of a Lie group leads to a contracted group which is not isomorphic to the original group. On the other hand, representations of the original group can be contracted leading to well-defined contracted states.

Therefore, it is possible to overcome the two difficulties mentioned above by means of a contraction of the Poincaré group, and we shall keep in mind this philosophy in this work, so that the final scope of it will be to introduce a well-defined contraction procedure over the Poincaré group, and its irreducible representations, implementing in a natural manner the limit $P_3 \to \infty$ and leading to an ultrarelativistic covariance group and well-defined ultrarelativistic covariance group and well-defined ultrarelativistic states.[7]

Now, the consistent definition of a contraction procedure for representations requires us[5,6] to find a Hilbert space where both the Poincaré and the contracted group are simultaneously realized. In other words, we need to find a complete set of commuting observables (CSCO) remaining unaltered in the limit $P_3 \to \infty$. The search of this CSCO and the study of their properties as well as the representations of the Poincaré group in the corresponding basis of eigenstates is the main purpose of this paper.

In order to illustrate this point let us clarify why the canonical basis is not appropriate to our end. For massive particles, irreducible representations of the Poincaré group with $P^2 > 0$, the canonical basis is made up of states defined by

$$|\mathbf{p},\sigma\rangle = B_p |\mathring{p}_i \sigma\rangle, \tag{2}$$

where B_p is the pure Lorentz transformaiton connecting $\mathring{p} = (m,\mathbf{0})$ with $(\omega_i \mathbf{p})$. given explicitly by

$$B_p = \exp\left\{ith^{-1}\frac{|\mathbf{p}|}{\mathbf{P}_0}\frac{1}{|\mathbf{p}|}\mathbf{p}\cdot\mathbf{k}\right\}. \tag{3}$$

The canonical states (2) are common eigenstates of the CSCO

[a)] On leave of absence from Instituto de Estructura de la Materia, Serrano, 119, Madrid-6, Spain.

 0022-2488/81/020412-10$1.00

$\{\mathbf{P}, S_3(\mathbf{p})\}$, where $S_3(\mathbf{p})$ is the third component of the canonical spin

$$\mathbf{S}(\mathbf{p}) = B_p \mathbf{J} B_p^{-1}. \tag{4}$$

An arbitrary state $|\phi\rangle$ can be expanded in terms of the canonical basis (2) as

$$|\phi\rangle = \sum_\sigma \int \frac{d^3p}{\omega} \phi(\mathbf{p},\sigma)|\mathbf{p},\sigma\rangle, \tag{5}$$

in such a way that its projection over an element of the canonical basis gives the Wigner function

$$\phi(\mathbf{p},\sigma) = \langle \mathbf{p},\sigma|\phi\rangle. \tag{6}$$

However the space of Wigner functions $\phi(\mathbf{p},\sigma)$, that is the canonical basis $|\mathbf{p},\sigma\rangle$, is not appropriate to describe the high momentum limit because these functions, or states, are not well defined in the limit $p_3 \to \infty$.

Thus the problem to be solved is three-fold:

(i) To obtain a CSCO unaltered in the infinite momentum limit;

(ii) to obtain the common eigenstates of the CSCO and to project over them the state space;

(iii) to study the representation of the Poincaré group over the Hilbert space subtended by the new states.

In Sec. 2 we shall show that the suitable CSCO for spin zero representations is $\{P_i, K_3\}(i = 1,2)$ while for nonzero spin representations we shall choose to add, as fourth operator, the third component of the null plane spin $\mathcal{J}_3$. In the following of this introduction, we shall briefly review the main features of the null plane basis which will be often used throughout this paper.

In the null plane basis the Poincaré generators are divided into two classes[8]:

(a) The kinematical generators

$$E_1 = \tfrac{1}{2}(K_1 + J_2), \quad E_2 = \tfrac{1}{2}(K_2 - J_1),$$
$$P_+ = \tfrac{1}{2}(P_0 + P_3), \quad K_3, J_3, P_1, P_2 \tag{7a}$$

leaving invariant the null plane initial surface $x^- = \frac{1}{2}(x_0 - x^3) = 0$: generating the stability group of the null plane considered as initial surface;

(b) the dynamical generators

$$P_- = P_0 - P_3, \quad F_1 = K_1 - J_2, \quad F_2 = K_2 + J_1, \tag{7b}$$

which do not belong to the stability group (P_- translates the null plane while F_1 and F_2 rotate it around the light cone $x^2 = 0$), describing thus the dynamics of the system.

Basis vectors are defined as

$$|p_+, \mathbf{p}_T, \rho\rangle = K_p |\tfrac{1}{2}m, \mathbf{0}_T, \rho\rangle, \tag{8}$$

where K_p is the Kogut and Soper boost, defined as[9]

$$K_p = \exp\left\{ -i \frac{\mathbf{P}_T}{P_+} \mathbf{E}_T \right\} \exp\left\{ -i \ln \frac{2P_+}{m} K_3 \right\}. \tag{9}$$

The null plane states (8) are common eigenstates of the CSCO $\{P_+, P_1, P_2, \mathcal{J}_3\}$, where $\mathcal{J}_3$ is a Casimir operator of the stability group of the null plane, given by

$$\mathcal{J}_3 = J_3 + \frac{E_1 p_2 - E_2 p_1}{P_+}, \tag{10}$$

which is the third component of the null plane spin, defined by

$$\mathcal{J} = K_p \mathbf{J} K_p^{-1}. \tag{11}$$

The choice (8) for the basis vectors fixes the representation of the Poincaré generators (7) as

$$K_3 | p_+, \mathbf{p}_T, \rho\rangle = i p_+ \frac{\partial}{\partial p_+} \Big| p_+, \mathbf{p}_T, \rho \Big\rangle,$$

$$J_3 | p_+, \mathbf{p}_T, \rho\rangle = \left[i\left(p_1 \frac{\partial}{\partial p_2} - p_2 \frac{\partial}{\partial p_1} \right) + \mathcal{J}_3 \right] \Big| p_+, \mathbf{p}_T, \rho \Big\rangle,$$

$$P_+ | p_+, \mathbf{p}_T, \rho\rangle = p_+ | p_+, p_T, \rho\rangle;$$

$$P_- | p_+, p_T, \rho\rangle = \frac{m_T^2}{2p_+} \Big| p_+, \mathbf{p}_T, \rho \Big\rangle,$$

$$m_T^2 = p_T^2 + m^2, \tag{12}$$

$$P_i | p_+, \mathbf{p}_T, \rho\rangle = p_i | p_+, \mathbf{p}_T, \rho\rangle \quad (i = 1,2),$$

$$E_i | p_+, \mathbf{p}_T, \rho\rangle = i p_+ \frac{\partial}{\partial p_i} \Big| p_+, \mathbf{p}_T, \rho \Big\rangle,$$

$$F_i | p_+, \mathbf{p}_T, \rho\rangle = \left[i p_i \frac{\partial}{\partial p_+} + i \frac{m_T^2}{2p_+} \frac{\partial}{\partial p_i} - \epsilon_{ij} \left(\frac{p_i}{p_+} \mathcal{J}_3 + \frac{m}{p_+} \mathcal{J}_i \right) \right] \Big| p_+, p_T, \rho \Big\rangle,$$

where $\mathcal{J}$ are the $(2j+1)$-dimensional matrix representation of the SU(2) algebra acting on the index ρ.

It goes without saying that the null plane basis is not appropriate to perform the high momentum limit for the same reasons applied above to the canonical basis. However the prominent role played by the null plane basis lies in the fact, as we shall see later in this paper, that the appropriate—new—basis is related to it by a Mellin transform.

II. STATES AND WAVEFUNCTIONS IN THE K_3-basis

In this section we shall develop points (i) and (ii) of the program presented in Sec. 1.

First of all we shall obtain a CSCO which is not altered by the infinite momentum limit. We shall begin with the simple case of spinless particles whose physical degrees of freedom are described, in the null plane basis, by the CSCO $\{P_+, P_i\}$. Obviously, as we have mentioned in the introduction, this set of operators is not appropriate to describe the high-energy limit $p_+ \to \infty$.

Therefore, we must replace the operator p_+ by another one, commuting with p_i and having a finite, smooth limit as $p_+ \to \infty$. Then if we choose this operator belonging to the Poincaré algebra, we are led, in an unambiguous way, to the pure Lorentz transformation along the third axis K_3. In short, for spin zero representations the new CSCO we get is $\{K_3, P_i\}$.

In the more general, arbitrary spin, case we must complete the set $\{K_3, P_i\}$ with another operator, describing the spin degrees of freedom and commuting with $\{K_3, P_i\}$. The simplest choice[10] is to take, as fourth operator, the third component of the null plane spin $\mathcal{J}_3$, Eq. (10).

The common eigenstates of the CSCO, $\{K_3, \mathcal{J}_3, p_i\}$, will be denoted by $|\lambda_1 p_T, \rho\rangle$, the eigenvalue corresponding to the operator K_3 and, thus, satisfying the equation

$$K_3 |\lambda, \mathbf{p}_T, \rho\rangle = \lambda |\lambda, \mathbf{p}_T, \rho\rangle. \tag{13}$$

The basis of states $|\lambda, \mathbf{p}_T, \rho\rangle$ will be called, hereafter, the

K_3 basis.

The relationship between K_3 and null plane states can be shown, from the action of K_3 over the null plane states, Eq. (12), and the defining property of K_3-states, Eq. (13), as

$$\left|\lambda,\mathbf{p}_T,\rho\right\rangle = \frac{1}{\sqrt{2\pi}}\int_0^\infty \frac{dp_+}{p_+}\left(\frac{p_+}{\kappa}\right)^{-i\lambda}\left| p_+,\mathbf{p}_T,\rho\right\rangle, \tag{14a}$$

so that the K_3-basis is nothing else than the Mellin transformation[11] of the null plane basis. In Eq. (14) we have been led to introduce, for dimensional reasons, the constant K with dimensions of momentum. We shall see later that this dimensional constant, which appears in a natural way when we change the basis, is able to implement the limit $p_+\to\infty$, *through the redefinition* $P_+ = \kappa\eta$ and the limit $K\to\infty$. In this way K is only a scale for momenta so that it plays the same role as C did in the nonrelativistic limit which, we remember, was implemented through the limit $C\to\infty$. We shall come back to this point in the following section.

The inverse Mellin transformation of (14) is given explicitly by

$$\left| p_+,p_T,\rho\right\rangle = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} d\lambda\ (p_+/\kappa)^{+i\lambda}\left|\lambda,\mathbf{p}_T,\rho\right\rangle. \tag{14b}$$

Using the orthogonality of the null plane basis

$$\langle p_+,\mathbf{p}_T,\rho|p'_+,\mathbf{p}'_T,\rho'\rangle = p_+\delta(p_+-p'_+)\delta^{(2)}(\mathbf{p}_T-\mathbf{p}'_T)\delta_{\rho\rho'}, \tag{15a}$$

we get the orthogonality relations for the K_3-basis

$$\langle\lambda,\mathbf{p}_T,\rho|\lambda',\mathbf{p}'_T,\rho'\rangle = \delta(\lambda-\lambda')\delta^{(2)}(\mathbf{p}_T-\mathbf{p}'_T)\delta_{\rho\rho'}, \tag{15b}$$

while the projection of the K_3-basis over the null plane basis is given by

$$\langle p_+,\mathbf{p}_T,\rho|\lambda,\mathbf{p}'_T,\rho'\rangle = \frac{1}{\sqrt{2\pi}}\left(\frac{p_+}{\kappa}\right)^{-i\lambda}\delta^{(2)}(\mathbf{p}_T-\mathbf{p}'_T)\delta_{\rho\rho'}. \tag{16}$$

The completeness relations of the null plane basis

$$\mathbf{1} = \sum_\rho\int\frac{dp_+}{p_+}d^2p_T\ |\,p_+,\mathbf{p}_T,\rho\rangle\langle p_+,\mathbf{p}_T,\rho\,| \tag{17a}$$

translates into the completeness relation of the K_3-basis as

$$\mathbf{1} = \sum_\rho\int d\lambda\, d^2p_T\ |\lambda,\mathbf{p}_T,\rho\rangle\langle\lambda,\mathbf{p}_T,\rho\,|. \tag{17b}$$

We can alternatively use the language of wavefunctions. Let $\phi(p_+,\mathbf{p}_T,\sigma)$ be the null plane wavefunction, or projection of the state $|\phi\rangle$ over the null plane state $|p_+,p_T,\sigma\rangle$

$$\phi(p_+,\mathbf{p}_T,\sigma) = \langle p_+,\mathbf{p}_T,\sigma|\,\phi\,\rangle. \tag{18a}$$

We define the corresponding K_3-function as the projection

$$\phi(\lambda,\mathbf{p}_T,\sigma) = \langle\lambda,\mathbf{p}_T,\sigma|\,\phi\,\rangle. \tag{18b}$$

The relations between null plane and K_3-functions are obtained from

$$\phi(\lambda,\mathbf{p}_T,\rho) = \frac{1}{\sqrt{2\pi}}\int_0^\infty\frac{dp_+}{p_+}\left(\frac{p_+}{\kappa}\right)^{i\lambda}\phi(p_+,\mathbf{p}_T,\rho) \tag{19a}$$

and

$$\phi(p_+,\mathbf{p}_T,\rho) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} d\lambda\ (p_1/\kappa)^{-i\lambda}\phi(\lambda,\mathbf{p}_T,\rho). \tag{19b}$$

The scalar product over null plane functions

$$(\phi_1,\phi_2) = \sum_\rho\int_0^\infty\frac{dp_+}{p_+}\phi_1^*(p_+,\mathbf{p}_T,\sigma)\,\phi_2(p_+,\mathbf{p}_T,\sigma) \tag{20a}$$

translates into the following scalar product for K_3-functions:

$$(\phi_1,\phi_2) = \sum_\rho\int d\lambda\, d^2p_T\,\phi_1^*(\lambda,\mathbf{p}_T,\sigma)\,\phi_2(\lambda,\mathbf{p}_T,\sigma). \tag{20b}$$

Let us finally remark that convergence properties of null plane functions translate into analyticity properties of K_3-functions in the complex λ-plane. For instance, if the space of null plane functions is restricted to functions of finite norm, or square integrable functions,

$$\|\phi\|^2 = \int_0^\infty\frac{dp_+}{p_+}d^2p_T\sum_\sigma|\,\phi(p_+,\mathbf{p}_T,\sigma)|^2<\infty, \tag{21a}$$

as was implicitly supposed throughout this section, then the corresponding K_3-functions are analytic in the complex λ-plane, on an infinitesimal strip around the real axis, defined by (14)

$$\begin{aligned}&|\mathrm{Im}\lambda\,|<\epsilon,\quad(\epsilon>0),\\ &-\infty<\mathrm{Re}\lambda<\infty,\end{aligned} \tag{21b}$$

so that the inverse Mellin transformation (19b) makes sense when the integration contour C is taken along the real axis, as was initially assumed.

Stronger convergence properties, or a more restricted space of null plane states, translate into analyticity properties on wider strips for K_3-functions, as we shall see in the following section. In fact there is a one-to-one correspondence between convergence properties of $\phi(p_+,\mathbf{p}_T,\rho)$ and domains of analyticity, in the complex λ-plane, of $\phi(p_+,\mathbf{p}_T,\rho)$.

III. REPRESENTATIONS OF THE POINCARÉ GROUP IN THE K_3-BASIS

In this section we shall develop point (iii) of Sec. I and compute the representation of infinitesinal and finite elements of the Poincaré group over the K_3-basis.

A. Representation of algebra generators

The representation of the algebra generators over the K_3-functions, (18b), is easily obtained from the action of Poincaré generators over null plane states, (12), and the relations, (14) and (19), between K_3 and null plane states and functions.

Let us denote the action of the generator O over K_3-functions by the same symbol O. We get the following representation:

$$K_3\phi(\lambda,\mathbf{p}_T,\rho)=\lambda\phi(\lambda,\mathbf{p}_T,\rho),$$

$$J_3\phi(\lambda,\mathbf{p}_T,\rho)=\left[i\left(p_1\frac{\partial}{\partial p_2}-p_2\frac{\partial}{\partial p_1}\right)+\mathscr{T}_3\right]\phi(\lambda,\mathbf{p}_T,\rho),$$

$$P_+\phi(\lambda,\mathbf{p}_T,\rho)=K\phi(\lambda-i,\mathbf{p}_T,\rho),$$

$$P_i\phi(\lambda,\mathbf{p}_T,\rho)=p_i\phi(\lambda,\mathbf{p}_T,\rho),$$

$$P_-\phi(\lambda,\mathbf{p}_T,\rho)=(m_T^2/2K)\,\phi(\lambda+i,\mathbf{p}_T,\rho),$$

$$E_i\phi(\lambda,\mathbf{p}_T,\rho)=iK\frac{\partial}{\partial p_i}\phi(\lambda-i,\mathbf{p}_T,\rho),$$

$$F_i\phi(\lambda,\mathbf{p}_T,\rho)=\frac{1}{K}\left[(\lambda+i)+i\frac{m_T^2}{2}\frac{\partial}{\partial p_i}-\epsilon_{ij}(p_j\mathscr{T}_3+m\mathscr{T}_j)\right]\phi(\lambda+i,\mathbf{p}_T,\rho). \quad (22)$$

Thus in the representation (22) of the infinitesimal generators of the Poincaré group we can already see the two main virtues we had required for a representation to be appropriate to describe ultrarelativistic situations:

(i) In the K_3-representation, unlike in the canonical or null plane representations, there appears explicitly a parameter K making it possible to implement the limit $p_+\to\infty$ by means of the dimensional contraction $\kappa\to\infty$. In this way the K_3-representation plays, with respect to the infinite momentum limit, the same role that the canonical representation, where the parameter C appears explicitly,[6] plays in the nonrelativistic limit

(ii) We shall take the view that the K_3-functions do not depend on the parameter K, thus being well defined in the infinite momentum limit. Accordingly the canonical and null plane wavefunctions do depend on K (Eq. 19b).

B. Equations for matrix elements of group operators

In this section we shall compute the equations satisfied by the matrix elements, between K_3-states, of group elements $\exp(iaA)$ of the Poincaré group, where A is some generator of the Lie algebra and a the corresponding parameter. We shall denote these matrix elements by

$$\psi^{(A)}_{a,\lambda',\mathbf{p}'_T,\rho'}(\lambda,\mathbf{p}_T,\rho)\equiv\langle\lambda,\mathbf{p}_T\,\rho|e^{iaA}|\lambda',\mathbf{p}'_T,\rho'\rangle. \quad (23)$$

As we shall see, the fact that we are taking states diagonal in K_3 will enable us to compute the matrix elements (23). The boost K_3 plays a very singular role in the null plane decomposition of the Poincaré algebra. Any operator A obeying

$$[K_3,A]=i\gamma A \quad (24)$$

is referred to as an operator of goodness γ.[8] Thus the Poincaré generators can be classified according to (24): P_+,E_i are "good" generators ($\gamma=+1$); K_3,J_3,P_i are "bad" generators ($\gamma=0$) and p_-, F_i are "terrible" generators ($\gamma=1$).

Equation (24) can be generalized to

$$[K_3,A^n]=in\gamma A^n, \quad (25)$$

so that we get the following operatorial identity:

$$(K_3+\gamma aA)\exp(iaA)=\exp(iaA)K_3, \quad (26)$$

from which, by taking expectation values between K_3-states, we immediately obtain

$$(\lambda'-\lambda)\psi^{(A)}_{a,\lambda,p_t,\rho}(\lambda',\mathbf{p}'_T,\rho')=-\gamma a\langle\lambda',\mathbf{p}'_T,\rho'|A\exp(iaA)|\lambda,\mathbf{p}_T\,\rho\rangle. \quad (27)$$

The a-dependence of the function $\psi^{(A)}$ can be factorized, writing Eq. (27) as

$$\left\{(\lambda'-\lambda)-i\gamma a\frac{\partial}{\partial a}\right\}\psi^{(A)}_{a,\lambda,p_T,\rho}(\lambda',\mathbf{p}'_T,\rho')=0, \quad (28)$$

so that we can cast the general solution of (28) as

$$\psi^{(A)}_{a,\lambda,p_T,\rho}(\lambda',\mathbf{p}'_T,\rho')=a^{-i\gamma(\lambda'-\lambda)}\phi^{(A)}_{\lambda,p_T,\rho}(\lambda',\mathbf{p}'_T,\rho'). \quad (29)$$

The equation satisfied by $\phi^{(A)}$ depends on the particular representation of the generator A, given by (22), and the different cases will be studied in the following section.

A general property satisfied by the function $\psi^{(A)}$ is

$$\psi^{(A)*}_{a,\lambda,p_T,\rho}(\lambda',\mathbf{p}'_T,\rho')=\psi^{(A)}_{-a,\lambda,p_T,\rho}(\lambda',\mathbf{p}'_T,\rho'), \quad (30)$$

which comes from the very definition (23) and the Hermiticity of A. This property will be widely used.

C. Representation of finite group elements

The aim of this section is to compute the action of finite group elements $\exp(iaA)$ over K_3-functions $\phi(\lambda,p_T,\rho)$ as given by

$$[e^{iaA}\phi](\lambda,\mathbf{p}_T,\rho)=\langle\lambda,\mathbf{p}_T,\rho|e^{iaA}|\phi\rangle. \quad (31)$$

This action can be written in two ways:

(i) Series representation, obtained by expanding the exponential in a power series as

$$[e^{iaA}\phi](\lambda,\mathbf{p}_T,\rho)=\sum_{n=0}^{\infty}\frac{(ia)^n}{n!}A^n\phi(\lambda,\mathbf{p}_T,\rho), \quad (32a)$$

where the action of A^n over ϕ is known from the iteration of Eq. (22).

Let us remark that if we want the series (32a) to make sense we must require, over the space of functions, stronger conditions than those postulated in (21a) and (21b) for square integrable functions. In particular, if we restict ourselves to the subspace of square integrable functions such that the expectation value of p_+^n is finite,

$$\langle\phi|P_+^n|\phi\rangle<\infty \quad (n=0,1,2,\ldots), \quad (33a)$$

then the space of K_3-functions $\phi(\lambda,p_T,\rho)$ is restricted to analytic ones, in the complex λ-plane, on the finite strip defined by (15)

$$-\infty<\mathrm{Re}\lambda<\infty;\quad -n/2-\epsilon<\mathrm{Im}\lambda<\epsilon. \quad (33b)$$

Similarly, if the space of square integrable functions is restricted to functions with finite expectation value of p_+^{-n}

$$\langle\phi|P_+^{-n}|\phi\rangle<-\infty \quad (n=0,1,2,\ldots), \quad (34a)$$

then the K_3-functions must be analytic on the strip (15)

$$-\infty<\mathrm{Re}\lambda<\infty;\quad -\epsilon<\mathrm{Im}\lambda<n/2+\epsilon. \quad (34b)$$

(iii) Integral representation, obtained inserting the completeness relation (17b) into (3.1), as

$$[e^{iaA}\phi](\lambda,p_T,\rho)=\sum_{\rho'}\int d\lambda'\,d^2\mathbf{p}'_T\,\psi^{(A)}_{a,\lambda',\mathbf{p}'_T,\rho'}(\lambda,\mathbf{p}_T,\rho)\phi(\lambda',\mathbf{p}'_T,\rho'), \quad (32b)$$

where the function $\psi^{(A)}$ is given by (23). The integral representation is more general than the series representation because we only require the function ϕ to be square integrable, or analytic along the real axis, as in (21b).

It goes without saying that both representations must coincide in the cases where the series representation makes sense.

In the remainder of this section we shall compute the (general) integral representation (32b) for the finite elements of the Poincaré group and prove the coincidence with the series representation. Because finite elments are obtained by exponentiation of Lie algebra generators, $\exp(iaA)$ with $A\in\{P_\pm, J_3, K_3, P_i, E_i, F_i\}$, we can divide the calculation according to the different algebra generators. We shall proceed by grouping the generators which share similar features, goodness zero generators, $\{P_+, P_-\}$, $\{E_i\}$, and $\{F_i\}$.

1. Goodness zero generators

Goodness zero generators K_3, J_3, p_i, are represented over K_3-functions in a local way. Verification of the following equations is immediate:

$$\begin{aligned}
&[e^{ia_i P_i}\phi\,](\lambda,p_T,\rho) = e^{ia_i p_i}\phi(\lambda,p_T,\rho),\\
&[e^{i\alpha_3 J_3}\phi\,](\lambda,p_T,\rho) = e^{i\alpha_3[i(p_1\partial/\partial p_2 - p_2\partial/\partial p_1)+\rho]}\phi(\lambda,p_T,\rho)\\
&\qquad = e^{i\alpha_3\rho}\phi(\lambda,R^{-1}p_T,\rho), \qquad (35)\\
&[e^{i\beta_3 K_3}\phi\,](\lambda,p_T,\rho) = e^{i\beta_3\lambda}\phi(\lambda,p_T,\rho).
\end{aligned}$$

2. The generators P_+ and P_-

Using the general Eqs. (27) and (29) for $A = P_+$ together with the action of P_+ over the K_3-basis we get the following functional equation for $\phi^{(P_+)}$, (hereafter we shall omit the superscript (A) of the functions $\psi^{(A)}$ and $\phi^{(A)}$ when there is no danger of confusion):

$$(\lambda' - \lambda)\phi_{\lambda,P_+,\rho}(\lambda',\mathbf{p}'_T,\rho') + K\phi_{\lambda,p_T,\rho}(\lambda' - i)\mathbf{p}'_T,\rho') = 0. \qquad (36)$$

The solution of the functional equation (36) together with the general property (30) gives the function

$$\begin{aligned}
\psi_{a_+,\lambda,p_T,\rho}(\lambda',\mathbf{p}'_T,\rho') = &-\frac{1}{2\pi i}\Gamma(i(\lambda'-\lambda))\\
&\times(-ia_-\kappa)^{-i(\lambda'-\lambda)}\delta^{(2)}(\mathbf{p}_T - \mathbf{p}'_T)\delta_{\rho\rho'} \qquad (37)
\end{aligned}$$

and the integral representation

$$\begin{aligned}
[e^{iaP_+}\phi\,](\lambda,\mathbf{p}_T,\rho) = &-\frac{1}{2\pi i}\int_{-\infty}^{\infty} d\lambda'\,\Gamma(i(\lambda-\lambda'))\\
&\times(-ia_-\kappa)^{i(\lambda'-\lambda)}\phi(\lambda',\mathbf{p}_T,\rho). \qquad (38a)
\end{aligned}$$

The series representation $\Sigma^{(ia)^{(n/n!)}}p_+^n$ acting over K_3-functions makes sense only for states such that $\langle p_+^n\rangle < \infty$, for any $n\geqslant 0$, so that the functions $\phi(\lambda,p_T,\rho)$ are restricted to be analytic in the lower half-plane $\mathrm{Im}\lambda < \epsilon$ in agreement with (33b). Under these conditions of analyticity we can write

$$[e^{ia_+P_+}\phi\,](\lambda,\mathbf{p}_T,\rho) = \sum_{n=0}^{\infty}\frac{(ia_-\kappa)^n}{n!}\phi(\lambda - in,\mathbf{p}_T,\rho). \qquad (38b)$$

The integrand of (38a) has a singularity at $\lambda' = \lambda$, due to the Γ function, so that the contour of integration in (38a) must be understood as going from $-\infty$ to $+\infty$ along the real axis, but avoiding the pole at $\lambda' = \lambda$ by means of a small semicircle, of radius $\delta < \epsilon$, in the upper half-plane. This distortion is possible because the function of (38a), $\phi(\lambda',\mathbf{p}_T,\rho)$, is supposed to be analytic on the strip $|\mathrm{Im}\lambda'| < \epsilon$. Furthermore, if the function $\phi(\lambda',\mathbf{p}_T,\rho)$ is also analytic on the lower half-plane, $\mathrm{Im}\lambda' < \epsilon$, the only singularities of the integrand of (38a) are the poles of $\Gamma(i(\lambda-\lambda'))$, at $\lambda' = \lambda - in$ $(n = 0,1,2,\ldots)$ with residues $(-1)^n/n!$, and we may displace the contour of integration downwards, to "minus infinity", and parallel to the real axis. The integral over the displaced contour will vanish while a series of contributions is obtained from the poles of the Γ function which has been crossed over. This series reproduces exactly the series representation (38b), as was required. In fact, the normalization factor $-1/2\pi i$ was chosen in (37) so as to cancel the factor $-2\pi i$ from the Cauchy theorem.

A straightforward application of the methods just described enables us to compute $\exp\{ia_+p_-\}\,\phi$ with some modifications which will be pointed out. The series representation

$$[e^{ia_+P}\,\phi\,](\lambda,\mathbf{p}_T,\rho) = \sum_{n=0}^{\infty}\left[\frac{ia_+m_T^2}{2\kappa}\right]^n\frac{1}{n!}\phi(\lambda + in,\mathbf{p}_T,\rho) \qquad (39a)$$

requires the function $\phi(\lambda',\mathbf{p}_T,\rho)$ to be analytic in the upper half-plane $\mathrm{Im}\lambda > -\epsilon$, in agreement with (34b).

The integral representation can be written as

$$\begin{aligned}
[e^{ia_+P}\,\phi\,](\lambda,\mathbf{p}_T,\rho) = &\frac{1}{2\pi i}\int_{-\infty}^{\infty} d\lambda'\,\Gamma(i(\lambda'-\lambda))\\
&\times\left(\frac{-a_+m_T^2}{2\kappa}\right)^{i(\lambda-\lambda')}\phi(\lambda',\mathbf{p}_T,\rho), \qquad (39b)
\end{aligned}$$

where the contour of integration avoids the pole $\lambda' = \lambda$ of the Γ function by a small semicircle in the lower half-plane. Again, if ϕ is analytic in the upper half-plane, we may displace the contour of integration upwards, and the series of contributions from the poles of $\Gamma(i(\lambda'-\lambda))$ at $\lambda' = \lambda + in$ which are crossed over reproduces the series representation (39a).

3. The generators F_i

Using the action of the generators E_i over K_3-functions, Eq. (22), the series representation for $\exp\{i\alpha_i E_i\}\,\phi$ can be evaluated as

$$[e^{i\alpha_i E_i}\phi\,](\lambda,\mathbf{p}_T,\rho) = \sum_{n=0}^{\infty}\frac{(i\alpha_i)^n}{n!}\left(ik\frac{\partial}{\partial p_i}\right)^n\phi(\lambda - ni,\mathbf{p}_T,\rho), \qquad (40a)$$

where the function ϕ is supposed to be analytic in the lower half-plane $\mathrm{Im}\lambda < \epsilon$.

In order to compute the general integral representation (32b) let us first evaluate the matrix elements (23), satisfying Eqs. (27)–(29). Using (12) we can see that the function ϕ, Eq. (29), satisfies the following differentiofunctional equation:

$$(\lambda' - \lambda)\phi_{\lambda,p_T,\rho}(\lambda',p'_T,\rho') + i\kappa\frac{\partial}{\partial p_i}\phi_{\lambda,p_T,\rho}(\lambda' - i,\mathbf{p}'_T,\rho'), \qquad (41)$$

whose solution depends on a parameter b with dimensions of (momentum)$^{-1}$ as

$$\Gamma(i(\lambda'-\lambda))(-i\kappa b)^{-i(\lambda'-\lambda)}e^{ib(\mathbf{p}'_j - p_j)}\delta(p'_j - p_j)\delta_{\rho\rho'}. \qquad (42)$$

In (42) we use the convention $j = 2$ for $i = 1$, and $j = 1$ for

$i=2$, and the general condition (30) has already been taken into account. Any function (42) with any value of b, is a solution of (41) so that the general solution of (41) must be given by a linear combination of functions (42). The physical significance of b is the impact parameter conjugate to the transverse momentum p_i. Because we are representing the Poincaré group over functions $\phi(\lambda,\mathbf{p}_T,\rho)$ of well-defined transverse momentum, the action of the group cannot depend on the impact parameter, by the Heisenberg uncertainty principle, and we must integrate the function (42) over b. On the other hand, the measure of integration db restores the dimensionality, (momentum)$^{-2}$, which the functions $\phi_{\lambda p_T \rho}(\lambda' \mathbf{p}'_T \rho')$ must exhibit, as can be seen from (15b).

In this way the matrix element function can be written as

$$\psi_{\alpha i,\lambda,p_T,\rho}(\lambda',\mathbf{p}'_T,\rho') = -\frac{1}{4\pi^2 i}\Gamma(i(\lambda'-\lambda))(-i\alpha_i)^{-i(\lambda'-\lambda)} \times\delta(p'_j-p_j)\delta_{\rho\rho'}I(\lambda'-\lambda), \tag{43}$$

where the function I is given by the Fourier transform

$$I(\lambda'-\lambda)=\int_{-\infty}^{\infty} db\,(\kappa b)^{-i(\lambda'-\lambda)}e^{ib(p'_i-p_i)}, \tag{44}$$

and the integral (44) can be evaluated as follows,[12] see Eq. (A27):

$$I=\frac{i}{\kappa}e^{n/2(\lambda'-\lambda)}\Gamma(-i(\lambda'-\lambda)+1) \times\left\{\left(\frac{p'_i-p_i}{\kappa}+io\right)^{i(\lambda'-\lambda)-1}-\left(\frac{p'_i-p_i}{\kappa}-io\right)^{i(\lambda'-\lambda)-1}\right\} \tag{45a}$$

for $-i(\lambda'-\lambda)\neq -n(n=1,2,...)$, while

$$I=i^n\frac{\pi}{(n-1)!}\operatorname{sign}\left(\frac{p'_i-p_i}{\kappa}\right)\left(\frac{p'_i-p_i}{\kappa}\right)^{n-1} \tag{45b}$$

for $-i(\lambda'-\lambda)=-n$. *The functions* $x\pm io$ must be understood in the sense of generalized functions.[12]

Thus we can explicitly cast the integral representation of $\exp\{i\alpha_i E_i\}$ into the following form:

$$[e^{i\alpha_i E_i}\phi\,](\lambda,p_T,\rho)=\int d\lambda'\,dp'_i\,\psi_{\alpha_i,\lambda',p'_i}(\lambda,p_i)\,\phi(\lambda',p'_i,p_j,\rho), \tag{40b}$$

where

$$\psi_{\alpha_i,\lambda',p'_i}(\lambda,p'_i)=-\tfrac{1}{2}(-i\alpha_i\kappa)^{i(\lambda'-\lambda)}e^{(\pi/2)(\lambda-\lambda')} \times\frac{1}{\sin\pi[i(\lambda-\lambda')]} \times\{(p_i-p'_i+io)^{i(\lambda-\lambda')-1} -(p_i-p'_i-io)^{i(\lambda-\lambda')-1}\} \tag{46a}$$

if $i(\lambda'-\lambda)\neq -1,2,...$, and

$$\psi_{\alpha_i,\lambda+in,p'_i}(\lambda,p_i)=\frac{1}{(2\pi)^2}i(\alpha_i)^{-n}\left(\frac{p_i-p'_i}{\kappa}\right)^{n-1} \tag{46b}$$

for $-i(\lambda'-\lambda)=-1,-2,...$.

Let us remark, from (46), that the function $\psi_{\alpha_i,\lambda',p'_i}(\lambda,p_i)$ has poles at $\lambda'=\lambda-in, n=0,1,2,...$. The contour of integration of λ' in (40b) must be distorted so as to avoid the singularity of the integrand at $\lambda'=\lambda$.

Using the following property[12] of generalized functions:

$$(x+io)^{-n-1}-(x-io)^{-n-1}=(-2\pi i)((-1)^n/n!)\delta^{(n)}(x), \tag{47}$$

the residue of the function (46), at the location of the poles, can be extracted as

$$\operatorname{Res}\psi_{\alpha_i,\lambda',p'_i}(\lambda,p_i)=-\frac{1}{2\pi i}\frac{(\alpha_i-\kappa)^n}{n!}\delta^{(n)}(p'_i-p_i). \tag{48}$$

Hence, if the function ϕ is analytic in the lower half-plane, the contour of integration may be displaced downward. The only singularities of the integrand of (40b) are the poles of (46a), whose residues, Eq. (48), must be added when the poles are crossed over. The series of residues of (40b) reproduces the series representation (40a). The proof is straightforward.

4. THE GENERATORS F_i

In this section we will merely quote the results concerning the series representation and the integral representation of the elements $\exp\{i\beta_i F_i\}$ of the Poincaré group, acting over K_3-functions. The computational details as well as the equivalence between the representations will be relegated to the Appendix.

First of all we can put the generators F_i into the following form:

$$F_i=e^{i\theta_i\mathscr{T}_3}\left(\frac{1}{p_+}O_i\right)e^{-i\theta_i\mathscr{T}_3}\quad(i=1,2), \tag{49}$$

where

$$O_i=p_iK_3+i\frac{m_T^2}{2}\frac{\partial}{\partial p_i}-\epsilon_{ij}(p_j^2+m^2)^{1/2}\mathscr{T}_3 \tag{50}$$

and

$$\theta_1=\arctan(m/p_2);\quad \theta_2=\arctan(m/p_1). \tag{51}$$

Using the decomposition (49), the series representation over functions $\phi(\lambda,\mathbf{p}_T,\rho)$ analytic in the upper half-plane, $\operatorname{Im}\lambda>-\epsilon$, can be written as

$$[e^{i\beta_i F_i}\phi\,](\lambda,\mathbf{p}_T,\rho)=\sum_{n=0}^{\infty}\frac{(i\beta_i)^n}{n!}\kappa^{-n} \times\sum_{\sigma,\rho'}T^{(j)}_{\epsilon\sigma}(\delta_{i2}\tfrac{\pi}{2},\theta_i(p_T),-\delta_{i2}\tfrac{\pi}{2}) \times(m_T^2)^n\left\{\frac{\lambda+ni}{m_T^2}p_i+\frac{i}{2}\frac{\partial}{\partial p_i}-\frac{\epsilon_{ij}(p_j^2+m^2)^{1/2}}{m_T^2}\sigma\right\}^n \times T^{(j)-1}_{\sigma\rho'}\left(\delta_{i2}\frac{\pi}{2},\theta_i(p_T),-\delta_{i2}\frac{\pi}{2}\right) \tag{52a}$$

$$\phi(\lambda+ni,p_T,\rho') =\sum_{n=0}^{\infty}\frac{(i\beta_i)^n}{n!}K^{-n}\sum_{\sigma,\rho'}T^{(j)}_{\rho\sigma}\left(\delta_{i2}\frac{\pi}{2},\theta_i,-\delta_{i2}\frac{\pi}{2}\right) \times\left\{\frac{\lambda}{m_T^2}p_i+\frac{i}{2}\frac{\partial}{\partial P_i}-\frac{\epsilon_{ij}(p_j^2+m^2)^{1/2}}{m_T^2}\sigma\right\}^n(m_T^2)^n \times T^{(j)-1}_{\sigma\rho'}\left(\delta_{i2}\frac{\pi}{2},\theta_i,-\delta_{i2}\frac{\pi}{2}\right)\phi(\lambda+n_i,p_T,\rho'). \tag{52b}$$

The mathematical steps leading from (49) to (52) can be found in the Appendix, Eqs. (A3)–(A13). In (52) the functions $T^{(j)}_{mn}$ are matrix elements for the spin j irreducible representation of the rotations group, parametrized as[13]

$$R(\psi,\theta,\phi) = e^{i\psi J_3} e^{i\theta J_1} e^{i\phi J_3}. \tag{53}$$

We may express the explicit integral representation of $\exp\{i\beta_i F_i\}, (i = 1,2)$, as

$$[e^{i\beta_i F_i}\phi\,](\lambda, \mathbf{p}_T, \rho) = \sum_{\rho'} \int d\lambda'\, dp_i'\, \psi_{\beta,\lambda' p_i' \rho'}(\lambda, \mathbf{p}_T, \rho)\, \phi(\lambda', p_i', p_j', \rho'), \tag{54}$$

where the function $\psi_{\beta,\lambda' p_i' \rho'}(\lambda, \mathbf{p}_T, \rho)$ is given by

$$\psi_{\beta,\lambda' p_i' \rho'}(\lambda, p_T, \rho) = \frac{1}{4\pi^2 i} \sum_{\sigma} T^{(j)}_{\rho\sigma}\left(\delta_{i2}\frac{\pi}{2}, \theta_i, -\delta_{i2}\frac{\pi}{2}\right) \times T^{-1(j)}_{\sigma\rho'}\left(\delta_{i2}\frac{\pi}{2}, \theta_i, -\delta_{i2}\frac{\pi}{2}\right) e^{i\lambda \ln(m_i^2/K^2) - i\lambda' \ln(m_i'^2/K^2)} \times e^{2i\sigma[\arctan p_i/\epsilon_i(p_i + m^2)^{1/2} - \arctan p_i'/\epsilon_i(\mathbf{p}_i^2 + m^2)^{1/2}]} \times (-i\beta_i)^{-i(\lambda' - \lambda)} \Gamma(i(\lambda' - \lambda))\, I(\lambda', \lambda) \tag{55}$$

$$I = \left(\frac{2}{\kappa}\right)^{i(\lambda' - \lambda) - 1} \frac{i}{\kappa} \Gamma(-i(\lambda' - \lambda) + 1) \times \{(p_i' - p_i + i\varepsilon)^{i(\lambda' - \lambda) - 1} - (p_i' - p_i - i\varepsilon)^{i(\lambda' - \lambda) - 1}\} e^{\pi/2(\lambda' - \lambda)} \tag{56a}$$

for $i(\lambda' - \lambda) \neq n (n = 1,2,...)$; while for $i(\lambda' - \lambda) = n$ we have

$$I(\lambda - in) = (2/\kappa)^{n-1} \frac{i^n \pi}{(n-1)!} \operatorname{sign}\left(\frac{p_i' - p_i}{\kappa}\right)(p_i' - p_i)^{n-1}. \tag{56b}$$

The calculations leading to the integral representation (54) are somewhat lengthy and the interested reader can find them in the Appendix, Eqs. (A14)–(A29). Let us note that the function ψ defined in Eq. (55) has singularities at $\lambda' = \lambda + in$ so that the integration contour in λ' in (54) must be distorted by a small semicircle in order to avoid the pole at $\lambda' = \lambda$. If the function ϕ is supposed to be analytic in the upper half-plane, the integration contour in λ' may be displaced upward, and the series of poles at $\lambda' = \lambda + in$ crossed over in (54) reproduces the series representation (52), as can again be found in the Appendix, Eqs. (A30)–(A32).

IV. CONCLUSION

The aim of this paper was to obtain a representation of the Poincaré group—or equivalently a basis—where the infinite momentum limit could be implemented in a natural, well-defined, fashion. It is usually thought that the null plane basis, whose kinematical algebra leaves invariant the null plane $x^0 = x^3$, is the best framework to describe high-energy situations and it has been successfully applied to parton models of high energy hadrons.[1,2] Nevertheless there are some "bad" features of the null plane basis to which we wish to call attention: (a) The absence of a parameter "measuring" high momenta (this role is played by C in the nonrelativistic limit). (b) As a consequence, the only way of implementing the infinite momentum limit is through an inner automorphism, or infinite momentum frame, so that, as pointed out by Bacry and Chang,[3] the contracted group is isomorphic to the Poincaré group, and no mathematical simplification is obtained. (c) The null-plane states are ill-defined in the infinite momentum limit, and a complete description of infinite momentum states would need to include those with infinite mass[3] (gigamomenta).

In this paper we have proposed the following CSCO: $\{\mathbf{p}_T, K_3, \mathcal{T}_3\}$, where $\mathcal{T}_3$ is the third component of the null plane spin. Let us briefly review the main features of the representation of the Poincaré group over the new basis, called the K_3-basis:

a) The spin operator is the same as in the null plane basis so that the kinematical algebra is spin independent and the spin degrees of freedom appear only in the dynamical part of the group.

b) K_3-states are well defined in the infinite momentum limit. In fact they are given as Mellin transforms of the null plane states.

c) In the representation of the Poincaré algebra, and the Poincaré group, there appears a parameter K, with dimensions of momentum, which enables the infinite momentum limit to be implemented as the limit $k \to \infty$ in a contraction procedure.

d) The convergence properties of null plane wave functions translate into (more transparent) analyticity properties of the K_3-functions in the complex λ (eigenvalue of K_3-plane.

Once we have the contraction parameter K, we can contract the Poincaré algebra and group by means of a suitable redefinition of the infintesimal generators and the limit $k \to \infty$. Two main contractions can be performed:

a) Let γ be the goodness of the operator O, and define

$$\tilde{O} = \kappa^{-\gamma} O. \tag{57}$$

This contraction is equivalent to the infinite momentum frame contraction of Bacry–Chang[3] and Kogut–Soper,[9] and the contracted group is isomorphic to the Poincaré group. With contraction (57) we will have the same features showed by the infinite momentum frame limit (as e.g., Galilean motion in the transverse plane, etc...) but the states will now be well defined.

b) The natural contraction dictated by the K_3-representation [see Eqs. (22)] would be given by the definitions

$$\tilde{E}_i = (1/\kappa)E_i, \quad \hat{P}_+ = (1/\kappa)P_+, \tag{58}$$

and the remaining generators unaltered. In this contraction the stability group of the null plane keeps the same structure, with rescaled variables, while the Hamiltonians F_1, F_2, P_- of the dynamic group go to zero, in the limit $\kappa \to \infty$, in the case of free particles. However, in the interacting case the dynamical algebra will no longer tend to zero, as it must be modified by interaction terms, and the description of the interaction could be simpler in the contracted, ultrarelativistic group, than in the original Poincaré group. This situation recalls a similar one in the nonrelativistic limit where the dynamics of a spin 1/2 particle is simply described by the nonrelativistic Pauli equation which is the limit $C \to \infty$ of the Dirac equation. The problem of the interaction, as well as such other related problems as local covariant realizations, equations of motion, and the position operator, are under consideration.

ACKNOWLEDGMENTS

One of the authors (M.Q.) wishes to thank Professor I.A. de Azcárraga for useful discussions about the null-plane basis. The other (J.R.M.) is grateful to Professors H.P. Dürr and W. Drechsler for the warm hospitality at the Max-Planck-Institut for Physik und Astrophysik, and the Max-Planck Institution for economical support.

APPENDIX

In this Appendix we shall compute the series and the integral representation of $\exp\{iaF_1\}\phi$, corresponding to Eqs. (49)–(56) of Sec. III. The representation of $\exp\{iaF_2\}$ is a straightforward application of the methods contained in this Appendix.

Let us first put F_1, as in Eq. (49)

$$F_1 = e^{i\theta_1\mathscr{T}_1}\left(\frac{1}{p_+}O_1\right)e^{-i\theta_1\mathscr{T}_1}, \tag{A1}$$

with

$$O_1 = P_1K_3 + i\frac{m_T^2}{2}\frac{\partial}{\partial p_1} - (p_2^2+m^2)\mathscr{T}_3\,. \tag{A2}$$

A. Series representation

In the series expansion of $\exp\{iaF_1\}$ there will appear $(1/P_+O_1)^n$ so that we will need the following two Propositions:

Proposition 1:

$$\left(\frac{1}{P_+}O_1\right)^n = \left(\frac{1}{P_+}\right)^n(m_T^2)^n\left(\frac{1}{m_T^2}O_1\right)^n, \tag{A3}$$

Proposition 2:

$$\left(\frac{1}{P_+}O_1\right)^n = \left(\frac{1}{m_T^2}O_1\right)^n\left(\frac{1}{P_+}\right)^n\left(\frac{1}{P_+}\right)^n. \tag{A4}$$

To prove these propositions, we will need the following:

Lemma:

$$\left(\frac{1}{m_T^2}O_1\right)^n\frac{1}{P_+} = \frac{1}{P_+}\left(O_1\frac{1}{m_T^2}\right)^n, \tag{A5}$$

Proof of the Lemma:

The lemma can be proven by induction over n:

1. That the lemma is true for $n=1$, that is

$$\frac{1}{m_T^2}O_1\frac{1}{P_+} = \frac{1}{P_+}O_1\frac{1}{m_T^2} \tag{A6}$$

follows from the explicit representation (A2) of O_1 and the commutation relation

$$\left[K_3,\frac{1}{P_+}\right] = -i\frac{1}{P_+}\,. \tag{A7}$$

2. Let us suppose (A5) true for n, then

$$\left(\frac{1}{m_T^2}O_1\right)^{n+1}\frac{1}{P_+} = \frac{1}{m_T^2}O_1\frac{1}{P_+}\left(O_1\frac{1}{m_T^2}\right)^n \tag{A8}$$

and using (A6) we get

$$\left(\frac{1}{m_T^2}O_1\right)^{n+1}\frac{1}{P_+} = \frac{1}{P_+}\left(O_1\frac{1}{m_T^2}\right)^{n+1} \tag{A9}$$

which proves Lemma (A5).

Proof of Proposition 1:

We shall again proceed by induction:

1. Equation (A3), for $n=1$, is trivial;

2. Supposing (A3) true for n, then

$$\left(\frac{1}{m_T^2}O_1\right)^{n+1} = \left(\frac{1}{P_+}\right)^n(m_T^2)^n\left(\frac{1}{m_T^2}O_1\right)^n\frac{1}{P_+}O_1 \tag{A10}$$

but by virtue of the Lemma (A5)

$$\left(\frac{1}{P_+}O_1\right)^{n+1} = \left(\frac{1}{P_+}\right)^n(m_T^2)^n\frac{1}{P_+}\left(O_1\frac{1}{m_T^2}\right)^nO_1 \tag{A11}$$

From where the Eq. (A5), for $n+1$, follows, proving thus Proposition 1.

The proof of Proposition 2 follows along the same lines as Proposition 1, using the result of Lemma (A5).

Now, using the decomposition (A1) and definition (31) we can cast the series representation of $\exp(iaF_1)$ as

$$[e^{iaF_1}\phi\,](\lambda,\mathbf{p}_T,\rho) = \sum_{n=0}^{\infty}\frac{(ia)^n}{n!}\sum_{\sigma,\sigma',\rho'}\int d\lambda'\,d^2\mathbf{p}_T'\,t^{(j)}_{\rho\sigma}(\theta_i(p_T'))\tag{A12}$$

$$\times\left\langle\lambda,\mathbf{p}_T,\sigma\left|\left(\frac{1}{P_+}O_1\right)^n\right|\lambda',\mathbf{p}_T',\sigma'\right\rangle t^{-1(j)}_{\sigma'\rho'}(\theta_1(p_T'))\,\phi(\lambda',\mathbf{p}_T',\rho')$$

where, by definition, $t^{(j)}(\theta) = T^{(j)}(O,\theta,O)$. It is easy to prove, using Proposition 1, that

$$\left\langle\lambda,\mathbf{p}_T,\sigma\left|\left(\frac{1}{P_+}O_1\right)^n\right|\lambda',\mathbf{p}_T',\sigma'\right\rangle$$

$$= (m_T^2)^nK^{-n}\left\{\frac{\lambda' p_1}{m_T^2}+\frac{i}{2}\frac{\partial}{\partial p_1}-\frac{(p_2^2+m^2)^{1/2}}{m_T^2}\sigma'\right\}^n$$

$$\times\delta(\lambda+ni-\lambda')\delta^{(2)}(\mathbf{p}_T-\mathbf{p}_T')\delta_{\sigma\sigma'}\,, \tag{A13a}$$

while the use of Proposition 2 gives

$$\left\langle\lambda,\mathbf{p}_T,\sigma\left|\left(\frac{1}{P_+}O_1\right)^n\right|\lambda',\mathbf{p}_T',\sigma'\right\rangle$$

$$= K^{-n}\left\{\frac{\lambda p_1}{m_T^2}+\frac{i}{2}\frac{\partial}{\partial p_1}-\frac{(p_2^2+m^2)^{1/2}}{m_T^2}\sigma\right\}^n$$

$$\times(m_T^2)^n\delta(\lambda+nk-\lambda')\delta^{(2)}(\mathbf{p}_T-\mathbf{p}_T')\delta_{\sigma\sigma'}\,. \tag{A13b}$$

Thus the insertion of (A13a) into (A12) proves the series representation (52a) while the insertion of (A13b) proves (52b). The equivalence between (52a) and (52b) can be proven directly by induction over n.

B. Integral representation

Using the action (22) of the generator F_1 over the K_3-states, the equation (27) for the matrix element $\psi^{(F_1)}$ leads to the following equation for the function ϕ, as defined in (29),

$$(\lambda'-\lambda)\phi_{\lambda,p_T,\rho}(\lambda',\mathbf{p}_T',\rho')$$

$$= \sum_{\sigma'\sigma}t^{(j)}_{\rho'\sigma'}(\theta_1(p_T'))K^{-1}\left\{(\lambda'+i)p_1'+i\frac{m_T^{12}}{2}\frac{\partial}{\partial p_1'}\right.$$

$$\left.-(p_2'^2+m^2)^{1/2}\sigma'\right\}$$

$$\times t^{-1(j)}_{\sigma'\sigma}(\theta_1(p_T'))\phi_{\lambda,p_T,\rho}(\lambda'+i,\mathbf{p}_T',\sigma)\,. \tag{A14}$$

Our task will be now to compute a solution to the differentio-functional equation (A14). This will be done by successive simplifications of the equation, by means of a series of *ansätze*.

Ansatz 1:

$$\phi_{\lambda,p_T,\rho}(\lambda',\mathbf{p}'_T,\rho') = \sum_{\mu} t^{(j)}_{\sigma\mu}(\theta(p'_T))\xi_{\lambda p_T\rho}(\lambda',\mathbf{p}'_T,\mu)\,. \tag{A15}$$

The substitution of (A15) into (A14) gives the following equation for the function ξ:

$$(\lambda'-\lambda)\xi_{\lambda,p_T,\rho}(\lambda',\mathbf{p}'_T,\mu) = K^{-1}\left\{(\lambda'+i)p'_1 + i\frac{m'^2_T}{2}\frac{\partial}{\partial p'_1} - (p'^2_2+m^2)^{1/2}\mu\right\}\xi_{\lambda,p_T,\rho}(\lambda'+i,\mathbf{p}'_T,\mu)\,. \tag{A16}$$

Ansatz 2:

$$\xi_{\lambda,p_T,\rho}(\lambda',\mathbf{p}'_T,\mu) = f(\lambda',\mathbf{p}'_T,\mu;b')\chi_{\lambda',p_T,\rho}(\lambda',\mathbf{p}'_T,b')\,, \tag{A17}$$

where f satisfies the eigenvalue equation

$$\left[\lambda' p'_1 + i\frac{m'^2_T}{2}\frac{\partial}{\partial p'_1} - (p'^2_2+m^2)^{1/2}\mu\right]f(\lambda',\mathbf{p}'_T,b';\mu) = k^2 b' f(\lambda',\mathbf{p}'_T,b';\mu)\,, \tag{A18}$$

whose solution can be explicitly computed as

$$f(\lambda',\mathbf{p}'_T,b';\mu) = \exp\left\{i\lambda'\ln\frac{m'^2_T}{\kappa^2} - 2i\frac{(\kappa^2 b' + \mu(p'^2_2+m^2)^{1/2})}{(p'^2_2+m^2)^{1/2}}\arctan\frac{p'_1}{(p'^2_2+m^2)^{1/2}}\right\}. \tag{A19}$$

Then, the equation satisfied by the function χ is

$$(\lambda'-\lambda)\chi_{\lambda,p_T,\rho}(\lambda',\mathbf{p}'_T,b') = \left(\frac{\kappa^2}{m'^2_T}\kappa b' + i\frac{\kappa}{2}\frac{\partial}{\partial p'_1}\right)\chi_{\lambda,p_T,\rho}(\lambda'+i,\mathbf{p}'_T;b')\,. \tag{A20}$$

Ansatz 3:

$$\chi_{\lambda,p_T,\rho}(\lambda',p'_T,b') = g(p'_T;b,b')\mathscr{H}(\lambda,\lambda')\mathscr{F}(\lambda,p_T,\rho)\,, \tag{A21}$$

where g is the solution of the following eigenvalue equation

$$\left\{\frac{\kappa^2}{m'^2_T}\kappa b' + i\frac{\kappa^2}{2\kappa}\frac{\partial}{\partial p'_1}\right\}g(p'_T;b,b') = \kappa b g(p'_T;b,b') \tag{A22}$$

or, explicitly,

$$g(p'_T;b,b') = e^{\frac{\kappa^2}{(p'^2_2+m^2)^{1/2}}\arctan\frac{p'_1}{(p'^2_2+m^2)^{1/2}}}\,. \tag{A23}$$

The substitution of (A21) into (A20) gives, for $\mathscr{H}(\lambda,\lambda')$, the simple functional equation

$$(\lambda'-\lambda)\mathscr{H}(\lambda,\lambda') = \kappa b\mathscr{H}(\lambda,\lambda'+1)\,, \tag{A24}$$

whose solution is $\mathscr{H}(\lambda,\lambda') = \Gamma(-i(\lambda'-\lambda))(-i\kappa b)^{i(\lambda'-\lambda)}$. (A25)

Using (A15)–(A25) we get

$$\phi_{\lambda,p_T,\rho'}(\lambda',\mathbf{p}'_T,\rho') = (i\kappa b)^{i(\lambda'-\lambda)}\Gamma(-i(\lambda'-\lambda))\sum_{\sigma} t^{(j)}_{\rho'\sigma}(\theta(p'_T))$$

$$\times\exp\left\{i\lambda'\ln\frac{m'^2_T}{\kappa^2} 2ibp'_1 - 2i\sigma\arctan\frac{p'_1}{(p'^2_2+m^2)^{1/2}}\right\}\delta(p_2-p'_2)\mathscr{F}(\mathbf{p}_T,\lambda,\rho)\,, \tag{A26}$$

which is independent on the parameter b'. Nevertheless the function in Eq. (A26) is explicitly dependent on b. Because the matrix element $\psi_{a,\lambda,p_T,\rho}(\lambda',\mathbf{p}'_T,\rho')$ may not be dependent on b, we must integrate over b in (A26), as in Sec. II.C.3. The integration can be performed with the aid of the Fourier transform[12]

$$\mathscr{F}[x^{\lambda}] = ie^{i\frac{1}{2}\pi\lambda}\Gamma(\lambda+1)\{(\sigma+io)^{-\lambda-1} - (\sigma-io)^{-\lambda-1}\} \quad \lambda\neq -1,-2,..., \tag{A27a}$$

$$\mathscr{F}[x^{-n}] = i^n\frac{\pi}{(n-1)!}\operatorname{sgn}\sigma(\sigma^{n-1})\,. \tag{A27b}$$

The function $\mathscr{F}(\mathbf{p}_T,\lambda,\rho)$ must be fixed with the aid of the general property (30), giving thus

$$\mathscr{F} = t^{-1(j)}_{\sigma\rho}(\theta,(p_T))\exp\left\{-i\lambda\ln\frac{m^2_T}{\kappa^2} + 2ibp_1 + 2i\sigma\arctan\frac{p_1}{(p^2_2+m^2)^{1/2}}\right\}. \tag{A28}$$

Using (A27)–(A29) we can give for the matrix element, the following expression, for $i(\lambda'-\lambda)\neq -1,-2,...$:

$$\psi_{a,\lambda,p_T,\rho}(\lambda',p'_T,\rho') = \frac{e^{(\pi/2)i(\lambda-\lambda')}}{4\pi}\left(\frac{-ia\kappa}{2}\right)^{i(\lambda'-\lambda)}\frac{1}{\sin\pi[i(\lambda'-\lambda)]}\,e^{i\lambda'\ln(m_T'^2/\kappa^2)-i\lambda\ln(m_T^2/\kappa^2)}$$

$$\times\sum_\sigma t^{(j)}_{\rho'\sigma}(\theta_1(p'_T))e^{-2i\sigma[\arctan(p'_1/(p_2'^2+m^2)^{1/2})-\arctan(p_1/(p_2^2+m^2)^{1/2})]}t^{-1(j)}_{\sigma\rho}(\theta_1(p_T))$$

$$\times\{(p_1-p'_1+io)^{i(\lambda-\lambda'-1}-(p_1-p'_1-io)^{i(\lambda-\lambda')-1}\}\delta(p_2-p'_2), \tag{A29a}$$

while

$$\psi_{a,\lambda,p_T,\rho}(\lambda+in,\mathbf{p}'_T,\rho') = \frac{1}{4\pi i}\left(\frac{-am_T'^2}{2\kappa^2}\right)^{-n}e^{i\lambda\ln(m_T'^2/m_T^2)(p_1-p'_1/\kappa)^{n-1}}$$

$$\times\sum_\sigma t^{(j)}_{\rho'\sigma}(\theta_1(p'_T))e^{-2i\sigma[\arctan(p'_1/(p_2'^2+m^2)^{1/2})-\arctan(p_1/(p_2^2+m^2)^{1/2})]}t^{-1}_{\sigma\rho}(\theta_1(p_T))\delta_{(p_2p'_2)}\,, \tag{A29b}$$

for $i(\lambda'-\lambda)=-1,-2,\ldots$.

Now, using (32b), the integral representation (54) follows easily. Let us note that the function $\psi_{a,\lambda,p_T,\rho}(\lambda',p'_T,\rho')$ has singularities at $\lambda'=\lambda-in$ in the lower half-plane of the complex variable λ', due to the function $1/\sin\pi[i(\lambda-\lambda')]$, but is regular in the upper half-plane, see (A29b).

C. Equivalence of series and integral representation

Whenever the function ϕ is analytic in the upper half-plane, the integral representation (54) leads unambiguously to the series representation (52). The simplest way of proving this is to write (54) as an integral over b and to exchange the order of integration: i.e. to integrate first in λ' and then in b. The residues of the poles at $\lambda'=\lambda+in, n=0,1,2,\ldots$, are added as usual, so that (54) can be written as

$$[e^{iaF_1}\phi\,](\lambda,\mathbf{p}_T,\rho) = \frac{1}{\pi}\sum_{\rho'}\int d\rho'_1\,db\sum_{n=0}^{\infty}\frac{(ia)^n}{n!}\kappa^{-n}b^n e^{-2ib(p_1-p'_1)}$$

$$\times(m_T'^2)^n\sum_\sigma t^{(j)}_{\rho\sigma}(\theta_1(p_T))e^{-2i\sigma\arctan p_1/(p_2^2+m^2)^{1/2}-\arctan p'_1/(p_2^2+m^2)^{1/2}}$$

$$\times e^{i\lambda[\ln(m_T'^2/\kappa^2)-\ln(m_T^2/\kappa^2)]}t^{-1(j)}_{\sigma\rho'}(\theta_1(p_T))\,. \tag{A30}$$

The following property

$$\left\{\left(\frac{1}{m_T^2}O_1\right)^n-b^n\right\}e^{-2ibp_1}e^{i\lambda\ln(m_T^2/\kappa^2)}e^{-2i\sigma\arctan[p_1/(p_2^2+m^2)^{1/2}]}=0\,, \tag{A31}$$

which can be proved straightforwardly, and the integral over the remaining b-dependent function

$$\int db\,e^{-2ib(p_1-p'_1)}=\pi\delta(p_1-p'_1) \tag{A32}$$

leads to the series representation (52b).

[1]For a review of IMF techniques, see: I.B. Kogut and L. Susskind, Phys. Rep. **8** C, 75 (1973).
[2]I.B. Kogut and D.E. Soper, Phys. Rev. D **1**, 2901 (1970); I.D. Bjorken, I.B. Kogut, and E.E. Soper, Phys. Rev. D **3**, 1382 (1971).
[3]H. Bacry and N.P. Chang, Ann. Phys. (N.Y.) **47**, 407 (1968).
[4]E. Inönü and E.P. Wigner, Proc. Natl. Acad. Sci. **39**, 510 (1953).
[5]I. Mickelsson and J. Niederle, Comm. Math. Phys. **27**, 167 (1972).
[6]J. León, M. Quirós, and J. Ramirez Mittelbrunn, Nuovo Cimento **46** B, 109 (1978); J. Math. Phys. **20**, 1068 (1979); Nuovo Cimento **54** B, 153 (1979).
[7]An earlier attempt along this direction can be found in J. León, M. Quirós, and J. Ramirez Mittelbrunn, Nuovo Cimento **41** A, 141 (1977); Phys. Lett. **68** B, 247 (1977).
[8]H. Leutwyler and I. Stern, Ann. Phys. **112**, 94 (1978) and references contained therein.
[9]D.E. Soper, Phys. Rev. D **5**, 1956 (1972); Ph.D. thesis, Stanford University, SLAC Report No. 137, 1971 (unpublished).
[10]Other possibilities have been explored in: M. Quirós and J. Ramirez Mittelbrunn, Nuovo Cimento Lett. **26**, 633 (1979).
[11]Bateman Manuscript Project, *Table of Integral Transforms*, edited by A. Erdélyi (McGraw-Hill, New York, 1953).
[12]I.M. Gel'fand and G.E. Shilov, *Generalized Functions*, Vol. I (Academic, New York, 1964).
[13]N.I. Vilenkin, *Special Functions and the Theory of Group Representations*, (American Mathematical Society, Providence, RI, 1968).
[14]To be precise, this is only true if the functions $\phi(p_+)$ satisfy a slightly stronger condition than (21a), namely that they should decrease at $p_+=\infty$ at least as $p_+^{-\epsilon}$ and vanish at $p_+=0$ at least as p_+^{ϵ} for some $\epsilon>0$. For example, a behavior of $\phi(p_+)$ like $1/\log p_+$ is enough for (21a) at ∞ but it is not sufficient to guarantee the existence of the analyticity band (21b).
[15]Similar considerations as in (14) apply also here.

Affine transformations and the geometry of superspace

R. W. Tucker
Department of Physics, University of Lancaster, Lancaster, England

(Received 1 August 1979; accepted for publication 1 February 1980)

A graded Cartan-type connection is devised on a bundle of graded affine frames over superspace. The relation of the gauged graded affine group to the geometry of superspace is discussed in the context of bundle reduction to simulate spontaneous symmetry breakdown. A complex quaternionic calculus is used to simplify the algebraic analysis.

PACS numbers: 11.30.Pb, 12.25. + e

I. INTRODUCTION

Certain physical theories have often received clarification and a lucid formulation in terms of geometric concepts. Such concepts attempt to focus on the salient features of the theory in an intrinsic coordinate independent manner.[1] In particular, classical field theory finds an economical description in the modern language of fiber bundles.

In this article some aspects of the theory of simple supergravity are formulated in terms of the geometry of a reduced manifold.[2] A distinction will be made between those aspects of the formulation that belong essentially to the establishment of the structure group of the appropriate bundle and those features that result when a particular choice is made for a connection in that bundle.[3] Such a choice may for example be made by finding the extremum of an action form on the base manifold which is invariant under change of section in the bundle under consideration. The language of the fiber bundle is particularly suited to the geometrical description of spontaneous symmetry breaking. Indeed in a recent paper K. Stelle and P. West have argued that the reduction of an SO(3,2) bundle over space–time to an O(3,1) bundle enables them to formulate a theory of gravity as a spontaneously broken SO(3,2) gauge theory. This approach is motivated by a particular action which generates the equations for the geometrical fields in the theory. Although the establishment of an invariant affine connection on a homogeneous space may often be thought of as being triggered by some spontaneous symmetry mechanism, this aspect of the formulation will not be dwelt upon in this paper.

In the following, a fairly well established procedure[2] for reducing one bundle to another will be carried out in an attempt to draw together the notions of the gauged graded Poincaré group (graded affine group) and the transformations of simple supergravity.

Two essential viewpoints will be adopted. Firstly it is asserted that this (and any) formulation should ultimately be expressible in a coordinate independent way. Thus at any point the (passive) choice of a coordinate chart in whatever manifold is being considered should have no intrinsic significance. Secondly, a geometrical theory of gravity will be said to be defined once the linear connection form and canonical form of a suitable bundle of anholonomic frames is specified. If the space under consideration admits a metric of definite signature then orthonormal frames (with respect to this metric) will be chosen and the connection becomes a metric one. Thus the conventional theory of Einstein is specified in terms of a torsion free connection on OM, the bundle of orthonormal frames over space–time M with the structure group O(3,1) that preserves the Minkowski metric with signature $(-+++)$. This interpretation of a gravitational theory is partly motivated by the need to establish a rule for transporting tensors (as well as frames) in space since in conventional theories they will describe the other (matter) fields of physics.

Theories involving SL(2,C) spinors are usually presented with the spinor components forming vector arrays that are coupled into invariant SL(2, C) combinations with suitable γ matrices. In order to formulate the theory of manifolds with spinorial coordinates it has proved convenient to break with this tradition and embed the spinor components into the ring of complex quaternions. Historically quaternions have often been used to simplify calculations involving rotations in three or four dimensions. The complex quaternions form a homomorphic image of the SL(2, C) algebra and enable the SL(2, C) group to act in a succint manner. The reader is referred to Hestene's[5] book on space–time algebra for an analogous viewpoint and earlier papers on quaternions. An essentially self-contained description of spinors in a complex quaternionic basis that is designed for application to superspace tensor analysis is presented in the Appendix. Since SL(2, C) plays a fundamental role in the following discussion it is often convenient to embed other entities into the complex quaternionic algebra. A considerable freedom from Fierz rearrangement is afforded by pursuing these techniques and tedious algebra in the ring of γ matrices is replaced by operations with the simpler quaternionic algebra. Complex quaternionic valued differential p forms will be used extensively in the analysis. These and other tensors will be graded in general by grading both their components and their basis elements. The components will often be graded as elements of some unspecified (hidden) Grassmann algebra. When necessary the terms odd or even will refer to the Fermi–Bose grading.

In order to establish notations a brief description will be presented for a geometry of ordinary space–time M (with points coordinated by x^m) in terms of a complex quaternionic vector valued linear connection form $\hat{\omega}$ on the bundle $OM(M, \mathrm{SL}(2, \mathrm{C}))$ of orthonormal frames. An element of OM is an orthonormal ordered basis of the tangent space $T_x(M)$ at x. Writing an element of SL(2, C) as $Q(\hat{\lambda}) = e^{\hat{\lambda}}$, where $\hat{\lambda} = \hat{\alpha} + i\hat{\beta}$ the right action on the tangent frame $\bar{\partial} = \partial_a . \tilde{e}^a$

 0022-2488/81/020422-08$1.00

is defined by

$$\tilde{\partial}\to\overline{Q}^{\dagger}\tilde{\partial}\overline{Q}=\partial_a.\overline{Q}^{\dagger}\tilde{e}^{a}\overline{Q}\equiv\partial_a.e(\hat{\lambda})=\partial_a.X^{a}_{b}\,\tilde{e}^{b},$$

i.e.,

$$\tilde{e}^{a}\xrightarrow{Q}X^{a}_{b}\,\tilde{e}^{b}. \tag{1.1}$$

Thus the linear frame $u\equiv(\tilde{\partial})=(\partial_1,\partial_2,\partial_3,\partial_0)$ may be considered as a map from the antihermitian quaternions H_q into $T_x(M)$ with the Lorentz group action given from above:

$$uQ{:}H_q\xrightarrow{Q}H_q\xrightarrow{u}TM;\tilde{e}^{a}\to X^{a}_{b}\,\tilde{e}^{b}\to u(X^{a}_{b}\,\tilde{e}^{b})=\partial_b.X^{b}_{a}. \tag{1.2}$$

The bundle of orthonormal frames OM is made into a differentiable manifold by taking the ten numbers (x^m,α^j,β^j) $(j=1,2,3)$ as local coordinates. The canonical form $\mathbf{e}'$ will be taken as an antihermitian quaternionic 1 form $\in T_u{}^*(OM)$ satisfying

$$u\mathbf{e}'(\mathbf{X}^*)=\pi\mathbf{X}^*,$$

where $\mathbf{X}^*\in T_u(OM)$ and $\pi{:}OM\to M$ is the projection. Locally a vector at u with coordinates (X^a,ξ^k), $k=1,\ldots,6$, may be written[6]

$$\mathbf{X}^*=\sum_{a=0}^{4}\partial_a.X^{a}+\sum_{j=1}^{3}\left(\frac{\partial}{\partial\beta^{j}}.\xi^{j}+\frac{\partial}{\partial\alpha^{j}}.\xi^{j+3}\right). \tag{1.3}$$

One readily verifies that in terms of the antihermitian coframe

$$\mathbf{e}=\sum_{a=1}^{4}\mathbf{e}^{a}\tilde{e}_{a},\quad\text{where}\quad\mathbf{e}^{a}(\partial_b)=\delta^{a}_{b},$$

$$\mathbf{e}'=Q\mathbf{e}Q^{\dagger}, \tag{1.4}$$

and $\mathbf{e}'$ corresponds simply to an SL(2, C) rotated coframe. In order to establish a connection 1-form a set of fundamental vector fields $\mathbf{J}_i{}^*$ is first defined on the bundle. These must correspond homomorphically to the six SL(2, C) generators $\hat{J}_i\equiv(\hat{e}_1,\hat{e}_2,\hat{e}_3,i\hat{e}_1,i\hat{e}_2,i\hat{e}_3)$. They may be chosen to be dual to the complex q-vector valued Maurer–Cartan 1-form $\hat{\omega}_{mc}\equiv Qd\overline{Q}$. Thus writing

$$\hat{\omega}_{mc}=\sum_{n=1}^{6}\omega^{n}_{mc}\,\hat{J}_n$$

$$J_i{}^*=\sum_{j=1}^{3}\left[\frac{\partial}{\partial\beta^{j}}.J^{j}_{i}(\alpha,\beta)+\frac{\partial}{\partial\alpha^{j}}.J^{j+3}_{i}(\alpha,\beta)\right],\quad i=1,\ldots,6,$$

one solves for J^{j}_{i} the equation

$$\hat{\omega}_{mc}(J_i{}^*)=\hat{J}_i$$

or

$$\omega^{n}_{mc}(J_i{}^*)=\delta^{n}_{i},\quad i,n=1,\ldots,6, \tag{1.5}$$

where

$$\omega^{n}_{mc}=\sum_{j=1}^{3}\left[\omega^{n}_{j}(\alpha,\beta)\,d\beta^{j}+\omega^{n}_{j+3}(\alpha,\beta)\,d\alpha^{j}\right].$$

The $\mathbf{J}_n{}^*$ span the vertical space tangent to the fibers and any vertical vector $\mathbf{A}^*\in T_n(OM)$ can be expressed as

$$\mathbf{A}^*=\sum_{i=1}^{6}A^{i}(x,\alpha,\beta)\,J_i{}^*$$

Any complex q-vector valued 1-form$\in T^*{}_u(OM)$ of the type $\hat{\omega}=\hat{\omega}_m\,dx^m+\hat{\omega}_{mc}$ satisfies the fundamental condition

$\hat{\omega}(\mathbf{A}^*)=\hat{A}\equiv\Sigma_{i=1}^{6}A^{i}\hat{J}_i$. The rest of $\hat{\omega}$ is defined so that $\hat{\omega}(\mathbf{X})=\hat{X}\equiv v\hat{X}$ for any $\mathbf{X}$ (not just vertical vector) and this partitions $T_u(OM)$ so that $\mathbf{X}=h\,\mathbf{X}+v\mathbf{X}$, where $\hat{\omega}(h\,\mathbf{X})=0$. From these definitions the horizontal component may be expressed in terms of a connection as

$$h\,\mathbf{X}=\mathbf{X}-\sum_{i=1}^{6}\omega^{i}(\mathbf{X})\,J_i{}^*, \tag{1.6}$$

where $\hat{\omega}=\Sigma_{i=1}^{6}\omega^{i}\hat{J}^{i}$. A convenient expression for an antihermitian q valued torsion 1-form can now be computed in term of a section through OM. The bundle torsion is defined in terms of $\mathbf{e}'$ as

$$\mathbf{T}=d\mathbf{e}'.h, \tag{17}$$

i.e., has components $\mathbf{T}(\partial_a,\partial_b)=d\mathbf{e}'(h\,\partial_a,h\,\partial_b)$ in any basis. From the definition (7) and the identities $dQ(h\partial_m)=\hat{\omega}(\partial_m)Q$, $dQ^{+}(h\partial_m)=Q^{+}\hat{\omega}(\partial_m)$ one finds that

$$\mathbf{T}=d\mathbf{e}+2\mathscr{A}(\hat{\omega}\wedge\mathbf{e}) \tag{1.8}$$

(Strictly speaking this is $\sigma^*\mathbf{T}$ expressed in terms of $\sigma^*\hat{\omega}$, where $\sigma{:}\ M\to OM$ defines the section. Sectioned forms will be implied on the following and the pull back symbol will not be explicitly mentioned.) Under a change of section generated by the SL(2, C) element $Q(\hat{\alpha}(x),\hat{\beta}(x))$

$$\hat{\omega}\to Q\hat{\omega}\overline{Q}+Qd\overline{Q} \tag{1.9}$$

$$\mathbf{T}\to Q\,\mathbf{T}Q^{\dagger}. \tag{1.10}$$

In terms of an exterior covariant derivative D_ω the torsion can be written

$$\mathbf{T}=D_\omega\mathbf{e}=-\mathbf{T}^{\dagger}, \tag{1.11}$$

where the subscript on D denotes the connection under consideration. The curvature of this connection is the complex q-vector valued 2-form $\hat{\mathbf{R}}$ defined by

$$\hat{\mathbf{R}}=d\hat{\omega}+\hat{\omega}\wedge\hat{\omega}\to Q\hat{\mathbf{R}}\overline{Q}, \tag{1.12}$$

with the usual $2\times3\times6=36$ real components. (The quaternions commute with the exterior multiplication.) Using definitions 8 and 12 and the condition $Q\overline{Q}=1$ the Bianchi identities follow immediately as

$$D_\omega\hat{\mathbf{R}}\equiv d\hat{\mathbf{R}}+2V(\hat{\omega}\wedge\hat{\mathbf{R}})=0, \tag{1.13}$$

$$D_\omega\mathbf{T}\equiv d\,\mathbf{T}+2\mathscr{A}(\hat{\omega}\wedge\mathbf{T})=2\mathscr{A}(\hat{\mathbf{R}}\wedge\mathbf{e}). \tag{1.14}$$

Spinors may be regarded as certain complex quaternionic valued tensors in a bundle associated with the orthonormal frame bundle. For a given SL(2, C) frame rotation

$$\mathbf{e}\to Q\mathbf{e}Q^{\dagger}, \tag{1.15}$$

the four types of spinors defined in the Appendix transform as

$$\phi_{\pm}\to Q\phi_{\pm}, \tag{1.16}$$

$$\dot{\phi}_{\pm}\to\dot{\phi}_{\pm}\,Q^{\dagger}. \tag{1.17}$$

As an example of the applicability of quaternionic forms in space-time the theory of simple supergravity[7] is recast into this language. Denoting the gravitino by an odd Majorana spinor valued 1-form $\boldsymbol{\chi}$ with $\boldsymbol{\chi}^{\dagger}=-i\boldsymbol{\chi}$ the action 4-form is

$$\Lambda(\mathbf{e},\hat{\omega},\boldsymbol{\chi})=\mathrm{Im}S\{k\hat{\mathbf{R}}\wedge\mathbf{e}\wedge\mathbf{e}^*-i\bar{\mathbf{e}}\wedge[\boldsymbol{\chi}\wedge D_\omega\dot{\boldsymbol{\chi}}+D_\omega\boldsymbol{\chi}\wedge\dot{\boldsymbol{\chi}}]\}, \tag{1.18}$$

where

$$D_\omega \chi = d\chi + \hat{\omega} \wedge \chi \to Q D_\omega \chi ,$$
$$D_\omega \dot{\chi} = d\dot{\chi} - \dot{\chi} \wedge \hat{\omega}^\dagger \to D_\omega \dot{\chi} Q^\dagger .$$

Making intrinsic variations in $\mathbf{e}$, $\dot{\chi}$, and $\hat{\omega}$, respectively, gives the equations:

$$\hat{\mathbf{R}} \wedge \mathbf{e} = (i/k) D_\omega \chi \wedge \dot{\chi}, \tag{1.19}$$
$$\bar{\mathbf{e}} \wedge D_\omega \chi = 0, \tag{1.20}$$
$$\mathbf{T} = (i/k)\, \chi \wedge \dot{\chi} , \tag{1.21}$$

where the last equation is used to fix the connection in terms of $\mathbf{e}$ and χ. Using the identities $\bar{\chi} \wedge \chi = \chi \wedge \bar{\chi} = 0$ the above action is readily verified to be invariant under the following variations with an odd spinor 0-form parameter ϵ:

$$\delta \mathbf{e} = 2\mathscr{R}(i\epsilon\dot{\chi}) , \tag{1.22}$$
$$\delta \chi = k D_\omega \epsilon , \tag{1.23}$$
$$\delta \dot{\chi} = k D_\omega \dot{\epsilon} . \tag{1.24}$$

These supersymmetry transformations will be returned to in the last section from the viewpoint of particular superspace diffeomorphisms.

2. GRADED AFFINE GROUP

Having established the notion of quaternionic valued forms and $OM(M, \mathrm{SL}(2,\mathrm{C})$ in the familiar context of simple supergravity this mode of description will be used to discuss the graded affine frame bundle $AN(N,G)$ over a supermanifold N. In addition to the SL(2,C) generator $\hat{\lambda}$ the other elements $(P_0, P_a, Q_\alpha, Q_{\dot{\alpha}})$ of the graded affine algebra are embedded into the complex quaternionic ring in the following way

$$P = -iP_0 - P_a \hat{e}_a , \tag{2.1}$$
$$S_\theta = S(\theta \bar{Q}_+) , \tag{2.2}$$
$$S_{\dot{\gamma}} = S(\dot{\gamma} \dot{Q}_+) . \tag{2.3}$$

In terms of these even elements the algebra is defined by

$$\mathscr{L}_{\hat{\lambda}_1} \hat{\lambda}_2 \equiv [\hat{\lambda}_1, \hat{\lambda}_2] ,$$
$$\mathscr{L}_{\hat{\lambda}} S(Pb) = S(P(\hat{\lambda} b + b\hat{\lambda})) ,$$
$$\mathscr{L}_{\hat{\lambda}} S_\beta = S_{\hat{\lambda}\beta} ,$$
$$\mathscr{L}_{\hat{\lambda}} S_{\dot{\beta}} = S_{\dot{\beta}\hat{\lambda}^\dagger} ,$$
$$\mathscr{L}_{S_\alpha} S_\beta = 2S(P\alpha\beta) . \tag{2.4}$$

For Majorana related spinorial elements the translation parameter b may be taken antihermitian. An affine group element will be parametrized as

$$\dot{G} = e^{S_\theta} e^{S_{\dot{\theta}}} e^{2SPb}\, e^{\hat{\lambda}} \equiv (\bar{Q}(\hat{\lambda}), b, \theta, \dot{\theta}) , \tag{2.5}$$

where $\bar{Q}(\hat{\lambda})$ is an abbreviation for the six coordinates $\hat{\lambda} = \hat{\alpha} + i\hat{\beta}$. Regarding the group manifold as being coordinated by $(Q,b,\theta,\dot{\theta})$ it may be verified from (2.4) that the group action is

$$G \times G \to G: (\bar{Q}_1, b_1, \alpha, \dot{\alpha})(\bar{Q}_2, b_2, \beta, \dot{\beta}) = (\bar{Q}_1 \bar{Q}_2, b_1 + \bar{Q}_1 b_2 \bar{Q}_1^\dagger + 2\alpha\beta\bar{Q}_1^\dagger , \alpha + \bar{Q}_1 \beta, \dot{\alpha} + \dot{\beta}\bar{Q}_1^\dagger) . \tag{2.6}$$

(The virtue of working with quaternionic parameters is particularly apparent in this composition rule.) It is also straightforward to verify that a left invariant 1-form in this parameterization is given by

$$\Omega_{mc} \equiv G^{-1} dG = Q d\bar{Q} + 2S(P(Q[db - \theta d\dot{\theta}] Q^\dagger)) + S_{d\dot{\theta} Q^\dagger} + S_{Qd\theta} . \tag{2.7}$$

As indicated in the first section such a form plays an important role in constructing a connection form with values in the graded affine Lie algebra. In particular it may be used to derive the corresponding fundamental vector fields on the group manifold.

Defining the generalized exterior product of the algebra valued even forms as

$$\alpha \wedge \beta \equiv \mathscr{L}_\alpha \wedge \beta = \sum_{i,j} \alpha_i \wedge \beta_j [J^i, J^j] , \tag{2.8}$$

where the general algebra is generated by J^i one may verify the Maurer–Cartan equation for (2.7):

$$d\Omega_{mc} = -\tfrac{1}{2} \Omega_{mc} \wedge \Omega_{mc} \tag{2.9}$$

which also completely specifies the graded Lie algebra.

3. GRADED AFFINE FRAME BUNDLE

Considerable work has been done recently in making the notion of a superspace precise.[8] For the purpose under discussion only the existence of a local chart with coordinate functions $Z^M = (x^m, z^\mu, z^{\dot{\mu}})$ will be required where the established conventions for the nature of the eight coordinates will be adopted. [The x^m may in fact be taken as the coordinates of a point in M which forms the base of an exterior vector bundle generated by the four elements $(z^\mu, z^{\dot{\mu}})$, i.e.,

$$N = \bigcup_x \sum_{k=0}^{4} \oplus \Lambda^k_x(V) , \tag{3.1}$$

where $\Lambda^r_x(V)$ is the space of rth exterior powers of V at x and V is the vector space spanned by $(z^\mu, z^{\dot{\mu}})$. For superfield applications this space is converted into a suitable module.]

Associated with each coordinate, directional derivatives $\partial/\partial z^M$ $M \equiv \partial_M = (\partial_m , \partial_\mu , \partial_{\dot{\mu}})$ are established together with their duals $(dx^m, dz^\mu, dz^{\dot{\mu}})$. These constitute a coordinate basis in $T_z(N)$ and $T^*_z(N)$, respectively. Writing $\mathbf{\Delta}_B = \partial_N \cdot E^N_B$ and $\mathbf{E}^A = E^A_M\, dZ^M$ quaternionic frames and coframes are defined (see Appendix)

$$\bar{\mathbf{\Delta}} = \bar{\partial} \oplus \bar{\pi}_+ \oplus \bar{\pi}_+ = \mathbf{\Delta}_A \tilde{e}^A , \tag{3.2}$$
$$\mathbf{E} = \mathbf{e} \oplus \epsilon \oplus \dot{\epsilon} = \mathbf{E}^A \tilde{e}_A . \tag{3.3}$$

Since all the elements of any bundle over superspace are graded, great care is needed to establish consistent ordering conventions. As different authors use different conventions in this respect it is useful to note at this point some of the differences that arise in graded differential geometry.

It is convenient to adopt a generalized Dirac notation. Thus a general (r,s) type supertensor field will be expressed in the previous basis as

$$\mathbf{K} = (\mathbf{\Delta}_{A_1} \otimes \mathbf{\Delta}_{A_2} \otimes \cdots \otimes \mathbf{\Delta}_{A_r}) K^{A_1 \cdots A_r}_{B_1 \cdots B_s} (\mathbf{E}^{B_1} \otimes \cdots \otimes \mathbf{E}^{B_s}) \tag{3.4}$$

This can be embedded naturally into the quaternionic frame

expansion by defining the quaternion components

$$\tilde{K} = K^{A_1\cdots A_r}_{B_1\cdots B_s}\,\tilde{e}_{A_1}\otimes\cdots\tilde{e}_{A_r}\otimes\tilde{e}^{B_1}\otimes\cdots\otimes\tilde{e}^{B_s} \tag{3.5}$$

and writing

$$\mathbf{K} = S(\bar{\mathbf{\Delta}}\otimes\cdots\otimes\bar{\mathbf{\Delta}}\check{K}\,\mathbf{E}\otimes\cdots\otimes\mathbf{E}), \tag{3.6}$$

where S here acts tensorwise in each space. Covariant tensor are regarded as evaluating vectors from the left. Thus

$$(\mathbf{E}^{A_1},\mathbf{E}^{A_2},\ldots,\mathbf{E}^{A_r})\,\mathbf{K}(\mathbf{\Delta}_{B_1},\mathbf{\Delta}_{B_2},\ldots,\mathbf{\Delta}_{B_s}) = K^{A_1\cdots A_r}_{B_1\cdots B_s}(-1)^{\epsilon}. \tag{3.7}$$

where

$$\epsilon = \sum_{j=1}^{s-1}\Big(\prod_{i=j}^{s}B_i\Big) + \sum_{j=2}^{r}\Big(\prod_{i=1}^{j}A_i\Big)$$

For any covariant tensor field K of degree s the graded alternating tensor field $A\,\mathbf{K}$ is defined by

$$A\,\mathbf{K}(\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_s) = \frac{1}{s!}\sum_{\pi}\epsilon(\pi)\,\mathbf{K}(\mathbf{X}_{\pi(1)},\mathbf{X}_{\pi(2)},\ldots,\mathbf{X}_{\pi(s)}), \tag{3.8}$$

where Σ_π is taken over all permutation π of $(1,2,\ldots,s)$ and $\epsilon(\pi)$ is the parity of the permutation taking into account the gradings of the arguments [e.g., if (1,2) are both odd and $\pi(1,2) = (2,1)$ then $\epsilon(\pi) = 1$]. The graded wedge product is then

$$\omega\wedge\omega' = A(\omega\otimes\omega'). \tag{3.9}$$

Hence if ω and ω' are r and s forms, respectively,

$$(\omega\wedge\omega')(\mathbf{Y}_1,\ldots,\mathbf{Y}_{r+s}) = \frac{1}{(r+s)!}\sum\epsilon(J;K)\,\omega(\mathbf{Y}_{J_1},\ldots,\mathbf{Y}_{J_r})\,\omega'(\mathbf{Y}_{K_1},\ldots,\mathbf{Y}_{K_s}), \tag{3.10}$$

where the sum is over all possible partitions of $(1,\ldots,r+s)$ into $(J_1,\ldots,J_r)$ and $(K_1,\ldots,K_s)$ and $\epsilon(J;K)$ is the graded sign of the permutation $(1,\ldots,r+s)\to(J_1,\ldots,J_r,K_1,\ldots,K_s)$.

The following are adopted for graded functions g, f^M, coordinates Z^M vector fields $\mathbf{X}$, $\mathbf{Y}$:

(a) $dz^m(\partial_N) = \delta^M_N$,

(b) $df^M = dZ^N(\partial_N f^M) = (-1)^{N(N+M)}(\partial_N f^M)\,dz^N$,

(c) $(\mathbf{X})f^M = (-1)^{fX}\,df^M(\mathbf{X}) = (-1)^{XN}X^N(\partial_N f^M)$,

(d) $\mathscr{L}_{\mathbf{Y}}f^M = (\mathbf{Y})f^M$,

(e) $\mathscr{L}_{\mathbf{Y}}\mathbf{X} = \{\mathbf{Y},\mathbf{X}\} = (\mathbf{XY} - (-1)^{XY}\mathbf{YX})$,

(f) $\mathscr{L}_{\mathbf{Y}}\circ d = d\circ\mathscr{L}_{\mathbf{Y}}$,

(g) $\mathscr{L}_{\mathbf{Y}}(fg) = (\mathbf{Y})(fg) = (\mathbf{Y})f\cdot g + (-1)^{Yf}f\cdot(\mathbf{Y})g$,

(h) $\mathscr{L}_{\mathbf{Y}}(df^M(X)) = (\mathscr{L}_{\mathbf{Y}}(df^M))\cdot\mathbf{X} + (-1)^{Yf}df^M(\mathscr{L}_{\mathbf{Y}}(\mathbf{X}))$,

and any superscript to (-1) is 0 or 1 according to whether the associated symbol is even or odd.

Returning to the construction of a bundle over N, a *linear* frame at $z\in N$ is an ordered basis $(\bar{\partial},\bar{\pi}_+,\bar{\bar{\pi}}_+)$ of $T_z(N)$. Define ON to be the set of all such frames at all points of N. The SL(2C) group acts on the frame at Z according to the rule

$$u\equiv(\bar{\partial},\bar{\pi}_+,\bar{\bar{\pi}}_+)\xrightarrow{Q}(\bar{Q}^{\dagger}\bar{\partial}\,\bar{Q},\bar{\pi}_+\bar{Q},\bar{Q}^{\dagger}\bar{\bar{\pi}}_+), \tag{3.11}$$

so the 14 local coordinates $(z,\hat{\lambda})$ serve to specify a frame in $ON(N,\mathrm{SL2C})$, the linear frame bundle over superspace with SL(2C) as structure group. A connection on this bundle is specified by giving a complex q-vector valued 1-form which may be locally pulled back to a 1-form $\hat{\omega}\in T^*(N)$ transforming under a change of section in $ON(N,\mathrm{SL2C})$ as

$$\hat{\omega}\to Q\hat{\omega}\bar{Q} + Qd\bar{Q}. \tag{3.12}$$

$\hat{\omega}$ is fixed in any local gauge by specifying 48 real superfields. The super torsion 2-form $\mathbf{T}$ is calculated in a manner analogous to that indicated in the first section:

$$\begin{aligned}\mathbf{T}\equiv D_\omega\mathbf{E} &= D_\omega\,\mathbf{e}\oplus D_\omega\epsilon\oplus D_\omega\dot{\epsilon}\\ &= d\mathbf{e} + 2\mathscr{A}(\hat{\omega}\wedge\mathbf{e})\oplus d\epsilon + \hat{\omega}\wedge\epsilon\oplus d\dot{\epsilon} - \dot{\epsilon}\wedge\hat{\omega}^{\dagger}\\ &\equiv \mathbf{T}_B\oplus\mathbf{T}_F\oplus\mathbf{T}_{\dot{F}}\end{aligned} \tag{3.13}$$

Analogously the super curvature 2-form is

$$\hat{\mathbf{R}} = d\hat{\omega} + \hat{\omega}\wedge\hat{\omega}. \tag{3.14}$$

Having established the notion of the linear frame bundle over N whose fibers locate the SL(2,C) orientation of the linear frame, attention is now turned to the definition of a graded affine frame bundle over N.

The basic idea is to regard the tangent vector space $T_z(N)$ at z as an affine space in which one can define an operation of translating the origin of the vector space. With this interpretation it is denoted by $A_z(N)$ and referred to as the tangent affine space. In addition to the ordered basis of tangent vectors used to define the linear frame a graded vector is now required to locate an "origin" for the linear frame within the affine tangent space. Specifying this completes the description of an affine frame at z. Denoting the components of the "origin" vector by $(p^a,p^\alpha,p^{\dot{\alpha}})$ and embedding them in the quaternionic algebra as $(p_b,p_f,p_{\dot{f}})$ the graded affine frame will be denoted by the set

$$(\bar{\mathbf{\Delta}},\mathbb{P}) \equiv (\bar{\partial},\bar{\pi}_+,\bar{\bar{\pi}}_+;S(\bar{\partial}\cdot p_b + \bar{\pi}_+\cdot p_f + \bar{\bar{\pi}}\cdot p_{\dot{f}})), \tag{3.15}$$

where $\mathbb{P}\in A_z(N)$. Such a frame may be identified with a graded affine transformation that maps a standard linear frame at $\mathbb{P} = 0$ into it. The right action of the graded group element $G = (Q,b,\theta,\dot{\theta})$ on the affine frame will be defined by

$$(\bar{\mathbf{\Delta}},\mathbb{P})\to[\bar{Q}^{+}\bar{\partial}\bar{Q},\bar{\pi}_+\bar{Q},\bar{Q}^{+}\bar{\bar{\pi}}_+;\ \mathbb{P} + S(\bar{\partial}\,.(b + p_f\dot{\theta}) + \bar{\pi}_+\theta + \bar{\bar{\pi}}_+\dot{\theta})] \tag{3.16}$$

and it may be verified that

$$((\bar{\mathbf{\Delta}},\mathbb{P})\,G_1)\,G_2 = (\bar{\mathbf{\Delta}},\mathbb{P})(G_1G_2). \tag{3.17}$$

Thus a point in the graded affine frame bundle $AN(N,G)$ can be coordinated by $(z,\hat{\lambda},b,\theta,\dot{\theta})$ and a group motion in the fiber at z induces the affine transformations

$$\begin{aligned}&p_b\to Q[p_b + b + P_f\dot{\theta}]\,Q^{\dagger},\\ &p_f\to Q[p_f + \theta],\\ &\dot{p}_f\to[p_f + \dot{\theta}]\,Q^{\dagger},\end{aligned} \tag{3.18}$$

in addition to the SL(2, C) transformations. Since $(Q,0,0,0)$ is a subgroup of G we may regard $(b,\theta,\dot{\theta})$ as coordinates on the coset space $G/\mathrm{SL}(2,C)$ and there is a correspondence between this coset space and the motion of the graded affine frame at z that has been quotiented by its SL(2, C) rotations. Indeed the coset space may be used as base space of "rigid" supersymmetry with b interpreted as a Minkowski space–

time coordinate. In Minkowski space all the $T_x(M)$ are isomorphic with each other and with M itself. Regarding the latter as an affine space it is clear that any coset frame in $AM(M, \mathrm{SL}(2,\mathrm{C})$ generated by SL(2, C) subgroup will serve as a space on which to define Poincaré transformations induced by P. The fiber symmetry is enlarged by using the "global" supersymmetry operations[10] (with parameter α) as fundamental vector fields

$$\mathbf{X}^{\pm}_{(\alpha)} = S(\partial_\theta \,.\alpha \pm \partial_b \,.\theta\dot{\alpha}), \quad \in T_u(G/\mathrm{SL}(2,\mathrm{C})),$$

in order to generate diffeomorphisms. With the gauging of SL(2C) however, b must regain the status of a group parameter in a fiber over the (curved) superspace base.

The relationship between a connection on $AN(N,G)$ and the linear connection on $ON(N,\mathrm{SL}(2,\mathrm{C}))$ will now be examined.

The graded generalized affine connection is introduced with the set of 1-forms $(\hat{\omega},\eta,\rho,\dot{\rho})\in T^*(N)$, where

$$\hat{\omega} = \omega_a \hat{e}_a\,, \quad \eta = \eta_k\, \tilde{e}_k\,,$$
$$\rho = \rho_\alpha \tilde{e}_\alpha\,, \quad \dot{\rho} = \rho_{\dot{\alpha}} \tilde{e}_{\dot{\alpha}}\,, \tag{3.19}$$

and may be written in a local section

$$\Omega = \hat{\omega} + 2S(P\eta) + S_\rho + S_{\dot{\rho}} \to G^{-1}\Omega G + G^{-1}dG. \tag{3.20}$$

Using the relations (2.4) and the Maurer–Cartan 1-form (2.7) one finds for the element

$$G = (Q,b,\theta,\dot{\theta}),$$
$$\hat{\omega}\to Q\hat{\omega}\bar{Q} + Qd\bar{Q},$$
$$\eta\to Q\,[\eta + D_\omega\, b - \theta D_\omega\, \dot{\theta} + 2\mathscr{H}(\rho\dot{\theta})]\, Q^\dagger,$$
$$\rho\to Q\,[\rho + D_\omega\theta\,],$$
$$\dot{\rho}\to[\,\dot{\rho} + D_\omega\dot{\theta}\,]\, Q^\dagger, \tag{3.21}$$

where covariant derivatives with respect to the connection $\hat{\omega}$ associated with the subgroup $(Q,0,0,0)$ have been used:

$$D_\omega\theta \equiv d\theta + \hat{\omega}\theta\,,$$
$$D_\omega\dot{\theta} \equiv d\dot{\theta} + \dot{\theta}\hat{\omega}^\dagger\,,$$
$$D_\omega b \equiv db + \hat{\omega}b + b\hat{\omega}^\dagger\,. \tag{3.22}$$

The curvature 2-form of this connection is

$$\mathbb{R} = d\Omega + \tfrac{1}{2}\Omega\wedge\Omega$$
$$\equiv \hat{R}_\omega + 2S(PR_\eta) + S_{R_\rho} + S_{R_{\dot{\rho}}} \tag{3.23}$$

where $(\hat{R}_\omega,\ R_\eta,\ R_\rho,\ R_{\dot{\rho}})$ are complex quaternionic valued. Using the generalized algebra exterior product (2.8) one finds

$$\hat{R}_\omega = d\hat{\omega} + \hat{\omega}\wedge\hat{\omega}\,,$$
$$R_\eta = D_\omega\eta + \rho\wedge\dot{\rho}\,,$$
$$R_\rho = D_\omega\rho\,,$$
$$R_{\dot{\rho}} = D_\omega\dot{\rho}\,, \tag{3.24}$$

where

$$D_\omega\eta = d\eta + \hat{\omega}\wedge\eta - \eta\wedge\hat{\omega}^\dagger\,,$$
$$D_\omega\rho = d\rho + \hat{\omega}\wedge\rho\,,$$
$$D_\omega\dot{\rho} = d\dot{\rho} - \dot{\rho}\wedge\hat{\omega}^\dagger\,. \tag{3.25}$$

The connection transformations (3.20) induce the following transformations on the algebra components of $\mathbb{R}$ under $(Q,b,\theta,\dot{\theta})$:

$$\hat{R}_\omega\to Q\hat{R}_\omega\bar{Q}\,,$$
$$R_\eta\to Q\,[R_\eta + D^2_\omega b - \theta D^2_\omega\dot{\theta} + 2\mathscr{H}(D_\omega\rho\dot{\theta})]\, Q^\dagger\,,$$
$$R_\rho\to Q\,[D_\omega\rho + D^2_\omega\theta\,]\,,$$
$$R_{\dot{\rho}}\to[D_\omega\dot{\rho} + D^2_\omega\dot{\theta}\,]\, Q^\dagger\,. \tag{3.26}$$

It may be noted that the second order exterior covariant derivatives in (3.26) generate the homogeneous Q curvatures:

$$D^2_\omega\theta = \hat{R}_\omega\theta\,,$$
$$D^2_\omega\dot{\theta} = \dot{\theta}\hat{R}^\dagger_\omega\,,$$
$$D^2_\omega b = \hat{R}_\omega b + b\hat{R}^\dagger_\omega\,. \tag{3.27}$$

At this point it should be stressed that the generalized affine connection 1-forms $(\hat{\omega},\eta,\rho,\dot{\rho})$ are completely independent of the affine coframe set $(\mathbf{e},\boldsymbol{\epsilon},\dot{\boldsymbol{\epsilon}},p_B,p_F,p_{\dot{F}})$ Now the fiber at z is isomorphic to the G group manifold. If this space is partitioned into its cosets generated by the SL(2, C) subgroup then the coset space $G/\mathrm{SL}(2,\mathrm{C})$ has the correct dimension to set up a correspondence between it and the tangent space $T_z(N)$. More precisely the bundle $AN(N,G)$ may be reducible to the bundle $ON(N,\mathrm{SL}(2,\mathrm{C}))$ by sectioning the fibers $G/\mathrm{SL}(2,\mathrm{C})$. Locally one may regard this as smoothly choosing the group parameters $(b,\theta,\dot{\theta})$ with z so that the associated affine frame is converted to a linear frame with a fixed origin for each z and the structure group is reduced to the residual SL(2, C). Globally the existence of this Higgs mechanism requires the existence of a global section that clearly depends on the nature of the bundle topology. To relate the generalized affine connection 1-forms and the affine frames a Cartan-type connection will be established relating $(\eta,\rho,\dot{\rho})$ to the linear coframes $(\mathbf{e},\boldsymbol{\epsilon},\dot{\boldsymbol{\epsilon}})$ in a covariant manner. Using the transformations (3.21), (3.18) one verifies that such a correspondence may be taken as:

$$\rho = \boldsymbol{\epsilon} + D_\omega\, p_F\,,$$
$$\dot{\rho} = \dot{\boldsymbol{\epsilon}} + D_\omega\, p_{\dot{F}}\,,$$
$$\eta = -i\mathbf{e} + D_\omega\, p_b - p_F D_\omega p_{\dot{F}} + 2\mathscr{H}(\boldsymbol{\epsilon}p_{\dot{F}})\,. \tag{3.28}$$

For a given $\hat{\omega}$ on $ON(N,\mathrm{SL}(2,\mathrm{C}))$ and canonical form $(\mathbf{e},\boldsymbol{\epsilon},\dot{\boldsymbol{\epsilon}})$ the forms $(\hat{\omega},\eta,\rho,\dot{\rho})$ fixed by this condition establish a graded Cartan connection on $AN(N,G)$. With the aid of the fundamental vector fields on this bundle one can study the parallel transport of affine frames on lifted curves in terms of the SL(2, C) rotations of the linear frames and their translations $(p_B,\ p_F,\ p_{\dot{F}})$ in the tangent affine space. The curvature of the graded Cartan connection can be now calculated in terms of the linear coframe fields:

$$\mathbf{R}_\eta = -i\mathbf{T}_B + \boldsymbol{\epsilon}\wedge\dot{\boldsymbol{\epsilon}} + D^2_\omega P_b - p_F\dot{p}_F\hat{\mathbf{R}}^\dagger_\omega + 2\mathscr{H}(\mathbf{T}_F\dot{p}_F)\,, \tag{3.29}$$

$$\mathbf{R}_\rho = \mathbf{T}_F + \hat{\mathbf{R}}_\omega p_F\,, \tag{3.30}$$

$$\mathbf{R}_{\dot{\rho}} = \mathbf{T}_{\dot{F}} + \dot{p}_F\hat{\mathbf{R}}^\dagger\,. \tag{3.31}$$

If the structure group is reduced to SL(2, C) with a section that sets $\mathbb{P} = 0$ the linear coframes become identified with

the graded translation components of the AN connection and the affine curvatures are simply related to the torsions on ON.

A relation has consequently been traced between the gauging of the graded affine group in superspace and the geometry established by reducing $AN(N,G)$ to $ON(N,\mathrm{SL}(2,\mathrm{C}))$. Within this residual bundle of graded l linear frames the SL(2, C) connection ω is entirely free at this point. This process of bundle reduction may be phrased in the language of spontaneous symmetry breakdown. Indeed by replacing the graded affine group by the graded de-Sitter group much of the discussion on development and reduction in Ref. 4 may be generalized to the situation discussed in this paper. However, a complete motivation along these would appear to demand a generalized action principle with solutions giving topological information.

To make contact with simple supergravity in a superspace geometry Wess and Zumino[11] have fixed the connection $\hat{\omega}$ in terms of the torsion $(\mathbf{T}_B,\ \mathbf{T}_F,\ \mathbf{T}_{\dot{F}})$:

$$\mathbf{T}_B = i\boldsymbol{\epsilon}\wedge\dot{\boldsymbol{\epsilon}}\,, \tag{3.32}$$

$$\mathbf{T}_F = \left[T^{\alpha}_{bc}\mathbf{e}^b\wedge\mathbf{e}^c + T^{\alpha}_{\beta c}\boldsymbol{\epsilon}^{\beta}\wedge\mathbf{e}^c + T^{\alpha}_{\beta c}\boldsymbol{\epsilon}^{\dot{\beta}}\wedge\mathbf{e}^c\right]\vec{e}_{\alpha}\,,$$

$$\mathbf{T}_{\dot{F}} = \left[T^{\dot{\alpha}}_{bc}\mathbf{e}^b\wedge\mathbf{e}^c + T^{\dot{\alpha}}_{\beta c}\,\boldsymbol{\epsilon}^{\beta}\wedge\mathbf{e}^c + T^{\dot{\alpha}}_{\dot{\beta}c}\boldsymbol{\epsilon}^{\dot{\beta}}\wedge\mathbf{e}^c\right]\vec{e}_{\dot{\alpha}}\,. \tag{3.33}$$

One observes that (3.32) implies the vanishing of $\mathbf{R}_{\eta}$ in the $\mathbb{P}=0$ gauge. This connection will be employed in the last section where a return is made to the local supersymmetry transformations discussed in Sec. 1.

4. CONNECTION PRESERVING SUPERSPACE DIFFEOMORPHISMS

The use of tensors over superspace to discuss supergravity of course relegates passive supercoordinate transformations at z to the status of labelling conventions. The fundamental nature of local supersymmetry transformations on the manifold N must be sought in the nature of particular transformations generated by certain vector fields. Having reduced the gauge group to SL(2, C) it may at first seem unlikely that transformations on the linear frames analogous to the graded affine transformations (3.21) could be given any natural formulation in terms of covariant operations involving the SL(2, C) connection $\hat{\omega}$. It will be shown, however, that at least within a certain choice of "gauge" (the Wess–Zumino gauge) such transformations can be identified with an intrinsic derivation.

The basic observation is to demand that once the torsion conditions have established an SL(2, C) connection on $ON(N,\mathrm{SL}(2,\mathrm{C}))$, transformations should be sought that leave this choice invariant. More precisely a diffeomorphism $f:N\to N$ that maps one point of N smoothly to another should induce no change in $\hat{\omega}$.

$$f^*\hat{\omega} = \hat{\omega}. \tag{4.1}$$

(By comparison one recalls that on a space with a metric tensor g, f is an isometry if $f^*g = g$. The generators of f are termed Killing vectors). Incidentally any transformation will induce a transformation on ON that leaves the canonical form invariant. Such a diffeomorphism is characterized by a vector field $\mathbf{X} = \partial_M\cdot X^M$ say which may be thought of as the velocity field of a steady flow on the manifold as each point displaces according to the map f. A particular characterization of $\mathbf{X}$ will be sought (by finding an equation for its components in some basis) so that the Lie derivative of $\hat{\omega}$ vanishes. Such a vector field is said to generate an affine transformation.[12]

In dealing with active motions on a manifold with connection it is preferable to translate the connection one form into a graded type-preserving Koszul connection ∇. Given $\hat{\omega}$ and with an arbitrary vector field $\mathbb{X}$:

$$\mathbb{X} = \mathbf{X} + \boldsymbol{\lambda} + \dot{\boldsymbol{\mu}} \equiv \partial_a\cdot X^a + \boldsymbol{\pi}_{\alpha}\cdot\lambda^{\alpha}\boldsymbol{\pi}_{\dot{\alpha}}\cdot\lambda^{\dot{\alpha}}\,. \tag{4.2}$$

∇ is defined by

$$\begin{aligned}
&\nabla_{\mathbf{X}}\bar{\partial} = 2\mathscr{A}(\bar{\partial}\cdot\hat{\omega}(\mathbf{X}))\,,\\
&\nabla_{\boldsymbol{\lambda}}\bar{\partial} = 2\mathscr{A}(\bar{\partial}\cdot\hat{\omega}(\boldsymbol{\lambda}))\,,\\
&\nabla_{\dot{\mu}}\bar{\partial} = 2\mathscr{A}(\bar{\partial}\cdot\hat{\omega}(\dot{\mu}))\,,\\
&\nabla_{\mathbf{X}}\bar{\boldsymbol{\pi}}_+ = \bar{\boldsymbol{\pi}}_+\hat{\omega}(\mathbf{X})\,,\\
&\nabla_{\boldsymbol{\lambda}}\bar{\boldsymbol{\pi}}_+ = \bar{\boldsymbol{\pi}}_+\hat{\omega}(\boldsymbol{\lambda})\,,\\
&\nabla_{\dot{\mu}}\bar{\boldsymbol{\pi}}_+ = \bar{\boldsymbol{\pi}}_+\hat{\omega}(\dot{\mu})\,,\\
&\nabla_{\mathbf{X}}\bar{\bar{\boldsymbol{\pi}}}_+ = \hat{\omega}(\mathbf{X})\bar{\bar{\boldsymbol{\pi}}}_+\,,\\
&\nabla_{\boldsymbol{\lambda}}\bar{\bar{\boldsymbol{\pi}}}_+ = \hat{\omega}(\boldsymbol{\lambda})\bar{\bar{\boldsymbol{\pi}}}_+\,,\\
&\nabla_{\dot{\mu}}\bar{\bar{\boldsymbol{\pi}}}_+ = \hat{\omega}(\dot{\mu})\bar{\bar{\boldsymbol{\pi}}}_+\,.
\end{aligned} \tag{4.3}$$

If one agrees to commute quaternionic elements across $\otimes$ these formula are summarized in the familiar form:

$$\nabla(\bar{\Delta}) = 2\mathscr{A}(\bar{\partial}\otimes\hat{\omega}) + \bar{\boldsymbol{\pi}}_+\otimes\hat{\omega} + \hat{\omega}^+\otimes\bar{\bar{\boldsymbol{\pi}}}_+ \tag{4.4}$$

In this language the (1,2) torsion tensor $\mathbb{T}$ is defined by

$$\mathbb{T}(\mathbf{X},\mathbf{Y}) = \nabla_{\mathbf{X}}\mathbf{Y} - \nabla_{\mathbf{Y}}\mathbf{X} - \{\mathbf{X},\mathbf{Y}\} \tag{4.5}$$

and the curvature operator of this connection is

$$\mathbb{R}(\mathbf{X},\mathbf{Y}) = \{\nabla_{\mathbf{X}},\nabla_{\mathbf{Y}}\} - \nabla_{\{\mathbf{X},\mathbf{Y}\}}\,, \tag{4.6}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are arbitrary graded vector fields on N. The derivation that is needed to determine the generator $\mathbf{X}$ of affine transformation is $A_X \equiv \mathscr{L}_X - \nabla_X$. For any vector fields $\mathbf{X}$, $\mathbf{Y}$, one sees from (4.5) that

$$A_{\mathbf{X}}\mathbf{Y} = -\nabla_{\mathbf{Y}}\mathbf{X} - \mathbb{T}(\mathbf{X},\mathbf{Y})\,. \tag{4.7}$$

Furthermore it is not difficult to show that the required generator $\mathbf{X}$ must satisfy

$$\nabla_{\mathbf{Y}}(A_{\mathbf{X}}) = \mathbb{R}(\mathbf{X},\mathbf{Y}) \tag{4.8}$$

for all $\mathbf{Y}$. Defining ∇_N by

$$\nabla_{\mathbf{Y}}(\partial_M\cdot X^M) = \partial_M\cdot(\nabla_N X^M)\,Y^N\,, \tag{4.9}$$

Eq. (4.8) implies that in a local coordinate system

$$\nabla_L(\nabla_J X^I + T^I_{KJ}X^K) + R^I_{JKL}X^K = 0\,. \tag{4.10}$$

In the Wess-Zumino gauge one can examine the action of $A_{\mathbf{X}}$ on ∂_M for $\mathbf{X}$ an odd affine generator and project the resulting vector onto an anholonomic basis. Thus with $\mathbf{E}^A(\mathbb{T}) = \mathbf{T}^A$,

$$\mathbf{E}^A(A_{\mathbf{X}}\partial_M) = -\nabla_M X^A - \mathbf{T}^A(\partial_B\cdot X^B, \partial_M)\,. \tag{4.11}$$

Comparing this result with Ref. 13 leads to the identification $\mathbf{E}^A(A_{\mathbf{X}}\,\partial_M) = \delta E^A_M$ and the calculations therein recover the local supersymmetry transformations together with their

minimal auxilliary field content. Thus in the Wess–Zumino gauge the coframe 1-forms $\mathbf{e}$ and $\dot{\mathbf{e}}$ contain the gravitino 1-forms χ and $\dot{\chi}$ introduced in Sec. 1 and the bosonic frame $\mathbf{e}$ contains the space-time tetrad (whose coordinate components are the conventional vierbeins e^a_m). In this gauge, for example, by taking components of (4.11) there results:

$$\begin{aligned}\delta E^{\ a}_m &= \delta e^a_m(x) = e^a(A_X\ \partial_{\mathbf{m}}) \\ &= -\mathbf{T}^a(\pi_\beta, X^\beta, \pi_\alpha)\,\chi^\alpha_m(x) - \mathbf{T}^a(\pi_{\dot\beta}, X^{\dot\beta}, \pi_{\dot\alpha})\,\chi^{\dot\alpha}_m(x), \\ \delta E^{\ \alpha}_m &= \delta\chi^\alpha_m(x) = e^\alpha(A_X\ \partial_{\mathbf{m}}) \\ &= -D_m\, X^\alpha - \mathbf{T}^\alpha(\pi_\beta, X^\beta, \partial_a)\, e^a_m \\ &\quad - \mathbf{T}^\alpha(\pi_{\dot\beta}, X^{\dot\beta}, \partial_a)\, e^a_m,\end{aligned} \tag{4.12}$$

where

$$D_m\, X^\alpha \equiv\equiv -e^a_m(\partial_a\, X^\alpha) + \sum_{\beta=\gamma,\dot\gamma} [\tilde e^\alpha \hat\omega(x)_a\ e^a_m(x)]_\beta\, X^\beta .$$

Thus the complete local supersymmetry transformations of supergravity appear in the Wess–Zumino superspace gauge to be intimately related to the derivation A_X which in turn may be used to specify the generator of affine transformation on the superspace manifold.

CONCLUSION

In this paper the relation between the gauging of the graded Affine group and the process of reduction to the SL(2, C) gauge group has been traced with the aid of a Cartan type connection on the graded Affine frame bundle. The methodology offers several generalizations to extended and unified theories of supergravity formulated in a superspace. Once a graded algebra has been chosen to extend the Affine group an extended bundle of Affine frames would appear to offer an attractive arena for the discussion of spontaneous bundle reduction. The dimension and grading of the base supermanifold can be adjusted so that a matching can be achieved between the coset spaces that prefigure in the spontaneous symmetry breakdown and the generalized superspace tangent spaces. The phenomenon of mass generation in matter fields would then be tied to the existence of a section that includes gravitational effects. Certainly it is to be expected that the interpretation of spontaneous symmetry breakdown will undergo a modification in the presence of gravity.

ACKNOWLEDGMENT

The author expresses thanks to Dr. C.T. Dodson for his patience.

APPENDIX

The 3 elements $\hat e_a$ $(a = 1,2,3,)$ in the ring of quaternions obey

$$\hat e_a \hat e_b = -\delta_{ab} + \epsilon_{abc}\hat e_c . \tag{A1}$$

A general complex quaternion is denoted

$$q = a_4 + \sum_{b=1}^{3} a_b \hat e_b , \tag{A2}$$

where a_4 and $a_b \in C$. The conjugate quaternion $\bar q$ is

$$\bar q = a_4 - \sum_{b=1}^{3} a_b \hat e_b . \tag{A3}$$

Complex conjugation commutes with quaternionic conjugation

$$q^* = a_4^* + \sum_{b=1}^{3} a_b^* \hat e_b \tag{A4}$$

and the Hermitian conjugate is $q^\dagger = \bar q^*$. In terms of these operations are defined the following:

$$\begin{aligned} &2\mathrm{Re}(q) = q + q^*, \\ &2i\mathrm{Im}(q) = q - q^*, \\ &2S(q) = q + \bar q, \\ &2V(q) = q - \bar q, \\ &2\mathscr{H}(q) = q + q^\dagger, \\ &2\mathscr{A}(q) = q - q^\dagger . \end{aligned} \tag{A5}$$

If $V(q) = 0\,[S(q) = 0]$ q will be called a q scalar (q vector). It is Hermitian (anti-Hermitian) if $q^\dagger = +(-)\,q$. Four real entities (b_0, b_a) may be embedded into an anti-Hermitian quaternion by writing

$$b = ib_0 + \sum_{a=1}^{3} b_a \hat e_a . \tag{A6}$$

If a 4-space has metric $g = -\mathbf{e}^o \otimes \mathbf{e}^o + \Sigma_{a=1}^3\, \mathbf{e}^a \otimes \mathbf{e}^a$ in terms of real 1-forms $(\mathbf{e}^o, \mathbf{e}^a)$ then an anti-Hermitian coframe is defined as

$$\mathbf{e} = i\mathbf{e}^o + \sum_{a=1}^{3} \mathbf{e}^a \hat e_a , \tag{A7}$$

with norm $\mathbf{e}\bar{\mathbf{e}} = \bar{\mathbf{e}}\mathbf{e} = ds^2$. If Q is a complex unit norm quaternion $Q\bar Q = \bar Q Q = 1$ then $\mathbf{e}' = Q\mathbf{e}Q^\dagger$ is also anti-Hermitian and has unit norm

$$\mathbf{e}'\bar{\mathbf{e}}' = Q\mathbf{e}Q^\dagger \bar Q^\dagger \bar{\mathbf{e}} \bar Q = Q\mathbf{e}\bar{\mathbf{e}}\bar Q = \mathbf{e}\bar{\mathbf{e}}, \tag{A8}$$

since $\mathbf{e}\bar{\mathbf{e}}$ is a q scalar. Hence Q generates proper local Lorentz transformations and it may be parametrized in terms of 3 angles (α_a) and 3 boosts (β_α):

$$Q = e^{\hat\alpha + i\hat\beta},$$

where

$$\hat\alpha = \sum_{a=1}^{3} \alpha_a \hat e_a ,$$

$$\hat\beta = \sum_{a=1}^{3} \beta_a \hat e_a ,$$

and the $\hat{}$ symbol will in general denote a q-vector.

A complex quaternionic p form A_p is a p form with complex quaternionic components in any basis of a real cotangent space. For graded forms there is the relation

$$\overline{A_p \wedge B_q} = (-1)^{pq+AB}\, \bar B_q \wedge \bar A_p , \tag{A10}$$

although $A_p \wedge B_q = (-1)^{pq+AB} B_q \wedge A_p$ only if one of the forms is q scalar. Under complex conjugation of odd o forms the rule

$$(\phi_1\phi_2)^* = \phi_2^*\phi_1^* = -\phi_1^*\phi_2^* \tag{A11}$$

is adopted. Hence for general odd forms

$$(\alpha \wedge \beta)^* = -\alpha^* \wedge \beta^* \tag{A12}$$

and for all forms (even or odd)

$$(A_p \wedge B_q)^\dagger = (-1)^{pq} B_q^\dagger \wedge A_p^\dagger . \tag{A13}$$

If A is a complex 2×2 matrix with $\det A = 1$ the complex SL(2, C) spinors are 2-component vectors transforming as

$$\begin{aligned} \phi'_r &= (A\phi)_r , \\ \phi'^r &= (\phi A^{-1})^r , \\ \psi'_{\dot r} &= (\psi A^{+})_{\dot r} , \\ \psi'^{\dot r} &= (A^{+-1}\psi)^{\dot r} , \end{aligned} \tag{A14}$$

where A^+ is the complex transpose of A. Writing

$$A = \begin{pmatrix} q_4 - iq_3 & -q_2 - iq_1 \\ q_2 - iq_1 & q_4 + iq_3 \end{pmatrix}, \tag{A15}$$

where $Q = q_4 + \Sigma_{a=1}^3 q_a \hat e_a$ then $Q\bar Q = 1$ implies $\det A = 1$. It may be verified that if q spinors are defined as

$$\phi_+ = \phi_1 U^1 + \phi_2 \hat U^2 , \tag{A16}$$

$$\phi_- = \phi^1 W^1 + \phi^2 \hat W^2 , \tag{A17}$$

and transform as $\phi_\pm \to Q\phi_\pm$ under SL(2, C), then the components $\binom{\phi_1}{\phi_2}$ and (ϕ^1, ϕ^2) transform as the index structure suggests (A14). The ideals are constructed with the elements

$$\begin{aligned} U^1 &= (1/\sqrt{2})(1 + i\hat e_3) , \\ \hat U^2 &= (1/\sqrt{2})(\hat e_2 + i\hat e_1) , \\ W^1 &= (1/\sqrt{2})(1 - i\hat e_3) , \\ \hat W^2 &= (1/\sqrt{2})(\hat e_2 - i\hat e_1) . \end{aligned} \tag{A18}$$

Similarly the dotted q spinors,

$$\dot\phi_+ = \dot\phi_{\dot 1} U^1 + \dot\phi_{\dot 2} \hat W^2 , \tag{A19}$$

$$\dot\phi_- = \dot\phi_{\dot 1} W^1 + \dot\phi^{\dot 2} \hat U^2 , \tag{A20}$$

transform as $\dot\phi_\pm \to \dot\phi_\pm Q^\dagger$. A q spinor λ without a subscript will be taken as λ_- by convention.

A typical tangent vector in superspace may be written in the $(\partial_a, \pi_\alpha, \bar\pi_{\dot\alpha})$ basis as

$$\begin{aligned} \mathbf{X} &= S(\bar{\mathbf{\partial}} X + \bar{\boldsymbol\pi}_+ \lambda + \bar{\dot{\boldsymbol\pi}}_+ \dot\mu) \\ &= \sum_{k=1}^4 \partial_k X^k + \sum_{\alpha=1}^2 \pi_\alpha \lambda^\alpha + \sum_{\dot\alpha=1}^2 \Pi_{\dot\alpha} \lambda^{\dot\alpha} , \end{aligned} \tag{A21}$$

where

$$\begin{aligned} X &= X^k \tilde e_k , \quad \bar{\mathbf{\partial}} = \partial_k \tilde e^k , \\ \lambda &= \lambda^\alpha \tilde e_\alpha \quad \bar{\boldsymbol\pi}_+ = \pi_\alpha \tilde e^\alpha , \\ \dot\mu &= \mu^{\dot\alpha} \tilde e_{\dot\alpha} , \quad \bar{\dot{\boldsymbol\pi}} = \pi_\alpha \tilde e^\alpha , \\ \tilde e_k &= (1, \hat e_a) , \quad \tilde e^k = (1, -\hat e_\alpha) , \\ \partial_4 &= -i\partial_0 , \quad \mathbf{e}^4 = i\mathbf{e}^0 \\ S(\tilde e_\alpha \tilde e^\beta) &= \delta_\alpha^\beta , \quad k,l = 1,2,3,4, \\ a &= 1,2,3, \quad \alpha,\dot\alpha = 1,2. \\ \tilde e_\alpha &= (W^1, \hat W^2) , \quad \tilde e^\alpha = (W^1, -\hat U^2) , \\ \tilde e_{\dot\alpha} &= (W^1, \hat U^2) , \quad \tilde e^{\dot\alpha} = (W^1, -\hat W^2) , \\ S(\tilde e_k \tilde e^l) &= \delta_k^l , \quad S(\tilde e_\alpha \tilde e^\beta) = \delta_\alpha^\beta , \end{aligned}$$

A complete duality between the quaternions $\tilde e^A = (\tilde e^k, \tilde e^\alpha, \tilde e^{\dot\alpha})$ and $\tilde e_A = (\tilde e_k, \tilde e_\alpha, \tilde e_{\dot\alpha})$ may be set up in an obvious manner so these basis quaternions need never appear in any practical calculations. For convenience the relative transformation properties of the various q spinors can be immediately determined from their notation.

A Majorana spinor p form is one where all four types of complex spinor p forms can be expressed in terms of two independent complex spinor p forms.

Denoting these by ϕ_A and ϕ_B for illustration a Majorana related set is

$$\begin{aligned} \phi_+ &= -\phi_B U^1 + \phi_A \hat U^2 , \\ \phi_- &= \phi_A W^1 + \phi_B \hat W^2 , \end{aligned} \tag{A22}$$

$$\begin{aligned} \dot\phi_+ &= i(\phi_B^* U^1 + \phi_A^* \hat W^2) , \\ \dot\phi_- &= i(\phi_A^* W^1 - \phi_B^* \hat U^2) , \end{aligned} \tag{A23}$$

In general they satisfy the relation

$$\phi^\dagger_\pm = \pm i\dot\phi_\pm . \tag{A24}$$

[1] A. Trautman, Bull. Acad. Pol. Sci. Ser. Math. Astron. Phys. **20**, 185, 503, 895 (1972); **21**, 345 (1973). D.V. Volkov and V.A. Soroka, JETP Lett. **18**, 312 (1973). A.H. Chamseddine and P.C. West, Nucl. Phys. B **129**, 39 (1977). Y. Ne'eman and T. Regge, Rivista del Nuovo Cimento **3**, (1978). S.W. Macdowell, Phys. Lett. A **80**, 212 (1979).

[2] K. Nomizu, Am. J. Math. **76**, 33 (1954). S. Kobayashi, Can. J. Math. **8**, 145 (1956).

[3] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. 4. C.T.J. Dodson, Int. J. Theor. Phys. **17**, 389 (1978).

[4] K.S. Stelle and P.C. West, ICTP/78-79/19 (Imperial College, 1979).

[5] D. Hestenes, *Space-Time Algebra* (Gordon and Breach, New York, 1966).

[6] Attention is drawn to the convention of ordering the vector basis to the left of the component. The reason is motivated in Sec. 3. The dot is designed to distinguish the vector $\partial_a X^a$ from the number $\partial_a X^a$).

[7] D.Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, Phys. Rev. **D 13**, 3214 (1976). S. Deser and B. Zumino, Phys. Lett. B **62**, 335 (1976).

[8] B. Kostant, in Proc. Symposium. Bonn. 1975, *Lecture Notes in Mathematics* 570 (Springer, Berlin, 1977). L. Smolin, HUTP-78/AO43 (Harvard College, 1978). A. Rogers, ICTP/78-79/15 (Imperial College, 1979).

[9] Although *ON* does not refer to any particular metric in superspace the term "orthonormal" will be taken as synonymous with anhalonomic ($dE \neq 0$) here.

[10] A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974).

[11] J. Wess and B. Zumino, Phys. Lett. B **74**, 51 (1978).

[12] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. 1, p. 230.

[13] J. Wess and B. Zumino, Phys. Lett. B **79**, 394 (1978).

Techniques for naturality isolation: I. With a Bosonic t channel

Richard Stacey[a)]
Department of Physics, Westfield College, University of London, London NW3 7ST, England

(Received 2 October 1979; accepted for publication 24 January 1980)

We develop the mathematical apparatus necessary to isolate pure naturality contributions from the helicity amplitudes for an arbitrary 2→2 scattering process (with a Bosonic t channel). We prove that certain combinations of the amplitudes and their (multiple) energy-derivatives suffice, and that no simpler methods exist. For high spin processes, where our methods might be overcomplicated, we develop a simple and accurate approximation using just a single energy-derivative.

PACS numbers: 11.50.Nk, 11.80.Fv

I. INTRODUCTION

In the next few years it is hoped that we will achieve a much more thorough understanding of the structure of scattering processes than we have at present. Experimental progress is being made in the measurement of a wide variety of spin-dependent observables[1]; the eventual aim will be complete amplitude analyses over a range of energies for the different processes. The new results should stimulate further theoretical developments. For the moment we must determine how to interpret the measurements in terms of the different types of t channel exchange participating.

Recently we published an account of how to isolate exchanges of definite naturality $[P(-)^J]$ from the helicity amplitudes for an arbitrary 2→2 scattering process.[2] In order to keep the analysis simple and practical we omitted to prove our results. Here we will repair that omission.

The problem is highly nontrivial and none the less important for having been swept under the carpet for many years. In Ref. 2 we spent some time demonstrating that conventional approximations to the same results are generally worthless at moderate energies. Here we will take this for granted. First of all we state the general problem in Sec. II A. Phrased mathematically, this involves finding J-independent functions $G(\lambda,\mu)$ such that

$$G(\lambda,\mu)d^J_{\lambda\mu}(z_t) = G(\lambda,-\mu)d^J_{\lambda\,-\mu}(z_t), \quad \forall J .$$

The solution turns out to depend greatly on whether the t channel is bosonic or fermionic (λ and μ integral or half-odd integral). In this paper we treat the bosonic case, the simpler of the two. The fermionic case is considered in an accompanying paper.[3]

We determine $G(\lambda,\mu)$ $(\lambda,\mu\in\mathbb{Z})$ in Sec. II B. It is a finite polynomial in $z_t(=\cos\theta_t)$ and $D(=d/dz_t)$ of order

$$m = \min(|\lambda|,|\mu|)$$

in D. In Sec. II C we develop our solution into a more useful form, and in Sec. II D we show that we need at least m derivatives to achieve an exact separation. That is, our solution is as good as it could be.

For large m our exact results may be rather difficult to use. In Sec. III we derive the general single-derivative approximation to naturality isolation and demonstrate that it is likely to be reliable at moderate energies.

Section IV contains a summary and some conclusions

II. GENERAL RESULTS

A. General problem

We start by considering a completely general 2→2 process

$$A + B \to C + D .$$

The t channel is labelled

$$\bar{D} + B \to C + \bar{A} .$$

Our program is to isolate from $f^t_{c\bar{a};\bar{d}b}$ and $f^t_{\,c\,-\,\bar{a};\bar{d}b}$ components of definite naturality. To do this we need the t-channel partial-wave expansions[4]

$$f^t_{c\bar{a};\bar{d}b} = \sum_J (\langle c\bar{a}|T^J(+)|\bar{d}b\rangle + \langle ca|T^J(-)|\bar{d}b\rangle)d^J_{\lambda\mu}(z_t) , \tag{1}$$

$$f^t_{\,c\,-\,\bar{a};\bar{d}b} = \xi\sum_J (\langle c\bar{a}|T^J(+)|\bar{d}b\rangle - \langle c\bar{a}|T^J(-)|\bar{d}b\rangle)d^J_{\lambda\,-\,\mu}(z_t) ,$$

where

$$z_t = \cos\theta_t, \quad \mu = c - \bar{a}, \quad \lambda = \bar{d} - b . \tag{2}$$

The $\langle c\bar{a}|T^J(p)|\bar{d}b\rangle$ are transition amplitudes of definite naturality (p) depending on t alone (not z_t). ξ is a phase factor.[5]

Our program is clearly equivalent to discovering functions $G(\lambda,\mu)$, independent of J, such that

$$G(\lambda,\mu)d^J_{\lambda\mu}(z_t) = G(\lambda,-\mu)d^J_{\lambda\,-\,\mu}(z_t) \tag{3}$$

for given such $G(\lambda,\mu)$ we can define

$$F(p) \equiv \tfrac{1}{2}[G(\lambda,\mu)f^t_{c\bar{a};\bar{d}b} + p\xi G(\lambda,-\mu)f^t_{\,c\,-\,\bar{a};\bar{d}b}]$$
$$= \sum_J \langle c\bar{a}|T^J(p)|\bar{d}b\rangle G(\lambda,\mu)d^J_{\lambda\mu}(z_t) \tag{4}$$

and $F(p)$ has contributions from naturality p alone.[6]

B. Central result

Theorem: There exist J-independent operators $G(\lambda,\mu)$ such that $\forall J$ $(\lambda,\mu\in\mathbb{Z})$

$$G(\lambda,\mu)d^J_{\lambda\mu} = G(\lambda,-\mu)d^J_{\lambda\,-\,\mu} .$$

If the sign of $\lambda\mu$ is positive, then

[a)]Present address: Department of Physics and Astronomy, University College London, Gower Street, London WCIE 6BT.

 0022-2488/81/020430-05$1.00

$$G(\lambda,\mu) = \prod_{r=1}^{m} [(z_t^2-1)D + z_t - M + m - 2r + 1\ ], \tag{5a}$$

while if the sign of $\lambda\mu$ is negative, then

$$G(\lambda,\mu) = (-)^{\mu} \prod_{r=1}^{m} [(z_t^2-1)D + z_t + M - m + 2r - 1\ ]. \tag{5b}$$

Proof: We use the notation of Sec. II A together with

$$z_{\pm} = z_t \pm 1$$

and

$$M(m) = \max(\min)(|\lambda|,|\mu|)\,.$$

Our starting point is the definition of the functions $e^J_{\lambda\mu}$ given in Ref. 7. These are related to the $d^J_{\lambda\pm\mu}$ we are interested in by

$$d^J_{\lambda\mu}(z_t) = z_+^{(1/2)|\lambda+\mu|}(-z_-)^{(1/2)|\lambda-\mu|}(e^J_{\lambda\mu} + e^J_{\lambda\mu})\,, \tag{6}$$

$$(-)^{\lambda+M} d^J_{\lambda-\mu}(z_t) = z_+^{(1/2)|\lambda-\mu|}(-z_-)^{(1/2)|\lambda+\mu|}(e^J_{\lambda\mu} - e^J_{\lambda\mu})\,.$$

First we will relate $e^J_{\lambda\mu}$; clearly then $d^J_{\lambda\pm\mu}$ are related. Trivial manipulations on Ref. 7 give, up to a common constant,

$$\begin{aligned} e^J_{\lambda\mu} + s e^J_{\lambda\mu} &= D^{M-m}(Dz_-D)^m P_J(z_t)\,, \\ e^J_{\lambda\mu} - s e^J_{\lambda\mu} &= D^{M-m}(Dz_+D)^m P_J(z_t)\,, \end{aligned} \tag{7}$$

[s = sign of $(\lambda\mu)$.] Since we are concerned only with linear relations on $e^J_{\lambda\mu}$ the constant makes no difference and we will ignore it. Now by induction,

$$(Dz_{\pm}D)^m = D^m z_{\pm}^m D^m\,.$$

Hence we can rewrite Eq. (7):

$$\begin{aligned} z_+^m (e^J_{\lambda\mu} + s e^J_{\lambda\mu}) &= z_+^m D^M z_-^m D^{m-M}(D^M P_J(z_t))\,, \\ z_-^m (e^J_{\lambda\mu} - s e^J_{\lambda\mu}) &= z_-^m D^M z_+^m D^{m-M}(D^M P_J(z_t))\,. \end{aligned} \tag{9}$$

Now (by induction again)

$$\begin{aligned} z_+^m D^M z_-^m D^{m-M} &= \prod_{r=1}^{m} [(z_t^2-1)D + (M-m+1)z_t \\ &\quad + M - m + 2r - 1\ ], \\ z_-^m D^M z_+^m D^{m-M} &= \prod_{r=1}^{m} [(z_t^2-1)D + (M-m+1)z_t \\ &\quad - M + m - 2r + 1]\,. \end{aligned} \tag{10}$$

Here we make a crucial observation. *All* the terms in *both* products in Eq. (10) differ from each other only by constants. Therefore they all commute, and hence so do the products. Thus

$$[z_+^m D^M z_-^m D^{m-M},\ z_-^m D^M z_+^m D^{m-M}] = 0\,. \tag{11}$$

Combining (9) and (11), we get

$$\begin{aligned} &z_-^m D^M z_+^m D^{m-M} z_+^m (e^J_{\lambda\mu} + s e^J_{\lambda\mu}) \\ &\quad = z_+^m D^M z_-^m D^{m-M} z_-^m (e^J_{\lambda\mu} - s e^J_{\lambda\mu})\,, \end{aligned} \tag{12}$$

a fundamental relation for our purposes.

Next we must combine Eqs. (12) and (6) to relate $d^J_{\lambda\pm\mu}$. If the sign of $\lambda\mu$ is positive, Eq. (6) becomes

$$d^J_{\lambda\mu}(z_t) = (z_t^2-1)^{(1/2)(M-m)} z_+^m (e^J_{\lambda\mu} + e^J_{\lambda\mu})(-)^{(1/2)(M-m)}\,, \tag{13}$$

$$(-)^{\lambda+M-m} d^J_{\lambda-\mu}(z_t) = (z_t^2-1)^{(1/2)(M-m)} z_-^m (e^J_{\lambda\mu} - e^J_{\lambda\mu}) \times (-)^{(1/2)(M-m)}\,.$$

Hence

$$I d^J_{\lambda\mu} = \eta K d^J_{\lambda-\mu}\,, \tag{14}$$

where

$$\begin{aligned} I &= z_-^m D^M z_+^m D^{m-M}(z_t^2-1)^{(1/2)(m-M)}\,, \\ K &= z_+^m D^M z_-^m D^{m-M}(z_t^2-1)^{(1/2)(m-M)}\,, \end{aligned} \tag{15}$$

and

$$\eta = (-)^{\lambda+M-m}\,.$$

If the sign of $\lambda\mu$ is negative, then the sign of $\lambda(-\mu)$ is positive, so

$$\eta K d^J_{\lambda\mu} = I d^J_{\lambda-\mu}\,. \tag{16}$$

Combining (14) and (16), we can now define

$$G(\lambda,\mu) = (z_t^2-1)^{(1/2)(M-m)}[\tfrac{1}{2}(I+\eta K) + s\tfrac{1}{2}(I-\eta K)]\,. \tag{17}$$

Then, quite generally,

$$G(\lambda,\mu) d^J_{\lambda\mu} = G(\lambda,-\mu) d^J_{\lambda-\mu}\,. \tag{18}$$

$G(\lambda,\mu)$ is defined by Eqs. (15) and (17) to be independent of J. To derive Eqs. (5) we use Eq. (10) and observe that

$$(z_t^2-1)^{(1/2)(M-m)}[(z_t^2-1)D + (M-m+1)z_t\,] \times (z_t^2-1)^{(1/2)(m-M)} = (z_t^2-1)D + z_t\,.$$

Hence Eqs. (17) and (5) are equivalent. We have proved the theorem. □

We observe that up to a constant

$$\begin{aligned} d^J_{\lambda\mu} &= (z_t^2-1)^{(1/2)(M-m)} z_+^m D^M z_-^m D^m P_J(z_t)\,, \\ d^J_{\lambda-\mu} &= (-)^{\lambda+M-m}(z_t^2-1)^{(1/2)(M-m)} z_-^m D^M z_+^m D^m \\ &\quad \times P_J(z_t)\,, \end{aligned} \tag{19}$$

and

Lemma:

$$G(\lambda,\mu) d^J_{\lambda\mu}(z_t) = (z_t^2-1)^{(1/2)(M+m)} D^{M+m}(z_t^2-1)^m D^m \times P_J(z_t)\,. \tag{20}$$

Proof: From Eqs. (17) and (15), if the sign of $\lambda\mu$ is positive,

$$\begin{aligned} &G(\lambda,\mu) d^J_{\lambda\mu}(z_t) \\ &= (z_t^2-1)^{(1/2)(M-m)} z_-^m D^M z_+^m D^{m-M}(z_t^2-1)^{(1/2)(m-M)} \\ &\quad \times (z_t^2-1)^{(1/2)(M-m)} z_+^m D^M z_-^m D^m P_J(z_t) \\ &= (z_t^2-1)^{(1/2)(M-m)}[z_-^m D^M z_+^m D^{m-M} z_+^m D^M z_-^m \\ &\quad \times D^{m-M} D^{M-m}] D^m P_J(z_t)\,. \end{aligned} \tag{21}$$

Thus the assertion of the lemma is equivalent to

$$\begin{aligned} &(z_-^m D^M z_+^m D^{m-M})\ (z_+^m D^M z_-^m D^{m-M}) D^{M-m} \\ &\quad = (z_t^2-1)^m D^{M+m}(z_t^2-1)^m\,. \end{aligned} \tag{22}$$

From Eq. (10) and the deduced commutation properties the left-hand side of Eq. (22) is

$$\prod_{r=1}^{m} \{ [(z_t^2-1)D + (M-m+1)z_t]^2 - (M-m+2r-1)^2 \} D^{M-m}.$$

The $r=1$ term in the product is

$$(z_t^2-1)[(z_t^2-1)D^2 + 2(M-m+2)z_t D + (M-m+1)(M-m+2)] = (z_t^2-1)D^{M-m+2}(z_t^2-1)D^{m-M} \tag{23}$$

(by induction). Thus the left-hand side of Eq. (22) equals

$$\begin{aligned}
&\prod_{2}^{m} \{ [(z_t^2-1)D + (M-m+1)z_t]^2 - (M-m+2r-1)^2 \} \\
&\quad \times (z_t^2-1)D^{M-m+2}(z_t^2-1) \\
&= (z_t^2-1)\Big[\prod_{2}^{m} \{ [(z_t^2-1)D + (M-m+3)z_t]^2 \\
&\quad - (M-m+2r-1)^2 \} D^{M-m+2} \Big](z_t^2-1) \\
&= (z_t^2-1)\Big[\prod_{1}^{m-1} \{ [(z_t^2-1)D + (M-m+3)z_t]^2 \\
&\quad - (M-m+2r+1)^2 \} D^{M-m+2} \Big](z_t^2-1)
\end{aligned} \tag{24}$$

The expression in square brackets (between the z_t^2-1 factors) in Eq. (24) is like the left-hand side of Eq. (22) with

$$M \to M+1, \quad m \to m-1.$$

Thus we can extract another factor of z_t^2-1 at each end and take

$$M \to M+2, \quad m \to m-2.$$

Doing this m times we have m factors of z_t^2-1 at each end, and between them the left-hand side of Eq. (22) with

$$M \to M+m, \quad m \to 0,$$

i.e.

$$D^{M+m}.$$

Thus the left-hand side of Eq. (22) equals

$$(z_t^2-1)^m D^{M+m}(z_t^2-1)^m,$$

the right-hand side of Eq. (22). We have established that Eq. (20) is true if the sign of $\lambda\mu$ is positive. From Eq. (18) this result is independent of the sign of $\lambda\mu$, so the lemma is proven.

C. Development of results

Here we develop the results derived in the previous section into a more useful form. We assume that the sign of $\lambda\mu$ is positive throughout this section.

(i) First (as is usual) we replace the f^t by

$$\begin{aligned}
f^+ &= \tfrac{1}{2}(f^t_{c\bar{a};\bar{d}b} + \eta f^t_{-c-\bar{a};\bar{d}b}), \\
f^- &= \tfrac{1}{2}(f^t_{c\bar{a};\bar{d}b} - \eta f^t_{-c-\bar{a};\bar{d}b}),
\end{aligned} \tag{25}$$

where

$$\eta = \xi(-)^{\lambda+M-m} = \eta_C \eta_{\bar{A}} (-)^{S_C+S_A}(-)^{c-\bar{a}}.$$

We can recast our earlier results in terms of $f^\pm$ rather than f^t. $F(p)$ [Eq. (4)] can be rewritten as

$$F(p) = G_p f^+ + G_{-p} f^- = G_+ f^p + G_- f^{-p}, \tag{26}$$

where

$$G_\pm = \frac{1}{2}\Big(\prod_{r=1}^{m} [(z_t^2-1)D + z_t - M + m - 2r + 1] \pm \prod_{r=1}^{m} [(z_t^2-1)D + z_t + M - m + 2r - 1] \Big) \tag{27}$$

As emphasized previously, all terms in Eq. (26) depend on the helicities $c,\bar{a},\bar{d},b$, though we do not make this explicit.

In general (if $m \neq 0$),

$$G_- \neq 0.$$

In fact G_- measures the extent to which the approximation

$$f^p \simeq \text{naturality } (p)$$

fails to hold.

(ii) In order to see how different contributions appear in $f^\pm$ and $F(p)$ [with a view to interpreting data on $F(p)$] we need some more notation. From Eqs. (1) and (19)

$$\begin{aligned}
f^\pm &= (z_t^2-1)^{(1/2)(M-m)} \sum_{p=\pm} \tfrac{1}{2}(z_+^m D^M z_-^m \pm p z_-^m D^M z_+^m) \\
&\quad \times D^m \sum_J t^J(p) P_J(z_t),
\end{aligned} \tag{28}$$

where

$$t^J(p) = \langle c\bar{a}|T^J(p)|\bar{d}b\rangle (-)^{(1/2)(M-m)}. \tag{29}$$

Thus, if we define

$$\begin{aligned}
A(p) &= (z_t^2-1)^{(1/2)(M-m)} \tfrac{1}{2}(z_+^m D^M z_-^m + z_-^m D^M z_+^m) \\
&\quad \times D^m \sum_J t^J(p) P_J(z_t), \\
B(p) &= (z_t^2-1)^{(1/2)(M-m)} \tfrac{1}{2}(z_+^m D^M z_-^m - z_-^m D^M z_+^m) \\
&\quad \times D^m \sum_J t^J(p) P_J(z_t),
\end{aligned} \tag{30}$$

we can write

$$f^+ = A(N) + B(U), \quad f^- = A(U) + B(N), \tag{31}$$

where

$$A(N) \equiv A(+), \quad B(U) \equiv B(+), \quad \text{etc.}$$

Now we can use our new notation to demonstrate the effect of a naturality $(p)J = \alpha$ (polelike) term. Observe that in the combination

$$z_+^m D^M z_-^m - z_-^m D^M z_+^m$$

the leading-order $O(z_t^{2m-M})$ terms vanish. Thus our $J=\alpha$ term will appear to higher order in $A(p)$ than in $B(p)$ (z_t^α versus $z_t^{\alpha-1}$) from Eq. (30). $A(p)$ will dominate $B(p)$ in the infinite-energy limit.

The equation for $F(p)$ to compare with Eq. (30) is [see Eqs. (1), (27), and (20)]

$$F(p) = (z_t^2-1)^{(1/2)(M+m)} D^{M+m}(z_t^2-1)^m \times \Big(D^m \sum_J t^J(p) P_J(z_t) \Big). \tag{32}$$

Our $J=\alpha$ term contributes to $A(p)$, $B(p)$, and $F(p)$ in the form

$$F(p)=\frac{(\alpha+m)!}{(\alpha-1)!}g(t)z_t^{\alpha+m}[1+O(z_t^{-2})],$$
$$A(p)=\alpha g(t)z_t^{\alpha}[1+O(z_t^{-2})], \tag{33}$$
$$B(p)=\lambda\mu g(t)z_t^{\alpha-1}[1+O(z_t^{-2})].$$

Equations (33), or (32) and (30), are the starting point for the application of our techniques; we refer the reader to Ref. 8 for further comments.

(iii) Here we give some examples of Eq. (27) for the cases $m=0$, 1, and 2.

(a) $m=0$ In this case

$$G_+=1,\quad G_-=0,$$

so the $f^{\pm}$ have definite naturality and $B(p)$ vanishes identically.

(b) $m=1$ Here

$$G_+=(z_t^2-1)D+z_t,\quad G_-=-M.$$

The results for $m=0$ and 1 together cover a wide variety of processes, for instance, all those with at least one vertex:

$$N\to N,\quad \text{e.g., } NN\to NN,$$

or

$$\gamma\to O^-,\quad \text{e.g., } \gamma p\to\pi\Delta,$$

or

$$O^-\to 1^{\pm},\quad \text{(e.g., } \pi N\to\rho\Delta).$$

Consider the process

$$\pi N\to\rho N.$$

Here there are four $m=0$ t-channel amplitudes $f^t_{c\bar a;\bar d b}$:

$$f^t_{0;+-}\equiv A_1\text{-like } Q \text{ numbers only } (UP),$$
$$f^t_{1;++}+f^t_{-1;++}\equiv\rho\text{-like } Q \text{ numbers } (NP),$$
$$f^t_{0;++}\equiv\pi\text{-like } Q \text{ numbers only } (UP),$$
$$f^t_{1;++}-f^t_{-1;++}\equiv\pi\text{-like } Q \text{ numbers } (NP).$$

The two mixed naturality ($m=1$) amplitudes are

$$f^{\pm}=\tfrac{1}{2}(f^t_{1;+-}\pm f^t_{-1;+-}),$$

and for $p=\pm$

$$F(p)=[(z_t^2-1)D+z_t]f^p-f^{-p}$$

isolate the pure naturality (ρ-like and A_1-like) contributions.

(c) $m=2$ For amplitudes with $m=2$

$$F(p)=G_+f^p+G_-f^{-p},$$

where

$$G_+=(z_t^2-1)D^2(z_t^2-1)+M^2,$$
$$G_-=-2M[(z_t^2-1)D+z_t].$$

D. Necessity for m derivatives

Here we will prove that we must use at least m derivatives to obtain pure naturality contributions. Thus the $F(p)$ of Sec. II C is the simplest possible function of amplitudes that does this.

For an $m=1$ pair of amplitudes the proof is simple. A $J=\alpha$ wrong naturality contribution appears in $A(-p)$ and $B(-p)$ [see Sec. II D, Eqs. (30)] like

$$A(-p)=\alpha z_t^{\alpha}[1+O(z_t^{-2})],$$
$$B(-p)=Mz_t^{\alpha-1}[1+O(z_t^{-2})].$$

Working to leading order, we want to cancel these terms in an α-independent linear combination. The only way to give $B(-p)$ an extra factor of α is to take

$$\frac{d}{dz_t}[z_tB(-p)],$$

i.e., we need at least one derivative.

For $m\geqslant 2$ the above argument is no use. Instead we consider eigenstates of $I=(z_t^2-1)D+z_t$, $Y(\lambda)$:

$$IY(\lambda)=\lambda Y(\lambda),\quad \lambda\in\mathbb{C}.$$

These exist:

$$Y\sim z_+^{-(1/2)(1+\lambda)}z_-^{(1/2)(\lambda-1)}.$$

Hence if we have a wrong naturality term

$$f^t_{c\bar a;\bar d b}=\prod_{r=1}^{m}(\lambda+M-m+2r-1)Y,$$

then [using Eq. (4)] in order that this may be eliminated in $F(p)$ we need a term

$$f^t_{-c-\bar a;\bar d b}=\prod_{r=1}^{m}(\lambda-M+m-2r+1)Y$$

up to a phase. The polynomials multiplying Y in the two cases have no common factors. Thus to eliminate the contribution we need to generate a factor

$$\prod_{r=1}^{m}(\lambda-M+m-2r+1)$$

from Y in $f^t_{c\bar a;\bar d b}$. We need to generate a polynomial of order m in λ from Y. The only way we can do this (in a linear fashion) is by differentiating m times (only one power of λ can be gained from one differential). Thus we *need* m derivatives of the amplitudes to isolate pure naturality contributions.

III. SINGLE-DERIVATIVE APPROXIMATION

In Sec. II we saw that (given adequate data) we do not need to make approximations to isolate pure naturality contributions for a bosonic t channel. However, we need m derivatives with respect to energy to do this and the data would have to be incredibly good to allow us to extract second- and higher-order derivatives. If we had a good approximation to exact separation, involving only a single derivative, this would be more useful.

Consider $G_{\pm}$ [Eq. (27)]. Since

$$D\sim O(z_t^{-1}),$$

we can write

$$G_+=(z_tDz_t)^m+O(z_t^{m-2}),$$
$$G_-=-\lambda\mu(z_tDz_t)^{m-1}+O(z_t^{m-3}).$$

Approximating $G_{\pm}$ by their leading orders and cancelling the common factor $(z_tDz_t)^{m-1}$, we get

$$\bar{F}(p) = \bar{G}_p f^+ + \bar{G}_{-p} f^- , \tag{34}$$

where

$$\bar{G}_+ = z_t Dz_t , \quad \bar{G}_- = -\lambda\mu . \tag{35}$$

$\bar{F}(p)$ approximately isolates naturality (p). From Eq. (33) it is easy to see that a $J = \alpha$ naturality $-p$ term appears in (34) to leading order

$$z_t^{\alpha_{-p}-2} .$$

The leading-order terms cancel out. By contrast a $J = \alpha_p$ naturality p term appears in $F(p)$ to higher order,

$$\sim z_t^{\alpha_p+1} .$$

Comparing this with the approximation of neglecting $B(-p)$ in f^p, we see that we are on much safer ground. The "wrong" naturality terms are suppressed by an extra factor of $O(z_t^{-2})$. For comparison with Eqs. (33) we give

$$\bar{F}(p) = \alpha(\alpha+1)\, g(t) z_t^{\alpha+1} [1 + O(z_t^{-2})] . \tag{36}$$

Of course for $m = 1$ this approximation is redundant: we can achieve exact separation with the same information. For $m \geqslant 2$ it should, however, be useful.

IV. SUMMARY

We have succeeded in exactly isolating the definite naturality contributions from helicity amplitudes for a general 2→2 process (with a bosonic t channel). Here we summarize our methods.

We write (for sign of $\lambda\mu$ positive)

$$\begin{aligned} f^+ &= \tfrac{1}{2}(f^t_{c\bar{a};\bar{d}b} + \eta f^t_{-c-\bar{a};\bar{d}b}) = A(N) + B(U) , \\ f^- &= \tfrac{1}{2}(f^t_{c\bar{a};\bar{d}b} - \eta f^t_{-c-\bar{a};\bar{d}b}) = A(U) + B(N) , \end{aligned} \tag{37}$$

where

$$\eta = \eta_C \eta_{\bar{A}} (-)^{S_A + S_C} (-)^{c-\bar{a}} .$$

We define terms of definite naturality

$$F(p) = G_p f^+ + G_{-p} f^- , \tag{38}$$

where

$$G_\pm = \frac{1}{2}\left(\prod_{r=1}^{m} [(z_t^2-1)D + z_t - M + m - 2r + 1] \pm \prod_{r=1}^{m} [(z_t^2-1)D + z_t + M - m + 2r - 1] \right) \tag{39}$$

A $J = \alpha$ term appears in Eqs. (37) and (38) in the form

$$\begin{aligned} A(p) &= \alpha g(t) z_t^\alpha [1 + O(z_t^{-2})] , \\ B(p) &= \lambda\mu g(t) z_t^{\alpha-1} [1 + O(z_t^{-2})] , \\ F(p) &= \frac{(\alpha+m)!}{(\alpha-1)!} g(t) z_t^{\alpha+m} [1 + O(z_t^{-2})] . \end{aligned} \tag{40}$$

These results should not be taken to imply that we are just concerned with a Regge model. They are simply useful to interpret a measured $F(p)$ in terms of the contributions to amplitudes $f^\pm$; they can be considered as a starting point for a more sophisticated analysis.

A simpler (and fairly accurate) approximation to naturality isolation involving only a single derivative has also been developed (see Sec. III).

We conclude that amplitude measurements over a range of energies can now give us, using the above results, a much better idea of the different naturality contributions than was possible previously for *every* 2→2 scattering process (with a bosonic t channel). The results are model independent.

ACKNOWLEDGMENTS

I would like to thank Professor E. Leader for his advice and encouragement in this work and the United Kingdom Science Research Council for financial support.

[1] e.g., $NN \to NN$ amplitude analysis project (Argonne) reported by A. Yokosawa ANL-HEP-PR-CP-47 (1977); and $\pi^- p \to \rho^0 n$ (CERN-Munich collaboration)—a recent reference is H. Becker *et al.*, Nucl. Phys. B **150**, 301 (1979).

[2] R. Stacey, Z. Phys. (1979) (to be published).

[3] R. Stacey, "Techniques for naturality isolation. II. With a fermionic t channel, J. Math. Phys. **21**, (1980).

[4] E. Leader, Phys. Rev. **166**, 1599 (1968).

[5] $\xi = \eta_C \eta_A (-)^{S_A + S_C}$, where the η are intrinsic parities.

[6] The $F(p)$ clearly depend on the helicities. For simplicity we have not made this explicit. The same goes for $A(p)$, $B(p)$, $f^\pm$, and $G_\pm$ in this section.

[7] M. Gell-Mann, M.L. Goldberger, F.E. Low, E. Marx, and F. Zachariasen, Phys. Rev. B **133**, 145 (1964).

[8] R. Stacey, "Naturality isolation for $0^- N \to 1^+ N$" (preprint 1979).

Techniques for naturality isolation. II. With a Fermionic t channel

Richard Stacey[a)]
Department of Physics, Westfield College, University of London, London NW3 7ST, England

(Received 2 October 1979; accepted for publication 24 January 1980)

We develop the mathematical apparatus necessary to isolate pure naturality contributions from the helicity amplitudes for an arbitrary 2→2 scattering process (with a Fermionic t channel). In their exact form they are unlikely to be useful; in fact we prove that an exact isolation is impossible using only a finite combination of the amplitudes and their energy derivatives. However, we derive an approximation series consisting of such finite combinations which converges quickly to an exact separation of naturalities. The first term in the series—the single derivative approximation—is likely to give a good approximation. We comment on the application of our techniques to πN backward scattering.

PACS numbers: 11.50.Nk, 11.80.Fv, 13.75.Gx

I. INTRODUCTION

In this paper we investigate how to isolate definite naturality exchanges for processes with a Fermionic t channel. [Here naturality $p = P(-)^{J-1/2}$] We remind the reader that the general problem (see Ref. 1 and Secs. II A and II C below) reduces to seeking J independent functions $G(\lambda,\mu)$ such that

$$G(\lambda,\mu)d^J_{\lambda\mu}(z_t) = G(\lambda,-\mu)d^J_{\lambda-\mu}(z_t)\,.$$

Here matters are more complicated than they were in the Bosonic case. There we were able to find $G(\lambda,\mu)$ essentially because

$$d^J_{\lambda\mu}(z_t) = N^J_{\lambda\mu}\times O_{\lambda\mu}(z_t,D)\times P_J(z_t)\,, \tag{1}$$

(where $N^J_{\lambda\mu}$ is a constant and $O_{\lambda\mu}$ is J independent) and

$$d^J_{\lambda-\mu}(z_t) = (-)^{J+\lambda}d^J_{\lambda\mu}(-z_t)\,. \tag{2}$$

Since $P_J(z_t)$ has the simple reflection property

$$P_J(-z_t) = (-)^J P_J(z_t)\,, \tag{3}$$

our problem reduces to finding functions $G_\pm$ such that

$$G_\pm O_{\lambda\mu}(z_t,D) = G_\pm O_{\lambda\mu}(-z_t,-D)\,. \tag{4}$$

Equation (4) is already J-independent, and can be solved. This simple behavior seems to be related to the simplicity of the spinless case ($O_{00}=1$).

Now if the t channel is Fermionic, Eq. (2) still holds but (1) is replaced by[2]

$$d^J_{\lambda\mu}(z_t) = N^J_{\lambda\mu}\times O_{\lambda\mu}(z_t,D)\times[P'_{J+1/2} - P'_{J-1/2}]\,. \tag{5}$$

Clearly we do not have a relation like (3) for $(P'_{J+1/2} - P'_{J-1/2})$. The simplest state (with $\lambda=\mu=\frac{1}{2}$) is *fundamentally* more complicated. To effect a separation we clearly need to find J-independent operators $g_\pm(z_t,D)$ such that

$$g_+ P_{J+(1/2)} = g_- P_{J-(1/2)}\,\forall J\,. \tag{6}$$

Once we have $g_\pm$ it is likely that we will be able to isolate naturalities by methods analogous to those developed for the Bosonic case.

Section II is devoted to our general results. First (Sec. II A) we prove that if we restrict $g_\pm$ to the class of polynomials in D (cf. Ref. 1) then we have no relation like (6). It follows simply that *no* combination of amplitudes and a *finite* number of energy derivatives can exactly isolate naturalities when the t channel is Fermionic. All is not lost, however. We discover in Sec. II B a formal method of exact isolation, which involves an infinite number of derivatives. It is unlikely to be useful as it stands, but it can be approximated sensibly in terms of a finite combination of derivatives, the approximation series converging quickly to the exact result (Sec. II C). In Sec. III we develop more fully the $m=\frac{1}{2}$ approximation series (Sec. III A), and consider briefly (Sec. III B) a process for which it is likely to be useful, viz:

$$\pi + p \to p + \pi\,.$$

In Sec. IV we treat the general single derivative approximation. Finally, we summarize our results and draw some conclusions (Sec. V). For the reader who wishes to use our methods, Sec. IIIB and IV are probably the most important.

II. GENERAL RESULTS

A. No separation with finite derivatives

Theorem: There do not exist functions f and g, being polynomials in $D = d/dz_t$ with coefficients functions of z_t, such that $[f\neq 0\neq g]$

$$f\cdot P_{J-1/2} = g\cdot P_{J+1/2}\,\forall J\,. \tag{7}$$

Proof: Suppose such functions did exist. We can take f and g to be polynomials of the same order $2m$ without loss of generality:

$$\begin{aligned} f(z,D) &= f_0 + f_1 D + \cdots + f_{2m}D^{2m}\,,\\ g(z,D) &= g_0 + g_1 D + \cdots + g_{2m}D^{2m}\,, \end{aligned} \tag{8}$$

where f_i and g_i are functions of z_t. By multiplying appropriate factors we can choose f and g such that f_i and g_i have no poles at $z_t=0$ and do not all vanish there. We will obtain a contradiction by showing that f_i and g_i *do* all vanish at $z_t=0$, and so establish the theorem.

We work with $J-\frac{1}{2}\in\mathbb{Z}$ only. We have

$$P_{2n} = C_n\sum_0^n (-)^r \frac{(2n+2r)!}{(2n)!}\,\frac{n!}{(n-r)!}\,\frac{n!}{(n+r)!}\,\frac{z_t^{2r}}{(2r)!}\,, \tag{9}$$

[a)]Present address: Department of Physics and Astronomy, University College London, Bower Street, London WCIE 6BT.

 0022-2488/81/020435-07$1.00

Techniques for naturality isolation. II. With a Fermionic t channel

Richard Stacey[a)]
Department of Physics, Westfield College, University of London, London NW3 7ST, England

(Received 2 October 1979; accepted for publication 24 January 1980)

We develop the mathematical apparatus necessary to isolate pure naturality contributions from the helicity amplitudes for an arbitrary 2→2 scattering process (with a Fermionic t channel). In their exact form they are unlikely to be useful; in fact we prove that an exact isolation is impossible using only a finite combination of the amplitudes and their energy derivatives. However, we derive an approximation series consisting of such finite combinations which converges quickly to an exact separation of naturalities. The first term in the series—the single derivative approximation—is likely to give a good approximation. We comment on the application of our techniques to πN backward scattering.

PACS numbers: 11.50.Nk, 11.80.Fv, 13.75.Gx

I. INTRODUCTION

In this paper we investigate how to isolate definite naturality exchanges for processes with a Fermionic t channel. [Here naturality $p = P(-)^{J-1/2}$] We remind the reader that the general problem (see Ref. 1 and Secs. II A and II C below) reduces to seeking J independent functions $G(\lambda,\mu)$ such that

$$G(\lambda,\mu)d^J_{\lambda\mu}(z_t) = G(\lambda,-\mu)d^J_{\lambda-\mu}(z_t)\,.$$

Here matters are more complicated than they were in the Bosonic case. There we were able to find $G(\lambda,\mu)$ essentially because

$$d^J_{\lambda\mu}(z_t) = N^J_{\lambda\mu}\times O_{\lambda\mu}(z_t,D)\times P_J(z_t)\,, \tag{1}$$

(where $N^J_{\lambda\mu}$ is a constant and $O_{\lambda\mu}$ is J independent) and

$$d^J_{\lambda-\mu}(z_t) = (-)^{J+\lambda}d^J_{\lambda\mu}(-z_t)\,. \tag{2}$$

Since $P_J(z_t)$ has the simple reflection property

$$P_J(-z_t) = (-)^J P_J(z_t)\,, \tag{3}$$

our problem reduces to finding functions $G_\pm$ such that

$$G_+O_{\lambda\mu}(z_t,D) = G_-O_{\lambda\mu}(-z_t,-D)\,. \tag{4}$$

Equation (4) is already J-independent, and can be solved. This simple behavior seems to be related to the simplicity of the spinless case ($O_{00}=1$).

Now if the t channel is Fermionic, Eq. (2) still holds but (1) is replaced by[2]

$$d^J_{\lambda\mu}(z_t) = N^J_{\lambda\mu}\times O_{\lambda\mu}(z_t,D)\times[P'_{J+1/2} - P'_{J-1/2}]\,. \tag{5}$$

Clearly we do not have a relation like (3) for $(P'_{J+1/2} - P'_{J-1/2})$. The simplest state (with $\lambda=\mu=\frac{1}{2}$) is *fundamentally* more complicated. To effect a separation we clearly need to find J-independent operators $g_\pm(z_t,D)$ such that

$$g_+P_{J+(1/2)} = g_-P_{J-(1/2)}\,\forall J\,. \tag{6}$$

Once we have $g_\pm$ it is likely that we will be able to isolate naturalities by methods analogous to those developed for the Bosonic case.

Section II is devoted to our general results. First (Sec. II A) we prove that if we restrict $g_\pm$ to the class of polynomials in D (cf. Ref. 1) then we have no relation like (6). It follows simply that *no* combination of amplitudes and a *finite* number of energy derivatives can exactly isolate naturalities when the t channel is Fermionic. All is not lost, however. We discover in Sec. II B a formal method of exact isolation, which involves an infinite number of derivatives. It is unlikely to be useful as it stands, but it can be approximated sensibly in terms of a finite combination of derivatives, the approximation series converging quickly to the exact result (Sec. II C). In Sec. III we develop more fully the $m=\frac{1}{2}$ approximation series (Sec. III A), and consider briefly (Sec. III B) a process for which it is likely to be useful, viz:

$$\pi + p \to p + \pi\,.$$

In Sec. IV we treat the general single derivative approximation. Finally, we summarize our results and draw some conclusions (Sec. V). For the reader who wishes to use our methods, Sec. IIIB and IV are probably the most important.

II. GENERAL RESULTS

A. No separation with finite derivatives

Theorem: There do not exist functions f and g, being polynomials in $D = d/dz_t$ with coefficients functions of z_t, such that $[f\neq 0\neq g]$

$$fP_{J-1/2} = gP_{J+1/2}\,\forall J\,. \tag{7}$$

Proof: Suppose such functions did exist. We can take f and g to be polynomials of the same order $2m$ without loss of generality:

$$\begin{aligned} f(z,D) &= f_0 + f_1D + \cdots + f_{2m}D^{2m}\,,\\ g(z,D) &= g_0 + g_1D + \cdots + g_{2m}D^{2m}\,, \end{aligned} \tag{8}$$

where f_i and g_i are functions of z_t. By multiplying appropriate factors we can choose f and g such that f_i and g_i have no poles at $z_t=0$ and do not all vanish there. We will obtain a contradiction by showing that f_i and g_i *do* all vanish at $z_t=0$, and so establish the theorem.

We work with $J-\frac{1}{2}\in\mathbb{Z}$ only. We have

$$P_{2n} = C_n\sum_0^n(-)^r\frac{(2n+2r)!}{(2n)!}\frac{n!}{(n-r)!}\frac{n!}{(n+r)!}\frac{z_t^{2r}}{(2r)!}\,, \tag{9}$$

[a)]Present address: Department of Physics and Astronomy, University College London, Bower Street, London WCIE 6BT.

 0022-2488/81/020435-07$1.00

From (20) we get

$$h^2 a_{r-2} = h\,[(r-1)(k+r-1)\,a_{r-1}\,] = (r-1)(k+r+1)\,h a_{r-1} = [(r!/(r-2)!\,](k+r+1)(k+r)a_r\,, \tag{22}$$

and hence (inductively)

$$h^r a_0 = r![(k+2r-1)!/(k+r-1)!\,]\,a_r\,, \tag{23}$$

in an obvious notation. Now, if $r < N$ then the expression

$$\frac{(k+2N-1)!}{(k+2r-1)!}\,\frac{(k+r-1)!}{(k-1)!} \tag{24}$$

represents a polynomial in k. Multiplying both sides of Eq. (23) by this factor gives

$$\frac{1}{r!}\,\frac{(k+2N-1)!}{(k-1)!}\,\frac{(k+r-1)!}{(k+2r-1)!}\,h^r a_0 = \frac{(k+2N-1)!}{(k-1)!}\,a_r\,. \tag{25}$$

Hence

$$\frac{(k+2N-1)!}{(k-1)!}\sum_0^N \frac{(k+r-1)!}{r!(k+2r-1)!}\,h^r a_0 = \frac{(k+2N-1)!}{(k-1)!}\sum_0^N a_r\,. \tag{26}$$

Equation (26) is the expression we were seeking.

(ii) We develop (26) to prove the theorem. Cancelling the common factor in (26) we obtain

$$\sum_0^N a_r = \sum_0^N \frac{(k+r-1)!}{r!(k+2r-1)!}\,h^r a_0\,. \tag{27}$$

The fact that inverse polynomials in $z_t D$ are well defined in their operation is discussed in Sec. II C. Since Σa_r converges sensibly to P_J we can take the limit

$$N \to \infty$$

in (27). Thus

$$P_J(z_t) = H a_0(J)\,, \tag{28}$$

where

$$H = \sum_0^\infty \frac{(k+r-1)!}{r!(k+2r-1)!}\,h^r\,. \tag{29}$$

H maps P_J onto its leading term for all J. This mapping is 1-1 so the inverse is well defined. We prove later (lemma following) that

$$HK = 1\,, \tag{30}$$

where

$$K = \sum_0^\infty \frac{k!}{j!(k+j)!}(-h)^j\,. \tag{31}$$

Hence

$$a_0(J) = KP_J \,\forall J\,. \tag{32}$$

From (18) we see easily that up to a common constant factor

$$a_0(J+\tfrac{1}{2}) = 2J z_t^{J+(1/2)}\,,$$

$$a_0(J-\tfrac{1}{2}) = (J+\tfrac{1}{2}) z_t^{J-(1/2)}\,,$$

and therefore

$$Da_0(J+\tfrac{1}{2}) = 2J(J+\tfrac{1}{2}) z_t^{J+(1/2)} = (2zD+1)a_0(J-\tfrac{1}{2})\,,$$

i.e.,

$$\tfrac{1}{2}Da_0(J+\tfrac{1}{2}) = ka_0(J-\tfrac{1}{2})\,. \tag{33}$$

Combining (33) and (32) we have

$$\tfrac{1}{2}DKP_{J+(1/2)} = kKP_{J+(1/2)}\,. \tag{34}$$

This is the assertion of the theorem, with (31) giving (17) with the values of k and h defined after Eq. (19).

There is one extra point to note: if $J \in \mathbb{Z} + \frac{1}{2}$ then a_r [Eq. (18)] can be infinite. If we exclude these points from the definitions of K we still have all the results in Sec. II C since there K is taken to act on a sum of P_N (N integral) before we generalize J.

We conclude this section by proving the result quoted in Eq. (30).

Lemma: Defining

$$H = \sum_0^\infty \frac{(k+r-1)!}{r!(k+2r-1)!}\,h^r$$

and

$$K = \sum_0^\infty (-)^r \frac{k!}{r!(k+r)!}\,h^r\,,$$

with

$$hk = (k+2)h$$

we have $KH = 1$.

Proof:

$$KH = \sum_{r,r'} (-)^r \frac{k!}{r!(k+r)!}\,h^r \frac{(k+r'-1)!}{r'!(k+2r'-1)!}\,h^{r'}\,.$$

Using the commutation properties

$$KH = \sum_{r,r'} (-)^r \frac{k!}{r!(k+r)!}\,\frac{(k+2r+r'-1)!}{r'!(k+2r+2r'-1)!}\,h^{r+r'} = \sum_{m=0}^\infty \left(\sum_{r=0}^m (-)^r \frac{k!(k+m+r-1)!}{r!(m-r)!(k+r)!(k+2m-1)!} \right) h^m\,. \tag{35}$$

The coefficient of h^0 is

$$k!(k-1)!/k!(k-1)! = 1\,.$$

The result of the lemma is now equivalent to proving that, for $m \geqslant 1$,

$$\sum_0^m (-)^r \frac{k!(k+m+r-1)!}{r!(m-r)!(k+r)!(k+2m-1)} = 0\,. \tag{36}$$

The expression on the left-hand side of (36) equals

$$\frac{k!}{(k+2m-1)!m!} \sum_{r=0}^m (-)^r \binom{m}{r} \frac{(k+m+r-1)!}{(k+r)!}\,. \tag{37}$$

Now if $k \in \mathbb{Z}^+$, then

$$D^{m-1}[x^{k+m-1}(1-x)^m] = D^{m-1}\left[\sum_{r=0}^m (-)^r \binom{m}{r} x^{r+k+m-1} \right] = \sum_{r=0}^m (-)^r \binom{m}{r} \frac{(k+m+r-1)!}{(k+r)!}\,x^{k+r}\,. \tag{38}$$

Putting $x = 1$ in (38), the left-hand side must be zero [at least one factor $(1-x)$], so

$$\sum_{r=0}^m (-)^r \binom{m}{r} \frac{(k+m+r-1)!}{(k+r)!} = 0 \tag{39}$$

for $k \in \mathbb{Z}^+$. Now (39) is a polynomial in k with infinitely many roots, so it must vanish identically, whatever k is. Thus (37) vanishes and (36) holds.

We have our result.

C. Formal exact separation (2)

In this section we explore the consequences of the theorem proved in Sec. II B (Eqs. (16) and (17)]. First we will justify some of our earlier analysis.

The presence in K [Eq. (17)] of inverse polynomials in $z_t D$ means that we cannot evaluate directly the result of operating it on an arbitrary function. Its operation is, nonetheless, perfectly well defined. For a wide class of amplitudes $f(z_t)$ we can write

$$f(z_t) = \int_{-\infty}^{\alpha_0} z_t^{\alpha} \tilde{f}(\alpha)\, d\alpha \,. \tag{40}$$

We have an upper bound α_0 because of the Froissart bound; our α_0 is chosen to be the lowest nontrivial bound. Now the operation of K on z_t^{α} is well defined for $|z_t| > 1$:

$$Kz_t^{\alpha} = \sum_{r=0}^{\infty} \frac{(\alpha - 2r + \frac{1}{2})!}{r!(\alpha - r + \frac{1}{2})!} \frac{\alpha!}{(\alpha - 2r)!} \frac{1}{4^r} z_t^{\alpha - 2r} \,. \tag{41}$$

The series on the right-hand side of (41) has good (uniform) convergence properties since as $r \to \infty$ the ratio of successive terms tends to

$$-\tfrac{1}{4} z_t^{-2} \,.$$

Thus we can commute the sum in (41) with the integral in (40), giving, with a little manipulation,

$$Kf(z_t) = z_t^{\alpha_0} \sum_{r=0}^{\infty} \left(\frac{1}{z_t}\right)^{2r} \int_{-\infty}^{0} z_t^{\alpha} \tilde{f}_r(\alpha)\, d\alpha \tag{42}$$

where

$$\tilde{f}_r(\alpha) = \frac{(\alpha + \alpha_0 - 4r + \frac{1}{2})!(\alpha + \alpha_0 - 2r)!}{r!(\alpha + \alpha_0 - 3r + \frac{1}{2})!(\alpha + \alpha_0 - 4r)!} \times \frac{1}{4^r} \tilde{f}(\alpha + \alpha_0 - 2r) \,.$$

We can draw two important conclusions from Eq. (42):

1. $Kf(z_t)$ is well defined: it exists.

2. The right-hand side of Eq. (42) is a good asymptotic expansion, effectively in z_t^{-2}. The coefficients $[\int_{-\infty}^{0} z_t^{\alpha} \tilde{f}_r(\alpha)\, d\alpha]$ have some z_t dependence, but as $z_t \to \infty$

$$|z_t|^{\epsilon} > O\left[\int_{-\infty}^{0} z_t^{\alpha} \tilde{f}_r(\alpha)\, d\alpha\right] > |z_t|^{-\epsilon} \;\forall\; \epsilon > 0 \,.$$

At low energies there can be more variation, but any energy-dependence in these factors should be totally swamped by the z_t^{-2} terms. The first few terms of Eq. (42) should give a good approximation to $Kf(z_t)$ in the usual "high energy" region (e.g. $P_L \geqslant 3$). This approximation is in fact obtained by taking the first few terms of K itself, as given by Eq. (17). "high energy" region (e.g. $P_L \geqslant 3$). This approximation is in fact obtained by taking the first few terms of K itself, as given by Eq. (17).

Conclusion (1) allows us to achieve a formal separation of naturalities; (2) allows us to approximate it more usefully, to arbitrary accuracy. The rest of this section is taken up with doing just that.

We have the expansion of $f^t_{c\bar{a};\bar{d}b}$ and $f^t_{-c\,\bar{a};\bar{d}b}$ in partial wave series:

$$f^t_{c\bar{a};\bar{d}b} = \sum_J [\langle c\bar{a}|T^J(+)|\bar{d}b\rangle + \langle c\bar{a}|T^J(-)|\bar{d}b\rangle] \times d^J_{\lambda\mu}(z_t) \,, \tag{43}$$

$$f_{-c-\bar{a};\bar{d}b} = \xi \sum_J \{\langle ca|T^J(+)|\bar{d}b\rangle - \langle c\bar{a}|T^J(-)|\bar{d}b\rangle\} d^J_{\lambda-\mu}(z_t) \,,$$

where $\xi = \eta_C \eta_{\bar{A}} (-)^{S_A + S_C - 1/2}$ [cf. Ref. 1, Eq. (1)]. Now up to a common constant [sign $(\lambda\mu) = +$][2]:

$$d^J_{\lambda\mu} = (z_t^2 - 1)^{1/2(M-m)}(z_t + 1)^m (z_t - 1)^{m - (1/2)} D^{m + (1/2)} \times D^{M - (1/2)}(P_{J + (1/2)} - P_{J - (1/2)}) \,, \tag{44}$$

$$(-)^{\lambda + M - m} d^J_{\lambda-\mu} = (z_t^2 - 1)^{(1/2)(M-m)}(z_t - 1)^m \times D^{M - (1/2)}(z_t + 1)^{m - (1/2)}(P_{J + (1/2)} + P_{J + (1/2)})$$

(cf. Ref. 1 for notation). Manipulating formally

$$D^{m + (1/2)} P_{J \pm (1/2)} = \tfrac{1}{2}[(z_t + 1)^{(1/2) - m} D^{(1/2) - M}(z_t - 1)^{-m} \times (z_t^2 - 1)^{(1/2)(m - M)}(-)^{\lambda + M - m} d^J_{\lambda-\mu} \pm (z_t - 1)^{(1/2) - m} D^{(1/2) - M} \times (z_t + 1)^{-m}(z_t^2 - 1)^{(1/2)(m - M)} d^J_{\lambda\mu}] \,. \tag{45}$$

The inverse powers of D are, again, not ill defined; but in the expressions we wish to use we will eliminate them (cf. Sec. III).

From Eqs. (16) and (17) we deduce that

$$\tfrac{1}{2} DK'(D^{m + (1/2)} P_{J + (1/2)}) = (z_t D + m + 1) K'(D^{m + (1/2)} P_{J - (1/2)}) \,, \tag{46}$$

where

$$K' = \sum_0^{\infty} \frac{(z_t D + m + 1)!}{r!(z_t D + m + r + 1)!} (\tfrac{1}{4} D^2)^r \,. \tag{47}$$

Hence, defining

$$G(\lambda,\mu) = (z_t D + \tfrac{1}{2} D + m + 1) K'(z_t - 1)^{(1/2) - m} D^{(1/2) - M} \times (z_t + 1)^{-m}(z_t^2 - 1)^{1/2(m - M)} \,, \tag{48}$$

$$G(\lambda, -\mu) = (z_t D - \tfrac{1}{2} D + m + 1) K'(z_t + 1)^{(1/2) - m} D^{(1/2) - M} \times (z_t - 1)^{-m}(z_t^2 - 1)^{1/2(m - M)} (-)^{\lambda + M - m} \,,$$

we can write [combining (45) and (46)]:

$$G(\lambda,\mu) d^J_{\lambda\mu}(z_t) = G(\lambda, -\mu) d^J_{\lambda-\mu}(z_t) \,. \tag{49}$$

The reader should compare this with Sec. II A of Ref. 1; we now proceed in a similar manner.

Using Eqs. (49) and (43) we can define a pure naturality p amplitude

$$F(p) = \tfrac{1}{2}[G(\lambda,\mu) f^t_{c\bar{a};\bar{d}b} + p\xi G(\lambda, -\mu) f^t_{-c-\bar{a};\bar{d}b}] = \sum_J \langle c\bar{a}|T^J(p)|\bar{d}b\rangle G(\lambda,\mu) d^J_{\lambda\mu}(z_t) \,. \tag{50}$$

We now go on to define $f^{\pm}$, $A(p)$, and $B(p)$ exactly as in Sec. II B of Ref. 1. In brief,

$$f^{\pm} = \tfrac{1}{2}[f^t_{c\bar{a};\bar{d}b} \pm \eta f^t_{-c-\bar{a};\bar{d}b}] = A\begin{pmatrix} N \\ U \end{pmatrix} + B\begin{pmatrix} U \\ N \end{pmatrix} , \tag{51}$$

where now

$$\eta = \eta_C \eta_{\bar{A}} (-)^{S_t + S_{\bar{t}} - (1/2)} (-)^{c - \bar{a}} \text{sign}(\lambda\mu) .$$

In terms of $f^{\pm}$

$$F(p) = G_p f^+ + G_{-p} f^- , \tag{52}$$

where

$$G_{\pm} = G(\lambda,\mu) \pm (-)^{m - \lambda - M} G(\lambda, -\mu) . \tag{53}$$

$A\,J = \alpha$ polelike term contributes to (51) and (52) as

$$\begin{aligned} A(p) &= \alpha g(t) z_t^{\alpha} [1 + O(z_t^{-2})] , \\ B(p) &= \lambda\mu g(t) z_t^{\alpha - 1} [1 + O(z_t^{-2})] , \\ F(p) &= \alpha(\alpha + 1)[(\alpha - M)!/(\alpha - \tfrac{1}{2})!] g(t) z_t^{\alpha - m} \\ &\quad \times [1 + O(z_t^{-2})] . \end{aligned} \tag{54}$$

This ends our general analysis. In Sec. III we develop in detail the $m = \frac{1}{2}$ approximation series, an example of what can be done in general (though not easily in a general manner).

III. APPROXIMATIONS FOR $m = \frac{1}{2}$

A. The m = ½ series

In line with point (2) following Eq. (42) we here develop a useful approximation series for the case

$$m = \tfrac{1}{2} .$$

Let us define modified amplitudes:

$$\begin{aligned} \bar{f}^t_{c\bar{a};\bar{d}b} &= (z_t + 1)^{-1/2} (z_t^2 - 1)^{1/2((1/2) - M)} f^t_{c\bar{a};\bar{d}b} , \\ \bar{f}^t_{-c-\bar{a};\bar{d}b} &= (z_t - 1)^{-1/2} (z_t^2 - 1)^{1/2((1/2) - M)} f^t_{-c-\bar{a};\bar{d}b} , \end{aligned} \tag{55}$$

where sign $(\lambda\mu) = +$ as usual. We modify the $G(\lambda,\mu)$ and $F(p)$ of Eqs. (48) and (50) by taking

$$\begin{aligned} G'(\lambda,\mu) &= D^{M - (1/2)} G(\lambda,\mu) , \\ F'(p) &= D^{M - (1/2)} F(p) . \end{aligned}$$

Equation (50) becomes

$$\begin{aligned} F'(p) = \tfrac{1}{2}[&(z_t D + \tfrac{1}{2}D + M + 1) K'' \bar{f}^t_{c\bar{a};\bar{d}b} \\ &+ p\eta (z_t D - \tfrac{1}{2}D + M + 1) K'' \bar{f}^t_{-c-\bar{a};\bar{d}b}] , \end{aligned} \tag{56}$$

where

$$\eta = \eta_C \eta_{\bar{A}} (-)^{S_t + S_{\bar{t}} - (1/2)} (-)^{c - \bar{a}}$$

and

$$K'' = \sum_{r=0}^{\infty} \frac{(z_t D + M + 1)!}{r!(z_t D + M + r + 1)!} (\tfrac{1}{4}D^2)^r . \tag{57}$$

Denoting by $F'_N(p)$ the first N terms in $F'(p)$ we can define

$$\begin{aligned} F''_N(p) &\equiv \frac{(z_t D + M + N + 1)!}{(z_t D + M + 1)!} F'_N(p) \\ &= (z_t D + M + 1) \sum_{r=0}^{N-1} \frac{(z_t D + M + N + 1)!}{r!(z_t D + M + r + 1)!} \\ &\quad \times (\tfrac{1}{4}D^2)^r \tfrac{1}{2}[\bar{f}^t_{c\bar{a}} + p\eta \bar{f}^t_{-c-\bar{a}}] \\ &\quad + \tfrac{1}{2}D(z_t D + M + 1) \sum_{r=0}^{N-1} \frac{(z_t D + M + N)!}{r!(z_t D + M + r + 1)!} \\ &\quad \times (\tfrac{1}{4}D^2)^r \tfrac{1}{2}[\bar{f}^t_{c\bar{a}} - p\eta \bar{f}^t_{-c-\bar{a}}] . \end{aligned} \tag{58}$$

For all N $F''_N(p)$ is a well-defined expression consisting of polynomials in D acting on the $\bar{f}^t$. The discussion following Eq. (42) establishes that $F''_N(p)$ converges to a pure naturality p amplitude as $N \to \infty$. The naturality $(-p)$ terms at each stage are suppressed relative to the naturality (p) terms by a factor of order

$$1/z_t^{2N+1} .$$

Now we will consider some examples of Eq. (58), for the cases $N = 1$ and 2.

1. $N = 1$.

In this (the simplest) case we have

$$\begin{aligned} &F''_1(p) \\ &= (z_t D + M + 2)[(z_t D + M + 1)\tfrac{1}{2}(\bar{f}^t_{c\bar{a}} + p\eta \bar{f}^t_{-c-\bar{a}}) \\ &\quad + \tfrac{1}{2}D \cdot \tfrac{1}{2}(\bar{f}^t_{c\bar{a}} - p\eta \bar{f}^t_{-c-\bar{a}}] . \end{aligned} \tag{59}$$

Apart from the irrelevant overall factor this is just Eq. (56) with $K'' \simeq 1$. We can rewrite $F''_1(p)$ in terms of the f^t [hence the $f^{\pm}$ of Eq. (51)] by taking the same order approximation we are using already in Eq. (55). Thus

$$\begin{aligned} \bar{f}^t_{c\bar{a};\bar{d}b} &\simeq z_t^{-M-1} (z_t - \tfrac{1}{2}) f^t_{c\bar{a};\bar{d}b} , \\ \bar{f}^t_{-c-\bar{a};\bar{d}b} &\simeq z_t^{-M-1} (z_t + \tfrac{1}{2}) f^t_{-c-\bar{a};\bar{d}b} . \end{aligned}$$

Continuing to retain only the leading two orders we get

$$F''_1(p) = z_t^{-M-1} (z_t D + 1)[(z_t D z_t) f^p - \lambda\mu f^{-p}] .$$

Apart from irrelevant factors this is just the single derivative approximation of Sec. IV for this special case.

2. $N = 2$.

This case is less trivial. From Eq. (58)

$$\begin{aligned} &F''_2(p) \\ &= (z_t D + M + 1)(z_t D + M + 3)[(z_t D + M + 2) + \tfrac{1}{4}D^2] \\ &\quad \times \tfrac{1}{2}[\bar{f}^t_{c\bar{a}} + p\eta \bar{f}^t_{-c-\bar{a}}] + \tfrac{1}{2}D(z_t D + M + 1) \\ &\quad \times [(z_t D + M + 2) + \tfrac{1}{4}D^2] \tfrac{1}{2}[\bar{f}^t_{c\bar{a}} - p\eta \bar{f}^t_{-c-\bar{a}}] . \end{aligned} \tag{60}$$

This last expression can of course be written in terms of $f^{\pm}$, just as $F''_1(p)$. The approximation is obtained by taking the leading four orders in z_t.

In general we need $2N$ energy-derivatives to evaluate $F''_N(p)$.

B. πN backward scattering

Here our main idea is just to observe that πN backward scattering is a process where our methods should be useful. The process is, in the s channel,

$$\pi + N \to N + \pi ,$$

with t channel

$$\bar{N} + \pi \to \pi + \bar{N} .$$

There are only two independent helicity amplitudes. If we write the t channel amplitudes $f^t_{c';b'}$ and the s channel amplitudes $f^s_{c;b}$, then

$$\begin{aligned} f^{s,t}_{+;+} &= -f^{s,t}_{-;-} \equiv N^{s,t} , \\ f^{s,t}_{+;-} &= -f^{s,t}_{-;+} \equiv F^{s,t} , \end{aligned} \tag{61}$$

in an obvious notation. We have one crossing angle χ, and

$$\begin{pmatrix} N^t \\ F^t \end{pmatrix} = \begin{bmatrix} \cos\chi & -\sin\chi \\ \sin\chi & \cos\chi \end{bmatrix} \cdot \begin{pmatrix} N^s \\ F^s \end{pmatrix}, \tag{62}$$

where

$$\cos\chi = \{(s+m^2-\mu^2)(t+m^2-\mu^2) - 4m^2(m^2-\mu^2)\} \times \{[s-(m-\mu)^2][s-(m+\mu)^2][t-(m-\mu)^2] \times [t-(m-\mu)^2][m+\mu)^2]\}^{-1/2}$$

$(m = m_N, \mu = m_\pi)$ and

$$\sin\chi = 2m[stu + (s+t)(m^2-\mu^2)^2 + 2(\mu^4 - m^4) \times (m^2-\mu^2)]^{1/2}\{[s-(m-\mu)^2][s-(m+\mu)^2] \times [t-(m-\mu)^2][t-(m-\mu)^2][m+\mu)^2]\}^{-1/2}.$$

Our amplitudes $f^{\pm}$ [Eq. (51)] are given by

$$f^+ = F^t + iN^t = e^{-i\chi}[F^s + iN^s], \qquad f^- = F^t - iN^t = e^{i\chi}[F^s - iN^s]. \tag{63}$$

The single derivative approximate separation scheme defined in the last section is

$$\bar{F}(p) = z_t \frac{d}{dz_t}(z_t f^p) - \tfrac{1}{4} f^{-p}. \tag{64}$$

The contributions of a $J = \alpha$ polelike term to the different amplitudes are

$$A(p) = \alpha z_t^{\alpha}[1 + O(z_t^{-2})], \quad B(p) = \tfrac{1}{4} z_t^{\alpha-1}[1 + O(z_t^{-2})], \quad F(p) = \alpha(\alpha+1) z_t^{\alpha+1}[1 + O(z_t^{-2})]. \tag{65}$$

This section [Eqs. (61)–(65)] can be taken by itself and used to get a good estimate of the pure naturality contributions. The $B(p)$ factors are not so important here as they are in any ($m \neq 0$) Bosonic t-channel process, due to the extra factor of $\frac{1}{4}(\equiv\lambda\mu)$. We still cannot rely on $B(p)$ being small for moderate energies. In our NN analysis $B(p)$ was generally comparable to $A(-p)$; a factor of $\frac{1}{4}$ will not make it negligible. Our single-derivative approximation should be very reliable.

IV. THE SINGLE-DERIVATIVE APPROXIMATION

In this section we derive the single-derivative approximation to exact naturality separation for a *general* Fermionic t channel. We have already done this in Sec. III A for the case $m = \frac{1}{2}$.

In fact the single-derivative approximation (≡SDA) is exactly the same here as in the Bosonic case (Ref. 1, Sec. III). There are two reasons:

1. The leading order ($J = \alpha$) expansions for $A(p)$ and $B(p)$ are the same in the Bosonic and Fermionic cases [Eq. (54) and Ref. 1, Eq. (31)].

2. The single derivative approximation simply cancels the leading order wrong-naturality terms.

These comments do not mean we are just considering a pole model, as might easily be thought. Any amplitude can be considered as being as convolution of polelike terms, or having the form of Eq. (40). If the leading naturality $(-p)$ term $[A(-p)]$ has the form

$$A(-p) = \int_{-\infty}^{\alpha_0} z_t^{\alpha} \tilde{f}(\alpha)\, d\alpha, \tag{66}$$

then the naturality $(-p)$ contribution to $\bar{F}(p)$ is

$$\bar{F}(p)[-p] = \int_{-\infty}^{\alpha_0 - 2} z_t^{\alpha} \tilde{\tilde{f}}(\alpha)\, d\alpha = \frac{1}{z_t^2}\left[\int_{-\infty}^{\alpha_0} z_t^{\alpha} \tilde{\tilde{f}}(\alpha - 2)\, d\alpha\right], \tag{67}$$

i.e., relatively smaller by a factor of $O(z_t^{-2})$ than $B(-p)$ is in f^p. This point is outside the main line of our argument, but it is important in order to avoid any misunderstanding.

Anyway, the SDA is the same for any 2→2 process, so that we have already considered it fully in Ref. 1. Nonetheless we will now show how to derive the technique from the general methods of Sec. II. This is a useful check on the validity of our manipulations.

The simplest approximation to $F(p)$ [Eq. (50)] is obtained by taking $K' \sim 1$ in Eq. (48):

$$\begin{aligned} G(\lambda,\mu) &= (z_t D + \tfrac{1}{2}D + m + 1)K'(z_t - 1)^{(1/2)-m} D^{(1/2)-M}(z_t + 1)^{-m}(z_t^2 - 1)^{(1/2)(m-M)} \\ &= (z_t D + \tfrac{1}{2}D + m + 1)(z_t^{(1/2)-m} + (m - \tfrac{1}{2}) z_t^{-(1/2)-m}) D^{(1/2)-M}(z_t^m - m z_t^{m-1}) z_t^{m-M}[1 + O(z_t^{-2})] \\ &= z_t^{(1/2)-m} D^{(1/2)-M} z_t^{-M-1}(z_t D z_t - mM)[1 + O(z_t^{-2})]. \end{aligned} \tag{68}$$

From the symmetry under $z_t \to -z_t$

$$G(\lambda, -\mu) \approx (-)^{\lambda+M-m} z_t^{(1/2)-m} D^{(1/2)-M} z_t^{-M-1}(z_t D z_t + mM). \tag{69}$$

We can define $F(p)$ by inserting these approximations into $G(\lambda, \pm\mu)$ in Eq. (50), and then we can use this $F(p)$ to define

$$\bar{F}(p) = z_t^{M+1} D^{M-(1/2)} z_t^{m-(1/2)} F(p).$$

Hence

$$\begin{aligned} \bar{F}(p) &= z_t D z_t \tfrac{1}{2}(f^t_{c\bar{a};\bar{d}b} + p\eta f^t{}_{-c-\bar{a};\bar{d}b}) - mM\tfrac{1}{2}(f^t_{c\bar{a};\bar{d}b} - p\eta f^t{}_{-c-\bar{a};\bar{d}b}) \\ &= z_t \frac{d}{dz_t}[z_t f^p] - \lambda\mu f^{-p}. \end{aligned} \tag{70}$$

This is just Eq. (34) of Ref. 1 again. As there, we can write as the contribution of a $J = \alpha$ naturality p term:

$$A(p) = \alpha z_t^{\alpha} g(t)[1 + O(z_t^{-2})], \quad B(p) = \lambda\mu z_t^{\alpha-1} g(t)[1 + O(z_t^{-2})], \quad \bar{F}(p) = \alpha(\alpha+1) z_t^{\alpha+1} g(t)[1 + O(z_t^{-2})]. \tag{71}$$

V. SUMMARY

In this paper we have done a great deal of mathematical work, leaving us with some simple techniques for extracting important information on pure naturality exchanges from measured amplitudes, the techniques being set in a well-understood mathematical framework.

We have a well-defined approximation series consisting of sums of amplitudes and a finite number of their energy derivatives, giving increasingly good approximations to pure naturality contributions. At every stage the error is reduced by a factor of order z_t^{-2}. The first term (SDA: Sec. IV) is likely to give a good approximation, the "wrong" naturality terms being suppressed by a factor of order z_t^{-3} relative to the "right" naturality terms. Better approximations are easily obtained (Sec. III) but are less simple.

The limit of the series exists (Sec. II) but in general we will not know how to express the amplitudes in a form where we can use our exact methods. No simpler method involving just a finite number of derivatives can achieve exact naturality isolation.

ACKNOWLEDGMENTS

I would like to thank Professor E. Leader for his encouragement in the course of this work, and the United Kingdom Science Research Council for financial assistance.

[1] R. Stacey, *Techniques for Naturality Isolation*: I. *With a Bosonic t channel* J. Math. Phys. **22**, 430 (1981), Companion paper.
[2] M. Gell-Mann, M.L. Goldberger, F.E. Low, E. Marx and F. Zachariasen, Phys. Rev. B **133**, 145 (1964).