Anomaly coefficients: Their calculation and congruences

H. W. Braden[a)]

Institute of Field Physics, Department of Physics and Astronomy and Department of Mathematics, University of North Carolina, Chapel Hill, North Carolina 27514

(Received 5 November 1987; accepted for publication 18 November 1987)

A new method for the calculation of anomaly coefficients is presented. For su(n) some explicit and general expressions are given for these. In particular, certain congruences are discovered and investigated among the leading anomaly coefficients. As an application of these congruences, the absence of global six-dimensional gauge anomalies is shown.

I. INTRODUCTION

A recurring problem in mathematical physics is the calculation of "anomaly coefficients." At its simplest we have a Weyl-invariant polynomial and we wish to express this in terms of a given basis of Weyl-invariant polynomials: the coefficients of this expansion are the anomaly coefficients.

The name arises as follows. Suppose we have a gauge theory with fermionic content in a representation λ of the gauge group G. That is, we have a vector bundle ξ with fiber V upon which G acts via the representation λ. By a now standard analysis,[1] the anomalies of this theory are associated with a secondary invariant of the principal G bundle P associated with G. Because the de Rham ring of P may be expressed in terms of G-invariant polynomials (whose coefficients are forms on the base space) these secondary invariants are also expressible in terms of G-invariant polynomials upon the use of the Chern–Weil homomorphism. The appropriate polynomial for a $2n$-dimensional theory is determined via the descent equations (or transgression) from $\mathrm{Tr}_\lambda\, F^{n+1}$, where the trace is over the representation λ of G and F is a curvature form on P. [This polynomial may also be viewed as resulting from the restriction of the GL(V) theory to G.] The work of Borel–Hirzebruch[2] shows that we can reduce our considerations to a maximal torus H of G and we then wish to express the Weyl-invariant polynomial $\mathrm{Tr}_\lambda\, F^{n+1}$ in terms of some given basis. When this basis corresponds to certain homotopy generators we actually get an expansion with integral coefficients, the anomaly coefficients. (The choice of basis will be elaborated upon in due course.)

In this paper we are going to present an alternate technique for calculating these anomaly coefficients, which, though in principle are known, are often tedious to obtain. The merit of this approach is that with a tractable method of computation we are able to discover a variety of congruences among these coefficients. One application we shall put to these congruences is to show the absence of six-dimensional global gauge anomalies.

An outline of the paper is as follows. In Sec. II we will describe our method of calculating these anomaly coefficients. After a preliminary section establishing notation and some elementary results we then describe both our and other approaches to these calculations. Although the techniques we describe in Sec. II are general for any group G we will focus our attention in Sec. III to su(n). Here again we give some preliminary remarks on A_{n-1} and symmetric polynomials before finding a general expression (3.37) for the anomaly coefficients in Sec. III C. In Sec. III D we give a variety of closed forms for the leading anomaly coefficient (which is used most often in physics). We give some examples in Sec. III E before proceeding to investigate the congruences of these leading anomaly coefficients in Sec. IV. After a discussion of su(2) in Sec. IV B some general results are presented for su(n) in Sec. IV C. A few other congruences are also noted in Sec. IV C that will be of use in Sec. V where we make some applications of these results. In particular we show the absence of six-dimensional global gauge anomalies. Finally in Sec. VI a review and some discussion are presented.

II. THE GENERAL CALCULATION OF ANOMALY COEFFICIENTS

In this section we will describe our approach to calculating anomaly coefficients, comparing this with other techniques. The specific implementation of this method and some examples will be given here; applications will be given in later sections.

A. Preliminaries

First we give preliminary notation and results. For this we will largely follow Ref. 3. Consider L a finite-dimensional semisimple Lie algebra with Cartan subalgebra H over a field of characteristic zero. Let Φ be a root system of L with $\Phi\subset E$, E a Euclidean space of dimension $l=\mathrm{rank}\, L$. Let Δ be a base of Φ and fix an ordering of the simple roots $\{\alpha_1,\ldots,\alpha_l\}$. Denote by Λ the weight lattice,

$$\Lambda=\{\lambda\in E\,|\,2(\lambda,\alpha)/(\alpha,\alpha)\in Z,\ \alpha\in\Phi\}. \qquad (2.1)$$

Let $\Lambda_r\subset\Lambda$ be the root lattice and W the Weyl group. A weight is called dominant (with respect to a fixed basis Δ) if all the integers $2(\lambda,\alpha)/(\alpha,\alpha)$ $(\alpha\in\Delta)$, are non-negative. It is called strongly dominant if these integers are positive. Denote by Λ^+ the set of dominant weights. We let $\lambda_1,\ldots,\lambda_l$ be the dual basis of Δ relative to the inner product on E: $2(\lambda_i,\alpha_j)/(\alpha_j,\alpha_j)=\delta_{ij}$. The λ_i are known as the fundamental dominant weights (relative to Δ) and form a basis of Λ. We set

[a)] Address after 1 January 1988: Department of Mathematical Sciences, University of Durham, Durham DH13LE, England.

$$\delta = \sum_{i=1}^{l} \lambda_i = \frac{1}{2} \sum_{\alpha > 0} \alpha. \tag{2.2}$$

Now let V be an L-module and denote by $\pi(V)$ the set of all of its weights. We will be working with finite-dimensional modules $V = V(\lambda)$ of highest weight λ; in this case we will also denote $\pi(V) = \pi(\lambda)$. Standard theory tells us that the map $\lambda \to V(\lambda)$ induces a one-to-one correspondence between Λ^+ and the isomorphism classes of finite-dimensional irreducible L-modules. Set $\pi^+(\lambda) = \pi(\lambda) \cap \Lambda^+$, $m_\lambda(\mu)$ the multiplicity of the weight μ in $V(\lambda)$, and denote by Stab λ the stabilizer of the weight λ in W:

$$\text{Stab}\,\lambda = \{\sigma \in \mathrm{W},\ \sigma(\lambda) = \lambda\}. \tag{2.3}$$

The stabilizer of a dominant weight (and thus any weight) is easily determined. Let

$$\lambda = \sum_{j=1}^{l} m_j \lambda_j \in \Lambda^+$$

and set $J(\lambda) = \{j | m_j = 0\}$. If we denote by r_j the Weyl reflection associated with α_j, then Stab λ is the group generated by $\{r_j,\ j \in J\}$. The groups Stab λ are again Weyl groups and are known as parabolic subgroups.[4] Clearly Stab λ is the Weyl group associated with the Dynkin diagram obtained from that of L by deleting all nodes and incident edges not indexed by the set J.

Finally we need some results on various algebras constructed from L and H. Recall that for a finite-dimensional vector space V we may define the tensor algebra $T(V)$ on V. The symmetric algebra $S(V)$ is formed from the (two-sided) ideal I in $T(V)$ generated by all $x \otimes y - y \otimes x$ $(x,y \in V)$, $S(V) = T(V)/I$. The algebra $P(V)$ of polynomials on V is given by $P(V) = S(V^*)$. Further, when the vector space has the additional structure of a Lie algebra (not necessarily finite-dimensional) we may construct the universal enveloping algebra $U(L)$. Here $U(L) = T(L)/J$ and the (two-sided) ideal J in $T(V)$ is generated by all $x \otimes y - y \otimes x - [x,y]$ $(x,y \in L)$. Note when L is Abelian the symmetric and universal enveloping algebras coincide, thus $S(H) = U(H)$. Let π be the canonical algebra homomorphism π: $T(L) \to U(L)$, and set $G = \text{Inn}\, L$, the inner automorphisms of L. Now G acts on $T(L)$ fixing the ideals I and J and the algebra $S(L)$. Thus the induced map $\hat{\pi}: S(L) \to U(L)$ is an isomorphism of G-modules. (Note $\hat{\pi}$ is only a linear map, not an algebra homomorphism. See Ref. 3 for remarks on this point.) Denote by $S(L)^G$ and $U(L)^G$ the G-invariant subspaces of $S(L)$ and $U(L)$, respectively. These are in fact subalgebras. By our earlier comments we may identify $S(L)^G$ with $P(L)^G$, the G-invariant polynomial functions on L using the nondegenerate Killing form.

Similar comments apply to H and we may identify $S(H)^{\mathrm{W}}$ with $P(H)^{\mathrm{W}}$, the Weyl-invariant polynomials. A theorem of Chevalley says that $P(L)^G$ and $P(H)^{\mathrm{W}}$ are isomorphic as algebras. In particular, because W is a finite reflection group, this algebra is finitely generated with l generators. Moreover, the degrees m_j of these generators are known. Let $d(L)$ be the set of these generators, $d(L) = \{m_i,\ i = 1,...,l\}$. Further the center $\mathfrak{Z}$ of $U(L)$ is precisely $U(L)^G$ and $\mathfrak{Z}$ contains the l independent Casimirs of L. The trace polynomials generate $P(L)^G$. We will express these as follows. Suppose ϕ: $L \to \mathrm{gl}[V(\lambda)]$ is an irreducible representation of highest weight λ. Then for $x \in L$ we have the trace polynomial $\mathrm{Tr}[\phi(x)^k]$. Sometimes we shall simply express this as $\mathrm{Tr}_\lambda\, x^k$, the notation readily including reducible representations as well.

B. The anomaly coefficients and their properties

We may now begin describing our approach to calculating anomaly coefficients. To do this we shall begin by reviewing some other investigations pertaining to the weight structure of representations of L. First we recall the notion of the $2n$th order index $I_{2n}(\lambda)$ of a representation $V(\lambda)$,

$$I_{2n}(\lambda) = \sum_{\mu \in \pi(\lambda)} (\mu,\mu)^n. \tag{2.4}$$

This was introduced[5] to generalize the notion of Dynkin index.[6] Here $I_0(\lambda) = \dim V(\lambda)$ and $I_2(\lambda)$ is, up to a constant multiple, the Dynkin index. [Note that some authors write $I_{2n}(\lambda)$ as $l_{2n}(\lambda)$.] The Dynkin index $I_2(\lambda)$ has both additive and multiplicative properties as well as a simple scaling for subalgebras:

$$I_2(\lambda_1 \oplus \lambda_2) = I_2(\lambda_1) + I_2(\lambda_2), \tag{2.5}$$

$$I_2(\lambda_1 \otimes \lambda_2) = I_2(\lambda_1) I_0(\lambda_2) + I_0(\lambda_1) I_2(\lambda_2), \tag{2.6}$$

$$I_2(\lambda^{L_1}) = \rho I_2(\lambda^{L_2}). \tag{2.7}$$

In (2.5)–(2.7) we have a semisimple subalgebra $L_2 \subset L_1$ and the representation λ^{L_1} branching to λ^{L_2}; the constant ρ that appears depends only on the embedding. In general, the higher-order indices retain only the first of these properties. For example, for L semisimple we have

$$I_4(\lambda_1 \otimes \lambda_2) = I_4(\lambda_1) I_0(\lambda_2) + I_0(\lambda_2) I_4(\lambda_2) + [2(l+2)/l] I_2(\lambda_1) I_2(\lambda_2). \tag{2.8}$$

Despite the loss of these additional properties, such higher-order indices have been useful in the decomposition of tensor products and computing branching ratios.[7] The second- and fourth-order indices have been extensively tabulated.[8]

With a desire to maintain the multiplicative as well as additive properties of the Dynkin index, Okubo[9,10] modified the above definition of index. With the definition

$$\tilde{I}_4(\lambda) = I_4(\lambda) - \frac{(l+2) I_0(\lambda_{\text{adj}})}{l\,[I_0(\lambda_{\text{adj}}) + 2]} \frac{I_2(\lambda)}{I_0(\lambda)} - \frac{1}{6} \frac{I_2(\lambda_{\text{adj}})}{I_0(\lambda_{\text{adj}})} I_2(\lambda), \tag{2.9}$$

where λ_{adj} denotes the adjoint representation of L, one finds[9] that (2.6) generalizes for $I_4(\lambda_1 \otimes \lambda_2)$. These modified indices are then observed to correspond to the higher, even-ordered Casimirs of $U(L)$. This is fairly evident when we express (2.4) in the form

$$I_{2n}(\lambda) = \mathrm{Tr}(g^{ij} \phi(h_i) \phi(h_j))^n. \tag{2.10}$$

Here $\phi(h_i)$ is the representation under λ of the element h_i in the Cartan subalgebra of L and g^{ij} is the inverse of the Killing form. Clearly Okubo is polarizing a given polynomial in $U(L)$. In order to understand this polarization we must discuss the basis of $U(L)^G$ he chooses.

As we mentioned in the first part of this section a trace polynomial such as (2.10) lies in the center $\mathfrak{Z}$ of $U(L)$ and $\mathfrak{Z}$ is finitely generated with l generators. It will be convenient in what ensues to introduce the polynomials $f_k(\mu)$ in $P(H)^W$ defined for each $\mu\in\Lambda$ and $k\in N$ by

$$f_k(\mu) = \frac{1}{|\mathrm{Stab}\,\mu|}\sum_{\sigma\in W}(\sigma\mu)^k. \tag{2.11}$$

Clearly $f_k(\mu)$ is a homogeneous polynomial with integer coefficients of degree k in the λ; and it suffices to take $\mu\in\Lambda^+$, each weight being conjugate under W to one and only one dominant weight. (This polynomial is referred to as Sym μ^k in Ref. 3.)

An interesting question to ask is the following: For which $\mu\in\Lambda^+$ does $\{f_k(\mu),\ k\in d(L)\}$ constitute a basis of $P(H)^W$ over Q? Let us call $F(L)$ the set of such μ forming a basis in this fashion. The quartic trace identity for the antisymmetric tensor representation (with Young tableaux $\boxminus$) of A_7,

$$f_4(\boxminus) = \tfrac{1}{12}[f_2(\boxminus)]^2, \tag{2.12}$$

shows, for example, that for some μ the f_k are not always independent. This identity is related to E_7.[9] Although we do not know the answer to our question in general, there are some widely used weights (which may coincide) that do yield a basis. One is the highest weight of the adjoint representation of L. For the classical groups the homotopy equivalence between the groups SO(n), U(n), Sp(n), and GL(n,F) with $F=R$, C, or H means that the weight associated with the vector (or fundamental) representation also yields a basis. When we are dealing with the algebra L rather than the group, this last representation need not be the representation of lowest Dynkin index whose highest weight yields a basis of $P(H)^W$. For example, the (algebra) isomorphisms $A_1\sim B_1$, $B_2\sim C_2$, $A_3\sim D_3$ show, in fact, that the spinor representations of so(3), so(5), and so(6) also provide bases via their highest weights, these representations having both lower dimension and Dynkin index than the corresponding vector representations.

In what follows we will denote by the *generating* representation of L that representation whose highest weight λ_{gen} is an element of $F(L)$ and whose Dynkin index is lowest. This is obviously only defined up to an algebra automorphism. Actually for so(7) there is ambiguity as both the spinor and vector representations have the same Dynkin index. Here we take the spinor representation as $f_4(\lambda_{\mathrm{vector}}) = -2f_4(\lambda_{\mathrm{spinor}})$ + products of lower f's. Clearly, the generating representation is just the vector representation of su(n) and sp(n) while for so(n) it depends on the value of n.

Returning to Okubo's work, he is essentially expressing polynomials in $P(H)^W$ in terms of the trace polynomials of the vector representation. Alternatively he is calculating the Chern characters of the given representation. The Chern characters have the splitting properties

$$\mathrm{ch}(\alpha\oplus\beta) = \mathrm{ch}\,\alpha\oplus\mathrm{ch}\,\beta, \tag{2.13}$$

$$\mathrm{ch}(\alpha\otimes\beta) = \mathrm{ch}\,\alpha\wedge\mathrm{ch}\,\beta, \tag{2.14}$$

and these give rise to the additive and multiplicative properties of the modified indices via

$$\mathrm{Tr}_\lambda\, e^{iF/2\pi} = \sum_{k=0}^{\infty}\frac{\bar I_k(\lambda)}{k!}\mathrm{Tr}_{\mathrm{vector}}\left(\frac{iF}{2\pi}\right)^k + \text{products of traces}. \tag{2.15}$$

These modified indices $\bar I_k$ are also known as the leading anomaly coefficients; the coefficients of the products of the traces give the nonleading anomaly coefficients. For a given representation these anomaly coefficients or the Chern characters are in principle known: a simple method for their calculation is, however, another matter. We shall now comment upon the calculation of these coefficients proposing a new technique.

First let us characterize these coefficients purely algebraically. To do this, we define polynomials $P_\pi(\mu,x)$ for $\mu\in\Lambda^+$, $x\in H$ and partitions of k, $\pi=(p_1^{\pi_1},\dots,p_r^{\pi_r})$ by

$$P_\pi(\mu,x) = (\mathrm{Tr}_\mu\, x^{p_1})^{\pi_1}\dots(\mathrm{Tr}_\mu\, x^{p_r})^{\pi_r}. \tag{2.16}$$

Thus when $\mu\in F(L)$ we may write

$$\mathrm{Tr}_\lambda\, x^k = \sum_\pi A_\pi(\lambda,\mu)P_\pi(\mu,x). \tag{2.17}$$

Here the sum is over partitions of π of k with $p_i\in d(L)$. As defined by (2.15) we have $\bar I_k(\lambda) = A_{(k)}(\lambda,\lambda_{\mathrm{vector}})$. We will be interested in calculating $A_\pi(\lambda,\lambda_{\mathrm{gen}})$, which are integers by our choice of λ_{gen}. Let us set $A_\pi(\lambda) = A_\pi(\lambda,\lambda_{\mathrm{gen}})$.

The calculation of anomaly coefficients is then the determination (in a given basis) of the coordinates $A_\pi(\lambda)$ for a given polynomial in $P(H)^W$. Several approaches[7,9–11] determine these from the character formula

$$\mathrm{ch}_{V(\lambda)} = \sum_{\mu\in\pi(\lambda)} m_\lambda(\mu)e(\mu) \tag{2.18}$$

$$= \sum_{\mu\in\pi^+(\lambda)} m_\lambda(\mu)\frac{1}{|\mathrm{Stab}\,\mu|}\sum_{\sigma\in W}e(\sigma\mu) \tag{2.19}$$

$$= \frac{\sum_{\sigma\in W}\mathrm{sn}(\sigma)e(\sigma[\lambda+\delta])}{\sum_{\sigma\in W}\mathrm{sn}(\sigma)e(\sigma\delta)}, \tag{2.20}$$

where the $e(\mu)$, $\mu\in\Lambda$, are the basis elements of $Z(\Lambda)$, the group ring of Λ over Z. These methods are essentially based on the form (2.20) and its expansions.

Rather than working with the characters directly to evaluate the $A_\pi(\lambda)$ we will break the problem up into a calculation for each of the weight spaces appearing in (2.17). First let us view the $e(\mu)$ appearing in (2.19) as a formal exponential.

Thus

$$\mathrm{ch}_{V(\lambda)} = \sum_{k=0}^{\infty}\frac{1}{k!}\sum_{\lambda\in\pi^+(\lambda)} m_\lambda(\mu)f_k(\mu). \tag{2.21}$$

If we express $f_k(\mu)$ in terms of a basis of $P(H)^W$ analogously to (2.16) and (2.17) for some $\nu\in F(L)$ as

$$f_k(\mu) = \sum_\pi a_\pi(\mu,\nu)\prod_i(f_{p_i}(\nu))^{\pi_i}. \tag{2.22}$$

Then we have

$$A_\pi(\lambda,\lambda_{\mathrm{gen}}) = \sum_{\mu\in\pi^+(\lambda)} m_\lambda(\mu)a_\pi(\mu,\lambda_{\mathrm{gen}}). \tag{2.23}$$

Again for simplicity set $a_\pi(\mu,\lambda_{\mathrm{gen}}) = a_\pi(\mu)$.

Our approach to calculating anomaly coefficients reduces then to being able to express the Weyl-invariant polynomial $f_k(\mu)$ in a given basis of $P(H)^W$. One might expect this problem was studied under classical invariant theory last century, though I am unaware of where such algorithms might appear. In the following section we will show how to evaluate $a_\pi(\mu)$ for su(n) giving a general expression for an arbitrary leading anomaly coefficient and explicit formulas up to the sixth-order coefficients. This considerably extends the formulas in the literature. We note this method depends on knowing the multiplicities $m_\lambda(\mu)$. These multiplicities may be evaluated by several known recursions and have been extensively tabulated.[12] The method is recursive and easily implemented. In a later section we will, in fact, observe some congruences the a_π's evidence and utilize these to discuss global anomalies.

We end this section by mentioning two other approaches to calculating the leading anomaly coefficients. Both are based on expressing a given representation in terms of tensor products of representations whose leading anomaly coefficient is known and then using the multiplicative properties of the $\bar{I}_k$. This involves knowing the branching rules of tensor products. Schellekens and Warner[13] work directly with the Chern characters utilizing (2.15). Most other authors do this for a fixed $\bar{I}_k$. A disadvantage of this approach is that, aside from the often nontrivial question of evaluating the branching rules for large representations, frequently several unknown representations will appear in calculating any tensor product and so various linear combinations of tensor products must be found to isolate any one unknown index. Also, the nonleading anomalies are not given in this approach.

III. CALCULATING SU(n) ANOMALY COEFFICIENTS

We shall now obtain the anomaly coefficients $a_\pi(\mu)$ [(2.23)] for an arbitrary representation of su(n) subject to $|\pi|\leqslant n$. This restriction will be discussed in due course. First we will recall some elementary properties of su(n) and discuss the basis of symmetric functions we will be using. After a brief interlude where we fix notation and establish some elementary results pertaining to symmetric polynomials we then go on to give expressions for $a_\pi(\mu)$. These formulas allow us to give some simple expressions for the leading anomaly coefficient. A short section of examples and use of the accompanying tables is included.

A. SU(n)

The Weyl group of $A_{n-1}\sim$su(n) is S_n, the symmetric group on n symbols. Following our discussion in Sec. II A we are interested in symmetric polynomials. We shall now establish the basis we will later be expanding in. Let us denote the inverse to the $(n-1)\times(n-1)$ Cartan matrix $A=(a_{ij})$ by $G=(g_{kj})$, where

$$g_{kj}=\begin{cases}(1/n)(n-j)k, & k\leqslant j,\\ (1/n)(n-k)j, & j\leqslant k.\end{cases} \tag{3.1}$$

The matrix G gives us a metric on weight space. If $\{x_1,\ldots,x_n\}$ form an orthonormal basis of R^n we may view the weight space as the subspace of R^n orthogonal to $\eta=x_1+\cdots+x_n$. Set $\bar{\eta}=\eta/n$. The Weyl group acts naturally on the basis of R^n:

$$w_i(x_j)=\begin{cases}x_{i+1}, & i=j,\\ x_i, & i=j-1,\\ x_j, & i\neq j,j-1.\end{cases} \tag{3.2}$$

In terms of this basis we have $\Phi=\{\alpha_i: \alpha_i=x_i-x_{i+1},\ 1\leqslant i\leqslant n-1\}$ and the elementary weights are

$$\lambda_k=\frac{1}{n}\Big[\sum_{j\leqslant k}(n-k)j\alpha_j+\sum_{k<j}(n-j)k\alpha_j\Big] \tag{3.3}$$

$$=x_1+\cdots+x_k-k\bar{\eta}. \tag{3.4}$$

We may associate with any dominant weight μ a Young diagram as follows. Suppose

$$\mu=\sum_{i=1}^{n-1}n_i\lambda_i \tag{3.5}$$

$$=\sum_{j=1}^{n-1}l_jx_j-c(\mu)\bar{\eta}, \tag{3.6}$$

where

$$l_j=\sum_{j\leqslant i}n_j, \tag{3.7}$$

$$c(\mu)=\sum_{j=1}^{n-1}jn_j. \tag{3.8}$$

Then μ is associated with the Young diagram with l_i the length of the ith row. When there is no confusion we will also denote the irreducible representation of which μ is the highest weight by the same Young diagram. The vector (or generating) representation is then given by $\square$.

Remark: In the notation of Ref. 8 we have μ given by $(n_n,\ldots,n_1)$ while in Ref. 12 it is given by

$$\begin{pmatrix}n_1\\ \vdots\\ n_n\end{pmatrix}.$$

With this notation we find

$$f_k(\square)=\lambda_1^k+(\lambda_2-\lambda_1)^k+\cdots+(\lambda_n-\lambda_{n-1})^k+(-\lambda_n)^k \tag{3.9}$$

$$=x_1^k+(x_2-\bar{\eta})^k+\cdots+(x_n-\bar{\eta})^k. \tag{3.10}$$

Because we are dealing with su(n), that is, the $\bar{\eta}=0$ subspace of R^n, we may simply set $\bar{\eta}=0$ in (3.10) and (3.11). [The u(n) trace or $\bar{\eta}\neq0$ component may be dealt with by suitable modifications of the ensuing discussion.] Thus for our purposes we may identify $f_k(\square)$ with the symmetric power function s_k:

$$f_k(\square)=s_k=\sum_{i=1}^{n}x_i^k. \tag{3.11}$$

For an arbitrary μ our technique for calculating the anomaly coefficients for su(n) reduces then to expressing the symmetric polynomial $f_k(\mu)$ in terms of the symmetric functions of powers. This we shall do after some general comments on symmetric functions.

B. Symmetric functions

We shall now fix some notation and establish a few simple results associated with symmetric polynomials. (For further results see Ref. 14.) Let $\pi=(\gamma_1,\gamma_2,\ldots,\gamma_q)=(p_1^{\pi_1},\ldots,p_r^{\pi_r})$, where $\gamma_1\geqslant\gamma_2\geqslant\cdots\geqslant\gamma_q>0$, $p_1>p_2>\cdots>p_r>0$, be a partition of $|\pi|$, the weight of π, with

$$|\pi|=\gamma_1+\cdots+\gamma_q=\pi_1 p_1+\cdots+\pi_r p_r. \qquad (3.12)$$

Here the length of the partition is $l(\pi)=q$. Set

$$g(\pi)=\pi_1!\pi_2!\cdots\pi_r!\,. \qquad (3.13)$$

Consider now the ring of polynomials with rational integral coefficients $Z[x_1,\ldots,x_n]$ in n independent variables $x_1,\ldots,x_n$. The symmetric group S_n acts naturally on this ring by permuting the variables and we have the symmetric polynomials forming the subring invariant under this action,

$$I_n=Z[x_1,\ldots,x_n]^{S_n}. \qquad (3.14)$$

This is a graded ring,

$$I_n=\underset{q}{\oplus} I_n^q, \qquad (3.15)$$

where I_n consists of the homogeneous symmetric polynomials of degree q together with the zero polynomial.

For each $\gamma=(\gamma_1,\gamma_2,\ldots,\gamma_n)\in N^n$ let us denote by x^γ the monomial

$$x^\gamma=x_1^{\gamma_1}x_2^{\gamma_2}\cdots x_n^{\gamma_n}. \qquad (3.16)$$

Then for any partition π of length $l(\pi)\leqslant n$ we define the symmetric monomial

$$m_\pi(x_1,\ldots,x_n)=\sum_{\substack{\text{distinct}\\ \text{permutations}}} x_1^{\gamma_1}\cdots x_k^{\gamma_k}. \qquad (3.17)$$

Here the summation is over distinct permutations of $\Pi_i x_i^{\gamma_i}$. When we do not require those permutations to be distinct we have the augmented monomial symmetric functions[15]

$$n_\pi(x_1,\ldots,x_n)=\sum_{\text{perms}} x_1^{\gamma_1}\cdots x_k^{\gamma_k}. \qquad (3.18)$$

Then

$$n_\pi(x_1,\ldots,x_n)=g(\pi)m_\pi(x_1,\ldots,x_n). \qquad (3.19)$$

We have that the set $\{m_\pi,\ |\pi|=k\}$ form a basis of I_n^k. There are several bases of I_n commonly employed. These include (a) the elementary symmetric polynomials $e_r=m_{(1^r)}$, which are algebraically independent over Z and have the generating function

$$E(t)=\sum_{r\geqslant 0} e_r t^r=\prod_{i\geqslant 1}(1+x_i t); \qquad (3.20)$$

(b) the complete symmetric, Wronski, or aleph functions

$$h_r=\sum_{|\pi|=r} m_\pi, \qquad (3.21)$$

which are again algebraically independent over Z and have the generating function

$$H(t)=\sum_{r\geqslant 0} h_r t^r=\prod_{i\geqslant 1}(1-x_i t)^{-1}; \qquad (3.22)$$

and (c) the symmetric power polynomials $s_r=m_{(r)}$, which are algebraically independent over Q and have the generating function

$$S(t)=\sum_{r\geqslant 0} s_{r+1}t^r=\frac{d}{dt}\ln H(t). \qquad (3.23)$$

These functions are defined to be zero for $r<0$. For our partition $\pi=(\gamma_1,\ldots,\gamma_q)$, we define

$$e_\pi=\prod_i e_{\gamma_i},\quad h_\pi=\prod_i h_{\gamma_i},\quad s_\pi=\prod_i s_{\gamma_i}.$$

The transition functions between the e_π, h_π, and Schur functions (all of which are Z-bases of the symmetric functions) are described in Ref. 14. For partitions of small weight these are tabulated in Ref. 15 along with the transition functions between the augmented symmetric monomials and the symmetric powers s_π. We note that the assumption of $q=l(\pi)\leqslant n$ enters when we use the above functions as bases for I_n^q. When this does not hold we have further relations between the monomials to be accounted for. This is the origin of the restriction $l(\pi)\leqslant n$ that will appear when we apply our results to su(n).

We will now establish a simple lemma pertaining to the symmetric monomials. First we introduce the symmetrization operator ζ as follows. Let $f(x_1,\ldots,x_n)$ be any function of $x_1,\ldots,x_n$ and set

$$\zeta f(x_1,\ldots,x_n)=\sum_{\sigma\in S_n} f(x_{\sigma(1)},\ldots,x_{\sigma(n)}). \qquad (3.24)$$

Thus, if $f(x_1,\ldots,x_n)$ is a symmetric function, we have

$$\zeta f(x_1,\ldots,x_n)=|\mathrm{Stab}\, f(x_1,\ldots,x_n)|\, f(x_1,\ldots,x_n), \qquad (3.25)$$

and, in particular,

$$|\mathrm{Stab}\, m_\pi|=(n-q)!g(\pi), \qquad (3.26)$$

$$|\mathrm{Stab}\, n_\pi|=(n-q)!\,. \qquad (3.27)$$

Then we have the following lemma.

Lemma:

$$\zeta m_\pi(l_1x_1,\ldots,l_nx_n)=(n-l(\pi))!m_\pi(l)n_\pi(x)$$
$$=(n-l(\pi))!n_\pi(l)m_\pi(x), \qquad (3.28)$$

$$\zeta n_\pi(l_1x_1,\ldots,l_nx_n)=(n-l(\pi))!n_\pi(l)n_\pi(x). \qquad (3.29)$$

Proof: This follows if upon writing

$$m_\pi(x)=\frac{1}{|\mathrm{Stab}\, x_1^{\gamma_1}\cdots x_q^{\gamma_q}|}\zeta(x_1^{\gamma_1}\cdots x_q^{\gamma_q})$$
$$=\frac{1}{(n-l(\pi))!g(\pi)}\zeta(x_1^{\gamma_1}\cdots x_q^{\gamma_q}),$$

then

$$\zeta m_\pi(l_1x_1,\ldots,l_nx_n)=\zeta\frac{1}{(n-l(\pi))!g(\pi)}$$
$$\times\sum_{\sigma\in S_n} l_{\sigma(1)}^{\gamma_1}\cdots l_{\sigma(n)}^{\gamma_n}x_{\sigma(1)}^{\gamma_1}\cdots x_{\sigma(n)}^{\gamma_n}$$
$$=\sum_{\sigma\in S_n} l_{\sigma(1)}^{\gamma_1}\cdots l_{\sigma(n)}^{\gamma_n}m_\pi(x)$$
$$=(n-l(\pi))!n_\pi(l)m_\pi(x).$$

C. A_{n-1} anomaly coefficients

We will now bring the results of Secs. III A and III B together to derive the anomaly coefficients of su(n). Let

$$\mu = \sum_{i=1}^{n} l_i x_i \tag{3.30}$$

be a weight. Thus $l_n = 0$ for su(n) while l_n can be nonzero for u(n). Viewing this as a polynomial in the x; we have

$$\mu^k = \sum_{|\pi| = k} \frac{k!}{\gamma_1! \cdots \gamma_k!} m_\pi (l_1 x_1, \ldots, l_n x_n). \tag{3.31}$$

Now we may write

$$f_k(\mu) = \frac{1}{|\mathrm{Stab}\,\mu|} \sum_{\sigma \in S_n} (\sigma\mu)^k \tag{3.32}$$

$$= (|\mathrm{Orbit}\,\mu|/n!) \zeta(\mu^k), \tag{3.33}$$

where ζ is the symmetrization operator introduced in Sec. III B. Thus (3.28) and (3.31) together yield

$$f_k(\mu) = \frac{|\mathrm{Orbit}\,\mu|}{n!} \sum_{|\pi| = k} \frac{k!(n - l(\pi))!}{\gamma_1! \cdots \gamma_k!} m_\pi(l) n_\pi(x). \tag{3.34}$$

Let us express the augmented symmetric monomials $n_\pi(x)$ in terms of the symmetric powers $s_{\pi'}$ via

$$n_\pi(x) = \sum_{\pi'} n_{\pi\pi'} s_{\pi'}, \tag{3.35}$$

where only now do we assume $|\pi| = |\pi'| \leqslant n$. We have shown that if (2.23) is given as

$$f_k(\mu) = \sum_{|\pi'| = k} a_{\pi'}(\mu) s_{\pi'}, \tag{3.36}$$

then the coefficient is

$$a_{\pi'}(\mu) = \frac{|\mathrm{Orbit}\,\mu|}{n!} \sum_{|\pi| = k} n_{\pi\pi'} \frac{k!(n - l(\pi))!}{\gamma_1! \cdots \gamma_k!} m_\pi(l). \tag{3.37}$$

Although (3.37) may look complicated, it is, in fact, very easy to compute low-order coefficients because the summation over different partitions is not that large. Thus

$$\begin{aligned} f_2(\mu) = {} & [|\mathrm{Orbit}\,\mu|/n(n-1)]\{[(n-1)m_{(2)}(l) \\ & - 2m_{(1^2)}(l)]s_2 + 2m_{(1^2)}(l)s_1^2\}, \end{aligned} \tag{3.38}$$

$$\begin{aligned} f_3(\mu) = {} & [|\mathrm{Orbit}\,\mu|/n(n-1)(n-2)] \\ & \times\{[(n-1)(n-2)m_{(3)}(l) \\ & - 3(n-2)m_{(2,1)}(l) + 12m_{(1^3)}(l)]s_3 \\ & + [3(n-2)m_{(2,1)}(l) - 18m_{(1^3)}(l)]s_2 s_1 \\ & + 6m_{(1^3)}(l)s_1^3\}. \end{aligned} \tag{3.39}$$

Let us define a reduced coefficient $\bar{a}_\pi(\mu)$ by

$$a_\pi(\mu) = [|\mathrm{Orbit}\,\mu|/n(n-1)\cdots(n-|\pi|)]\bar{a}_\pi(\mu). \tag{3.40}$$

Further, if we drop the explicit l dependence of the symmetric monomials we have

$$\begin{aligned} \bar{a}_{(4)} = {} & (n-1)(n-2)(n-3)m_{(4)} \\ & - 4(n-2)(n-3)m_{(3,1)} \\ & - 6(n-2)(n-3)m_{(2^2)} \\ & + 24(n-3)m_{(2,1^2)} - 144m_{(1^4)}, \end{aligned} \tag{3.41a}$$

$$\begin{aligned} \bar{a}_{(3,1)} = {} & 4(n-2)(n-3)m_{(3,1)} \\ & - 24(n-3)m_{(2,1^2)} + 192m_{(1^4)}, \end{aligned} \tag{3.41b}$$

$$\begin{aligned} \bar{a}_{(2^2)} = {} & 6(n-2)(n-3)m_{(2^2)} \\ & - 12(n-3)m_{(2,1^2)} + 72m_{(1^4)}, \end{aligned} \tag{3.41c}$$

$$\bar{a}_{(2,1^2)} = 12(n-3)m_{(2,1^2)} - 144m_{(1^4)}, \tag{3.41d}$$

$$\bar{a}_{(1^4)} = 24m_{(1^4)}, \tag{3.41e}$$

$$\begin{aligned} \bar{a}_{(5)} = {} & (n-1)(n-2)(n-3)(n-4)m_{(5)} \\ & - 5(n-2)(n-3)(n-4)[m_{(4,1)} + 2m_{(3,2)}] \\ & + 20(n-3)(n-4)[2m_{(3,1^2)} + 3m_{(2^2,1)}] \\ & - 360(n-4)m_{(2,1^3)} + 4!\times 5! m_{(1^5)}, \end{aligned} \tag{3.42}$$

$$\begin{aligned} \bar{a}_{(6)} = {} & (n-1)(n-2)(n-3)(n-4)(n-5)m_{(6)} \\ & - (n-2)(n-3)(n-4)(n-5) \\ & \times[6m_{(5,1)} + 15m_{(4,2)} + 20m_{(3^2)}] \\ & + 60(n-3)(n-4)(n-5) \\ & \times[m_{(4,1^2)} + 2m_{(3,2,1)} + 3m_{(2^3)}] \\ & - 360(n-4)(n-5)[2m_{(3,1^3)} + 3m_{(2^2,1^2)}] \\ & + 8640(n-5)m_{(2,1^4)} - 5!\times 6! m_{(1^6)}. \end{aligned} \tag{3.43}$$

Several comments are now in order.

(1) When we are dealing with su(n) rather than u(n), $s_1 = 0$ in all of the preceding equations; in this case, $a_\pi = 0$ whenever $l_n = 1$.

(2) Suppose the Young diagram we are calculating for has r rows,

$$l = (l_1, \ldots, l_r, 0, \ldots).$$

Then $m_\pi(l) = 0$ whenever $r < l(\pi)$.

This last comment means that for a Young diagram of only a few rows many of the terms appearing in $a_\pi(\mu)$ vanish. The formula (3.37) and its specializations (3.38)–(3.43) are easy to implement on a computer. We have tabulated some of these results in Tables I–III. We will conclude this section discussing these results, giving some examples of their use, and comparing them with existing results. Before doing this, however, it is useful to specialize (3.37) to the case of the leading coefficients $a_{(k)}(\mu)$, giving some simple expressions for these.

D. Leading anomaly coefficients

Our purpose now is to give some simple closed expressions for the coefficient $a_{(k)}(\mu)$ in the general formula (3.37), this being the term relevant for the leading anomaly coefficient. Similar reasoning may be applied to the nonleading coefficients though we do not pursue this here.

First, let us define the symmetric polynomial $F(k,j)$ by

TABLE I. The coefficients a_k for su(4) weights (d_i). The orbit size (O.S.) of this weight and the dimension (DIM) of the corresponding highest weight representation and its congruence class are also given.

d_3	d_2	d_1	DIM	O.S.	Class	a_2	a_3	a_4
0	0	1	4	4	1	1	1	1
0	1	0	6	6	2	2	0	−4
0	0	2	10	4	2	4	8	16
1	0	1	15	12	0	8	0	8
0	0	3	20	4	−1	9	27	81
0	1	1	20	12	−1	11	9	−13
0	2	0	20	6	0	8	0	−64
0	0	4	35	4	0	16	64	256
1	0	2	36	12	1	19	27	67
0	1	2	45	12	0	24	48	72
0	3	0	50	6	2	18	0	−324
0	0	5	56	4	1	25	125	625
0	2	1	60	12	1	27	15	−237
1	1	1	64	24	2	40	0	−80
1	0	3	70	12	2	36	96	312
0	0	6	84	4	2	36	216	1296
0	1	3	84	12	1	43	135	403
2	0	2	84	12	0	32	0	128
0	4	0	105	6	0	32	0	−1024
0	0	7	120	4	−1	49	343	2401
1	0	4	120	12	−1	59	225	947

$$F(k,j) = \sum_{\substack{|\pi| = k \\ l(\pi) = j}} \frac{k!}{\gamma_1! \cdots \gamma_j!} m_\pi . \tag{3.44}$$

We will return to properties of this polynomial in a moment; the following lemma shows its appearance.

Lemma:

$$a_{(k)}(\mu) = |\mathrm{Orbit}\,\mu| \sum_{j=1}^{k} (-1)^{j-1} \times \frac{(j-1)!(n-j)!}{n!} F(k,j)[\mu]. \tag{3.45}$$

TABLE II. The coefficients a_k for su(5) weights (d_i). The orbit size (O.S.) of this weight and the dimension (DIM) of the corresponding highest weight representation and its congruence class are also given.

d_4	d_3	d_2	d_1	DIM	O.S.	Class	a_2	a_3	a_4	a_5
0	0	0	1	5	5	1	1	1	1	1
0	0	1	0	10	10	2	3	1	−3	−11
0	0	0	2	15	5	2	4	8	16	32
1	0	0	1	24	20	0	10	0	10	0
0	0	0	3	35	5	−2	9	27	81	243
0	0	1	1	40	20	−2	16	18	4	−78
0	1	0	1	45	30	−1	21	9	−9	69
0	0	2	0	50	10	−1	12	8	−48	−352
0	0	0	4	70	5	−1	16	64	256	1024
1	0	0	2	70	20	1	24	34	84	154
0	1	1	0	75	30	0	30	0	−90	0
0	0	1	2	105	20	−1	34	76	154	196
0	0	0	5	126	5	0	25	125	625	3125
0	1	0	2	126	30	0	45	75	135	375
1	0	0	3	160	20	2	46	122	394	1178
0	0	2	1	175	20	0	40	50	−140	−1750
0	0	3	0	175	10	1	27	27	−243	−2673
1	0	1	1	175	60	2	78	36	−18	−516
2	0	0	2	200	20	0	40	0	160	0
0	0	0	6	210	5	1	36	216	1296	7776
0	2	0	1	210	30	2	54	−12	−234	1692

TABLE III. The coefficients a_k for su(6) weights (d_i). The orbit size (O.S.) of this weight and the dimension (DIM) of the corresponding highest weight representation and its congruence class are also given.

d_5	d_4	d_3	d_2	d_1	DIM	O.S.	Class	a_2	a_3	a_4	a_5	a_6
0	0	0	0	1	6	6	1	1	1	1	1	1
0	0	0	1	0	15	15	2	4	2	−2	−10	−26
0	0	1	0	0	20	20	3	6	0	−6	0	66
0	0	0	0	2	21	6	2	4	8	16	32	64
1	0	0	0	1	35	30	0	12	0	12	0	12
0	0	0	0	3	56	6	3	9	27	81	243	729
0	0	0	1	1	70	30	3	21	27	21	−45	−339
0	1	0	0	1	84	60	−1	34	8	−2	80	−146
0	0	0	0	0	105	15	−2	16	16	−32	−320	−1 664
0	0	1	0	1	105	60	−2	40	28	−8	−20	520
1	0	0	0	2	120	30	1	29	41	101	185	389
0	0	0	0	4	126	6	−2	16	64	256	1024	4 096
0	0	2	0	0	175	20	0	24	0	−96	0	4 224
0	1	0	1	0	189	90	0	72	0	−108	0	−828
0	0	0	1	2	210	30	−2	44	104	236	440	284
0	0	1	1	0	210	60	−1	58	26	−134	−430	898
0	0	0	0	5	252	6	−1	25	125	625	3125	15 625
0	1	0	0	2	280	60	0	72	108	216	540	792
1	0	0	0	3	315	30	2	56	148	476	1420	4 316
0	0	1	0	2	336	60	−1	82	152	286	560	1 822
1	0	0	1	1	384	120	2	128	88	80	−440	−2 272

This follows from (3.37) when we realize that

$$n_{(k)\pi} = (-1)^{l(\pi)-1}(l(\pi)-1)! . \tag{3.46}$$

The evaluation of the leading anomaly coefficient reduces then to knowing how to evaluate $F(k,j)[\mu]$. Let $\mu = (l_1,...,l_N)$, i.e., the number of nonzero coordinates l_i in (3.30) is N. Viewed as a Young diagram, μ has N rows. One can readily show the following lemma.

Lemma: If m_π is a symmetric monomial in $N \geqslant k$ variables, then

$$F(k,j) = \sum_{p=0}^{j-1} (-1)^p \binom{N-j+p}{p} \times \sum (x_1 + \cdots + x_{j-p})^k. \tag{3.47}$$

Corollary: We have

$$(x_1 + \cdots + x_N)^k = \sum_{p=0}^{k-1} (-1)^p \binom{N-k+p-1}{p} \times \sum (x_1 + \cdots + x_{k-p})^k. \tag{3.48}$$

Here (3.48) follows from (3.47) by summing over j from 1 to k. The second summations in (3.47) and (3.48) are understood to be over all distinct permutations of $j-p$ or $k-p$ variables.

Our second lemma means the evaluation of (3.45) is straightforward. First, homogeneity means that if $(\alpha\mu) = (\alpha l_1,...,\alpha l_N)$, then

$$a_{(k)}(\alpha\mu) = \alpha^k a_{(k)}(\mu). \tag{3.49}$$

We have, for example,

(i) $a_{(k)}(l_1) = l_1^k,$ (3.50)

(ii) $a_{(k)}(l_1,l_2) = \chi_{l_1,l_2}[n(l_1^k + l_2^k) - (l_1+l_2)^k],$ (3.51)

$$\text{(iii)}\ a_{(k)}(l_1,l_2,l_3) = \chi_{l_1,l_2,l_3}[n^2(l_1^k + l_2^k + l_3^k) - n[(l_1+l_2)^k + (l_1+l_3)^k + (l_2+l_3)^k + l_1^k + l_2^k + l_3^k] + 2(l_1+l_2+l_3)^k], \tag{3.52}$$

$$\text{(iv)}\ a_{(k)}(l_1,l_2,l_3,l_4) = \chi_{l_1,l_2,l_3,l_4}\left[n^3\left(\sum l_i^k\right) - n^2\left(\sum (l_1+l_2)^k + 3\sum l_i^k\right) + n\left(2\sum (l_1+l_2+l_3)^k + \sum (l_1+l_2)^k + 2\sum l_1^k\right) - 6(l_1+l_2+l_3+l_4)^k\right]. \tag{3.53}$$

The coefficients $\chi_{(l)}$ in the above are defined as follows:

$$|\text{Orbit}(l_1,\ldots,l_r)| = \chi_{(l_1,\ldots,l_r)}|\text{Orbit}(r,r-1,\ldots,1)|.$$

That is, χ accounts for the reduction in the orbit of μ when several of the l_i coincide. For example, $\chi_{l,l} = \frac{1}{2}$, $\chi_{l_1,l_1,l_3} = \chi_{l_1,l_2,l_2} = \frac{1}{2}$, $\chi_{l_1,l_1,l_1} = \frac{1}{6}$, and so on. Clearly if $\mu = (l_1,\ldots,l_N) = (q_1^{\mu_1},\ldots,q_s^{\mu_s})$, then $\chi(\mu) = 1/g(\mu)$, where $g(\mu)$ was defined in (2.5).

We now will calculate $a_{(k)}(1^N)$, where $\mu = (1^N)$ is the totally antisymmetric representation on N indices. Substitution in (3.47) yields

$$F(k,j)(1^r) = j!\binom{r}{j}S(k,j). \tag{3.54}$$

Here the $S(k,j)$ are the ubiquitous Stirling numbers of the second kind, these being the number of ways of putting k distinct objects into j identical boxes allowing no box to be empty.[16] These numbers may be viewed as given by the defining relation

$$x^n = \sum_{l=0} S(n,l)(x)_l, \tag{3.55}$$

where $(x)_l$ is the falling factorial. In obtaining (3.54) we have used the identity

$$S(k,j) = \frac{1}{j!}\sum_{p=0}^{j}(-1)^p(j-p)^k\binom{j}{p}. \tag{3.56}$$

[Note that $S(0,0) = 1$ and $S(k,0) = 0$ for $k>0$.] Alternately (3.54) follows directly from (3.44) using the number of terms appearing in the symmetric monomial $m_\pi(1^N)$,

$$m_\pi(1^N) = \binom{N}{l(\pi)}\frac{l(\pi)!}{\pi_1!\cdots\pi_r!}. \tag{3.57}$$

In this case

$$S(k,j) = \sum_{\substack{|\pi|=k\\ l(\pi)=j}}\frac{k!}{(p_1!)^{\pi_1}\cdots(p_r!)^{\pi_r}\pi_1!\cdots\pi_r!} \tag{3.58a}$$

$$= \sum_{\substack{|\pi|=k\\ l(\pi)=j}}\frac{k!}{\gamma_1!\cdots\gamma_j!\pi_1!\cdots\pi_r!}. \tag{3.58b}$$

Using (3.56) we may express $a_{(k)}(1^N)$ in the following equivalent forms:

$$a_{(k)}(1^N) = \sum_{j=1}(-1)^{j-1}\binom{n-j}{N-j}S(k,j) \tag{3.59a}$$

$$= \sum_{l=0}^{k-1}(-1)^{l-1}l!\binom{n-1-l}{N-1-l}S(k-1,l) \tag{3.59b}$$

$$= \sum_{l=0}(-1)^{N+l-1}(N-l)^{k-1}\binom{n}{l} \tag{3.59c}$$

$$= \sum_{s=0}(-1)^s\binom{n-k}{N-1-s}a_{k-1,s}. \tag{3.59d}$$

In going from (3.59a) to (3.59b) we use the recurrence relation of the binomial coefficients and

$$S(k,j) = S(k-1,j-1) + jS(k-1,j). \tag{3.60}$$

Using the identity (3.55) we may readily show the equivalence of (3.59c) and (3.59b). In (3.59d) the numbers $a_{k-1,s}$ are the Eulerian numbers[17] given by

$$a_{n,l} = \sum_k(-1)^{k-l}\binom{k}{l}(n-k)!S(n,n-k). \tag{3.61}$$

Using this identity (3.59d) follows simply from (3.59b). [Be careful. The Eulerian numbers are variously tabulated as either $a_{n,l}$ or $A(n,l)$, where $A_{n,l} = A(n,l+1)$.] Expression (3.59c) is that given by Okubo and Patera[10] and (3.59c) and (3.59d) are given by Frampton and Kephart[18]; the others are new.

E. Examples

We have shown thus far that the anomaly coefficient may be written as

$$A_\pi(\lambda) = \sum_{\mu\in\pi^+(\lambda)} m_\lambda(\mu)a_\pi(\mu), \tag{3.62}$$

and we have given a formula (3.37) for $a_\pi(\mu)$ valid for su(n) when $|\pi|\leqslant n$. Further, we have given simple expressions for the coefficients $a_{(k)}(\mu)$. We will now focus on the leading anomaly coefficient set $A_k(\mu)\equiv A_{(k)}(\mu)$ and $a_k(\mu)\equiv a_{(k)}(\mu)$. Because we will later be interested in congruence properties of the $a_k(\mu)$ and $A_k(\mu)$ we have tabulated $a_k(\mu)$, $k = 2,\ldots,6$, for the 20 lowest-dimensional representations of su(4), su(5), and su(6). These are given in Tables I, II, and III, respectively. [More extensive tables including also su(7) to su(10) have been constructed but only mention of these will be made later.]

In calculating with (3.62) we mention the following useful result. Let $\bar\mu$ be the weight obtained from μ by the Dynkin diagram automorphism; viewed as representations

$\overline{\mu}$ gives the complex conjugate representation to μ. Then

$$a_k(\overline{\mu}) = (-1)^k a_k(\mu). \tag{3.63}$$

An analogous result obviously holds for $A_k(\mu)$.

We further observe that the weights summed over in (3.62) belong to the same congruence class. As the diagram automorphism sends the class m to $-m$ for su(n) we only need to keep track of this when k is odd. We have included this class label on our tables. Also we give the orbit size (O.S.) of the highest weight for the representation under consideration, the dimension of this representation and the coordinates of the weight following Ref. 8.

Let us compute some examples.

(a) We calculate $A_k(\yng(2,1,1))$ for su(5). Here $\mu = (0101)$. We note $\overline{\mu}$ is in the same class as $\yng(1)$. Further $m_{\yng(2,1)}(\overline{\yng(1)}) = 3$, $m_{\yng(2,1)}(\yng(2,1,1)) = 1$. Thus

$$A_2(\yng(2,1,1)) = a_2(\yng(2,1,1)) + 3a_2(\yng(1)) = 21 + 3 = 24,$$

$$A_3(\yng(2,1,1)) = +a_3(\yng(2,1,1)) - 3a_3(\yng(1)) = +6,$$

$$A_4(\yng(2,1,1)) = -9 + 3 = -6, \quad A_5(\yng(2,1,1)) = 66.$$

These agree with the known results.[19]

(b) We will now calculate $A_k(\mu)$ for various low-dimensional representations using (3.50)–(3.53). Further we will need

$$A_k(2,1^{p-1}) = a_k(2,1^{p-1}) + pa_k(1^{p+1}), \tag{3.64}$$

$$A_k(\yng(2)) = a_k(\yng(2)) + a_k(\yng(1,1)), \tag{3.65}$$

$$A_k(\yng(3)) = a_k(\yng(3)) + a_k(\yng(2,1)) + a_k(\yng(1,1,1)), \tag{3.66}$$

$$A_k(\yng(2,2)) = a_k(\yng(2,2)) + a_k(\yng(2,1,1)) + 2a_k(\yng(1,1,1,1)), \tag{3.67}$$

$$A_k(\yng(3,1)) = a_k(\yng(3,1)) + a_k(\yng(2,2)) + 2a_k(\yng(2,1,1)) + 3a_k(\yng(1,1,1,1)), \tag{3.68}$$

$$A_k(\yng(2,2,1)) = a_k(\yng(2,2,1)) + 2a_k(\yng(2,1,1,1)) + 5a_k(\yng(1,1,1,1,1)). \tag{3.69}$$

Here the multiplicities are easily obtained directly or from Ref. 12.

We have then

$$a_k(\yng(1)) = A_k(\yng(1)) = 1; \tag{3.70}$$

$$a_k(\yng(1,1)) = A_k(\yng(1,1)) = n - 2^{k-1}; \tag{3.71}$$

$$a_k(\yng(2)) = 2^k, \quad A_k(\yng(2)) = n + 2^{k-1}; \tag{3.72}$$

$$a_k(\yng(1,1,1)) = A_k(\yng(1,1,1)) = 1/2n^2 - 1/2n(1+2^k) + 3^{k-1}; \tag{3.73}$$

$$a_k(\yng(3)) = 3^k,$$

$$A_k(\yng(3)) = 1/2n^2 + 1/2n(1+2^k) + 3^{k-1}; \tag{3.74}$$

$$a_k(\yng(2,1)) = n(1+2^k) - 3^k, \quad A_k(\yng(2,1)) = n^2 - 3^{k-1}; \tag{3.75}$$

$$a_k(\yng(1,1,1,1)) = A_k(\yng(1,1,1,1)) = -2^{2k-2} + (n/6)\,[n^2 - 3n(1+2^{k-1}) + 2(1+3^k+2^{k-1})]; \tag{3.76}$$

$$a_k(\yng(2,1,1)) = n^2[1+2^{k-1}] - n[1+2^k+3^k] + 2^{2k},$$

$$A_k(\yng(2,1,1)) = 1/2n(n-2^{k-1})[n^2 - n - 2^k]; \tag{3.77}$$

$$a_k(\yng(2,2)) = 2^k(n - 2^{k-1}),$$

$$A_k(\yng(2,2)) = (n/3)\,(n^2 - [1 + 3^k - 3\cdot 2^{k-1}]); \tag{3.78}$$

$$a_k(\yng(3,1)) = n(1+3^k) - 2^{2k},$$

$$A_k(\yng(3,1)) = 1/2n(n+2^{k-1})[n^2 + n - 2^k]; \tag{3.79}$$

$$a_k(\yng(1,1,1,1,1)) = A_k(\yng(1,1,1,1,1)) = (n/24)\,[n^3 - n^2 2(2^k+3) + n(11 + 3\cdot 2^{k+1} + 4\cdot 3^k) - 6(1+2^k) - 4(3^k+2^k)] + 5^{k-1}; \tag{3.80}$$

$$a_k(\yng(5)) = 5^k,$$

$$A_k(\yng(5)) = (n/24)\,[n^3 + n^2\cdot 2(2^k+3) + n(11 + 3\cdot 2^{k+1} + 4\cdot 3^k) + 6(1+2^k) + 4(3^k+2^k)] + 5^{k-1}; \tag{3.81}$$

$$a_k(\yng(2,1,1,1)) = (n/6)\,[n^2(3+2^k) - n3(3+2^{k+1}+3^k) + 6(1+2^{2k}) + 5(2^k+3^k)] - 5^k,$$

$$A_k(\yng(2,1,1,1)) = (n/6)\,[n^3 - n^2(3+2^k) + n(2+3^k) + 2^k + 3^k] - 5^{k-1}; \tag{3.82}$$

$$a_k(\yng(4,1)) = n(1+2^{2k}) - 5^k,$$

$$A_k(\yng(4,1)) = (n/6)\,[n^3 + n^2(3+2^k) + n(2+3^k) - 2^k - 3^k] - 5^{k-1}; \tag{3.83}$$

$$a_k(\yng(2,2,1)) = (n^2/2)\,(1+2^{k+1}) - (n/2)\,(1 + 2^{k+1} + 2^{2k} + 2\cdot 3^k) + 5^k,$$

$$A_k(\yng(2,2,1)) = (n/24)\,[5n^3 - n^2\cdot 2(3+2^k) - n(4\cdot 3^k - 3\cdot 2^{k+1} + 5) + 6(1+2^{2k}) - 4(2^k+3^k)]; \tag{3.84}$$

$$a_k(\yng(3,2)) = n(3^k+2^k) - 5^k,$$

$$A_k(\yng(3,2)) = (n/24)\,[5n^3 + n^2\cdot 2(3+2^k) - n(4\cdot 3^k - 3\cdot 2^{k+1} + 5) - 6(1+2^{2k}) + 4(2^k+3^k)]; \tag{3.85}$$

$$a_k(\yng(3,1,1)) = (n/2)\,[n^2(2+3^k) - n(2^{2k+1}+2^k+2+3^k)] + 5^k,$$

$$A_k(\yng(3,1,1)) = 1/4n^2(n^2-1-2^{k+1}) + 5^{k-1}. \tag{3.86}$$

Here we have written explicitly the leading anomaly coefficients of su(n) for Young diagrams of length 5 or less. Several of these results may be found in Ref. 10; many, however, are generalizations or new.

IV. ANOMALOUS CONGRUENCES

A. General

We now have several explicit expressions for the anomaly coefficients, leading anomaly coefficients, and the associated polynomials $a_k(\mu)$. The latter have been evaluated for su(4), su(5), and su(6) in Tables I, II, and III, respectively for several weights μ corresponding to low-dimensional representations. These tables show several unexpected congruences. In particular, these tables [and the analogous ones for su(7) to su(10) not included here] show that, for all of the weights examined,

$$a_2 \equiv a_4 \bmod 6, \quad \mathrm{su}(n)\ n{:}4\text{–}10, \tag{4.1}$$

$$a_3 \equiv a_5 \bmod 12, \quad \mathrm{su}(n)\ n{:}5\text{–}10, \tag{4.2}$$

$$a_2 \equiv a_6 \bmod 30, \quad \mathrm{su}(n)\ n{:}6\text{–}10, \tag{4.3}$$

$$a_4 \equiv a_6 \bmod 24, \quad \mathrm{su}(n)\ n{:}6\text{–}10, \tag{4.4}$$

$$a_2 \equiv a_3 \bmod 2, \quad \mathrm{su}(n)\ n{:}4\text{–}10, \tag{4.5}$$

together with mod 2 congruences between those pairs not included in (4.1)–(4.5). We will show in this section that these congruences hold generally and show how to construct further ones.

These congruences also hold with a_k replaced by A_k. In fact, we have the following lemma.

Lemma: $A_\pi(\mu) \equiv A_{\pi'}(\mu) \bmod q\,(\forall \mu \in \Lambda^+)$ if and only if $a_\pi(\mu) \equiv a_{\pi'}(\mu) \bmod q\,(\forall \mu \in \Lambda^+)$.

Proof: This is obvious upon setting

$$A_\pi(\mu) = \sum_{\mu \in \pi^+(\lambda)} m_\lambda(\mu) a_\pi(\mu). \tag{4.6}$$

Here we have $m_\lambda(\lambda) = 1$ and (4.6) may be viewed as a change of basis with $m_\lambda(\mu)$ being a lower triangular matrix with integer entries 1 on the diagonal. This may be inverted over Z so that

$$a_\pi(\mu) = \sum m_\lambda(\mu)^{-1} A_\pi(\lambda). \tag{4.7}$$

Now the lemma is obvious.

We remark that there is a dependence of q on π and π'. Further, to establish these congruences it is often easier to work with the polynomials $a_\pi(\mu)$ than $A_\pi(\mu)$; this will be apparent from the su(2) example to be considered next.

In this section we will establish the identities (4.1)–(4.5) and the analogous identities

$$a_2 \equiv \tilde{a}_{(2^2)} \bmod 12, \quad \mathrm{su}(2),\ \mathrm{su}(3), \tag{4.8}$$

$$a_2 \equiv \tilde{a}_{(2^2)} \bmod 6, \quad G_2, \tag{4.9}$$

$$a_2 \equiv a_4 \bmod 3, \quad \mathrm{so}(7). \tag{4.10}$$

Note that in (4.8) and (4.9) there is no fourth-order Casimir in $U(L)$ and we define the coefficient $\tilde{a}$ for these groups by

$$\mathrm{Tr}_\lambda\, x^4 = \tilde{a}_{(2^2)} \mathrm{Tr}_{\lambda_{\rm gen}}\, x^4. \tag{4.11}$$

The fact there is no fourth-order Casimir means $\mathrm{Tr}_{\lambda_{\rm gen}}\, x^4 = c(\mathrm{Tr}_{\lambda_{\rm gen}}\, x^2)^2$, where c is a constant. We will be interested in the integers $\tilde{a}$.

We shall use several of these identities when we consider global anomalies in the next section. To establish these congruences it is helpful to consider su(2) first. This example will give us an upper bound on the q that appears in the leading anomaly congruences. After considering this example we will then establish the remaining congruences needed.

B. The su(2) example

Although this is the simplest possible example we will obtain some useful results by considering it in detail. Any $\mu \in \Lambda^+$ may be expressed as $\mu = n\lambda_1$, $n \in Z$. The polynomial $f_k(\mu)$ [(2.22) or (3.32)] then becomes

$$f_k(n\lambda_1) = (n^k/2)[1 + (-1)^k]\lambda_1^k = n^k f_k(\lambda_1), \tag{4.12}$$

which clearly vanishes unless k is even. In this case, we have

$$f_k(n\lambda_1) = n^k [f_2(\lambda_1)]^{k/2}, \tag{4.13}$$

and consequently [note that for su(2), $\tilde{a}_{(2^{k/2})} = 2^{k/2-1} a_{(2^{k/2})}$]

$$\tilde{a}_{(2^{k/2})}(n\lambda_1) = n^k, \tag{4.14}$$

$$\tilde{A}_{(2^{k/2})}(n\lambda_1) = \frac{1}{2} \sum_{j=0}^{n} (n-2j)^k. \tag{4.15}$$

Utilizing our lemma, if we can find q (depending on k and k') such that all

$$n^k \equiv n^{k'} \bmod q, \tag{4.16}$$

then $\tilde{a}_{(2^{k/2})}(\mu) \equiv \tilde{a}_{(2^{k'/2})}(\mu)$ for all μ. As we remarked earlier, this is somewhat easier than dealing with the $\tilde{A}_{(2^{k/2})}$'s directly, the first few of which are

$$\tilde{A}_{(2)}(n\lambda_1) = \tfrac{1}{6} n(n+1)(n+2), \tag{4.17}$$

$$\tilde{A}_{(2^2)}(n\lambda_1) = \tfrac{1}{30} n(n+1)(n+2)[3n^2+6n-4], \tag{4.18}$$

$$\tilde{A}_{(2^3)}(n\lambda_1) = \tfrac{1}{42} n(n+1)(n+2) \times [3n^4 + 12n^3 - 24n + 16]. \tag{4.19}$$

Given k and $k' \in N$ (we assume $k > k'$) we are interested then in finding the largest integer q such that (4.16) is true for all integers. To find such a q first recall that the Euler–Fermat theorem tells us that given a prime p and a number a such that $(a,p) = 1$, then $a^{\phi(p^\alpha)} \equiv 1 \bmod p^\alpha$, where $\phi(p^\alpha) = p^\alpha - p^{\alpha-1}$. The number $\phi(p^\alpha)$ need not be the least integer δ that $a^\delta \equiv 1 \bmod p^\alpha$ when $(a,p) = 1$. This number δ is called the order of $a \bmod p^\alpha$. Clearly $\delta | \phi(p^\alpha)$.

Now suppose q has the prime factorization $q = p_1^{\alpha_1} \cdots p_l^{\alpha_l}$. If we consider n to be of the form $n = mp_1$, where $(m,p_1) = 1$, then from (4.16) we have that $n^k \equiv n^{k'} \bmod p_1^{\alpha_1}$ and therefore $k' \geqslant \alpha$ for each p_i. On the other hand, if $(n,p_1) = 1$ then we have $n^{k-k'} \equiv 1 \bmod p_1^{\alpha_1}$ and so the exponent of n divides $k - k'$. Further, if $r = 2, 4, p^\alpha$, or $2p^\alpha$ (here

p is now an odd prime) then there exist $\phi(\phi(r))$ numbers whose exponent is $\delta(r)$.[20] (Such a number is known as a primitive root.) Therefore, we can determine the odd prime content of q by finding p_i and the maximum α_i subject to

$$\text{(a)}\ \phi(p_i^{\alpha_i})\,|\,k-k', \tag{4.20}$$

$$\text{(b)}\ k'\geqslant\alpha_i. \tag{4.21}$$

For the two-prime content of q we have that if $(2,a)=1$ and $\beta\geqslant 3$, then $a^{2^{\beta-2}}\equiv 1 \bmod 2^\beta$, that is, $a^{1/2\phi(2^\beta)}\equiv 1 \bmod 2^\beta$. We then wish to solve for

$$\begin{aligned}&\text{(a)}\ \phi(2^\alpha)\,|\,k-k' \quad\text{if}\quad \alpha\leqslant 2,\\ &\qquad 1/2\phi(2^\alpha)\,|\,k-k' \quad\text{if}\quad \alpha\geqslant 3,\end{aligned} \tag{4.22}$$

$$\text{(b)}\ k'\geqslant\alpha. \tag{4.23}$$

As examples of solving these equations we obtain

$$\begin{aligned}&n^2\equiv n^4 \bmod 12, \quad n^4\equiv n^6 \bmod 24,\\ &n^2\equiv n^6 \bmod 60, \quad n^4\equiv n^8 \bmod 240,\\ &n^2\equiv n^8 \bmod 252, \quad n^6\equiv n^8 \bmod 24.\end{aligned} \tag{4.24}$$

Although not relevant for the su(2) example we note further

$$\begin{aligned}&n^3\equiv n^5 \bmod 24, \quad n^2\equiv n^3 \bmod 2,\\ &n\equiv n^3 \bmod 6, \quad n\equiv n^5 \bmod 30,\\ &n\equiv n^4 \bmod 2, \quad n\equiv n^2 \bmod 2,\\ &n^3\equiv n^4 \bmod 2, \quad n^2\equiv n^5 \bmod 2.\end{aligned} \tag{4.25}$$

At this stage we have shown how to obtain the congruences applicable to the su(2) theory. This analysis, however, is relevant to the leading anomalies of su(n). From Eq. (3.49) we see that $a_{(k)}(\alpha)=\alpha^k$, where this weight corresponds to a Young diagram with one row of length α. Clearly then, if the congruence $a_{(k)}(\mu)\equiv a_{(k')}(\mu) \bmod q$ holds for all weights μ, it is true, in particular, for the restricted weights we have just mentioned and so q divides that number obtained from (4.20)–(4.23). Therefore, we have obtained an upper bound on the leading anomaly congruences. In the next subsection we will show how to reduce this upper bound to obtain (4.1)–(4.5).

C. SU(n) congruences

Our consideration of su(2) has given an upper bound on the congruences that appear for su(n). In turning now to higher su(n) we shall see how this number is reduced. Consider su(4) for definiteness. We have, from (3.50)–(3.52),

$$a_{(k)}(r)=r^k, \tag{4.26}$$

$$a_{(k)}(r,s)=\begin{cases}4(r^k+s^k)-(r+s)^k, & r\neq s,\\ 4r^k-r^k\cdot 2^{k-1}, & r=s,\end{cases} \tag{4.27}$$

$$\begin{aligned}a_{(k)}(r,s,t)&=16(r^k+s^k+t^k)\\ &\quad-4([r+s]^k+[s+t]^k+[t+r]^k\\ &\quad+r^k+s^k+t^k)+2(r^k+s^k+t^k),\\ a_{(k)}(r,r,t)&=a_{(k)}(t,r,r)\\ &=8(2r^k+t^k)-2(2^kr^k+2[r+t]^k\\ &\quad+2r^k+t^k)+(2r+t)^k,\\ a_{(k)}(r,r,r)&=8r^k+3^{k-1}r^k-2(2^k+1)r^k.\end{aligned} \tag{4.28}$$

As we have mentioned already, a necessary condition for $a_{(k)}(\mu)\equiv a_{(k')}(\mu) \bmod q$ for all μ is that $m^k\equiv m^{k'} \bmod q$ for all m. From (4.27) and (4.28) we see further that we require $2^{k-1}\equiv 2^{k'-1}$ and $3^{k-1}\equiv 3^{k'-1} \bmod q$. These conditions modify the constraints (4.21) and (4.23) to give

$$(b')\ k'-1\geqslant\alpha. \tag{4.29}$$

The effect of these constraints alters only the two-prime content of q in this example and we obtain (4.1).

These constraints are clearly sufficient for the su(4) case. In general, by considering $a_{(k)}(1^N)$ we require $N^{k-1}\equiv N^{k'-1} \bmod q$, and so for su($n$) we get the constraint (4.29) for those primes $p<n$. These additional constraints actually yield the congruences (4.1)–(4.5). These conditions can, in fact, be shown to be sufficient although this will be proved elsewhere. We note in passing that extra constraints can only arise from those representations μ that have rows of equal lengths appearing. We have then the following theorem.

Theorem: Necessary and sufficient conditions to find the largest q such that $a_{(k)}(\mu)\equiv a_{(k')}(\mu) \bmod q$ for all representations μ of su(n) and $k,k'\leqslant n$ are

(i) $m^k\equiv m^{k'} \bmod q$, for all $m\in Z$,

(ii) $p^{k-1}\equiv p^{k'-1} \bmod q$, for $p<n$.

Remark: We have shown how to construct this number q in Sec. IV B.

We conclude our considerations of su(n) congruences by proving the su(3) identity (4.8). Because our restriction of $|\pi|<n$ is violated, we must use a different approach. We note that if $\lambda=(m_1,m_2)$ is a highest weight vector for su(3) then

$$\begin{aligned}d(\lambda)&\equiv\dim V(\lambda)\\ &=\tfrac{1}{2}(m_1+1)(m_2+1)(m_1+m_2+2),\end{aligned} \tag{4.30}$$

$$c(\lambda)=\tfrac{2}{3}(m_1^2+m_2^2+m_1m_2+3m_1+3m_2), \tag{4.31}$$

$$\begin{aligned}l_2(\lambda)/2&=\tfrac{1}{24}(m_1+1)(m_2+1)(m_1+m_2+2)\\ &\quad\times(m_1^2+m_2^2+m_1m_2+3m_1+3m_2),\end{aligned} \tag{4.32}$$

where $c(\lambda)$ is the eigenvalue of the quadratic Casimir on $V(\lambda)$ and $l_2(\lambda)/2$ is the second Dynkin index. With

$$\mathrm{Tr}_\lambda\, x^2=a_2(\lambda)\mathrm{Tr}\, x^2, \tag{4.33}$$

$$\mathrm{Tr}_\lambda\, x^4=\tilde{a}_{(2^2)}(\lambda)\mathrm{Tr}\, x^4, \tag{4.34}$$

we see $a_2(\lambda)=l_2(\lambda)/2$ and may compute[9]

$$\tilde{a}_{(2^2)}(\lambda)=\tfrac{3}{10}[l_2(\lambda)/d(\lambda)][4l_2(\lambda)-d(\lambda)]. \tag{4.35}$$

In particular $\tilde{a}_{(2^2)}(\lambda)-a_{(2)}(\lambda)=12\Delta(m_1+1,m_2+1)/6!$, where

$$\begin{aligned}\Delta(a,b)&=ab(a+b)(a^2+b^2+ab-3)\\ &\quad\times(a^2+b^2+ab-7).\end{aligned} \tag{4.36}$$

Thus if $6!\,|\,\Delta(a,b)$ for all integers a and b, then the congruence (4.8) holds. To see this holds we note

$$\begin{aligned}\Delta(a,b)&=\Delta(a+1,b-1)\\ &\quad+(a-2)(a-1)a(a+1)(a+2)(a+3)\\ &\quad-(b-3)(b-2)(b-1)b(b+1)(b+2).\end{aligned} \tag{4.37}$$

Here each of the additional terms appearing is the product of six consecutive integers and so divisible by 6!. By induction the result follows as $\Delta(a,0)=0$. We have then proved

$$a_{(2)}(\lambda)\equiv\tilde{a}_{(2^2)}(\lambda)\quad \bmod 12\quad \mathrm{su}(3). \tag{4.38}$$

D. Other congruences

We now turn to establishing the congruences (4.9) and (4.10) for G_2 and SO(7). Although the general method we have outlined in Sec. II would work for these groups, we will not develop this here. Rather, we will work from some existing results of Okubo[9] to show these congruences.

First, consider G_2. Let $\lambda=(m_1,m_2)$ be a highest weight. It is then straightforward to show

$$\begin{aligned}d(\lambda)&\equiv\dim V(\lambda)\\&=(1/5!)(m_1+1)(m_2+1)\\&\quad\times(m_1+m_2+2)(2m_1+m_2+3)\\&\quad\times(3m_1+m_2+4)(3m_1+2m_2+5),\end{aligned} \tag{4.39}$$

$$c(\lambda)=2m_1^2+2m_1m_2+6m_1+\tfrac{10}{3}m_2+\tfrac{2}{3}m_2^2, \tag{4.40}$$

$$\begin{aligned}l_2(\lambda)/2&=[d(\lambda)/21]\\&\quad\times[3m_1^2+m_2^2+3m_1m_2+9m_1+5m_2].\end{aligned} \tag{4.41}$$

Here $c(\lambda)$ is the eigenvalue of the quadratic Casimir of $V(\lambda)$ and $l_2(\lambda)/r$ the second Dynkin index. Our conventions have $d(0,1)=7$, $d(1,0)=14$, $l_2(0,1)/2=2$, and $l_2(1,0)/2=8$.

Now let us write

$$\mathrm{Tr}_\lambda\, x^2=a_2(\lambda)\mathrm{Tr}_7\, x^2, \tag{4.42}$$

$$\mathrm{Tr}_\lambda\, x^4=\tilde{a}_{(2^2)}(\lambda)\mathrm{Tr}_7\, x^4. \tag{4.43}$$

Clearly $a_2(\lambda)=l_2(\lambda)/4$. Because $\mathrm{Tr}_7 x^4=c(\mathrm{Tr}_7\, x^2)^2$ we have $a_{(2^2)}(\lambda)=\tilde{a}_{(2^2)}(\lambda)c$; we choose to work with the integer $\tilde{a}_{(2^2)}(\lambda)$ in what follows. Utilizing the results of Ref. 9 it is not difficult to show

$$\begin{aligned}\tilde{a}_{(2^2)}(\lambda)=\tfrac14 a_{(2)}(\lambda)[3m_1^2+m_2^2+3m_1m_2\\+9m_1+5m_2-2]\end{aligned} \tag{4.44}$$

and in particular $\tilde{a}_{(2^2)}(\lambda)-a_{(2)}(\lambda)=\Delta(\lambda)/(4\times7!)$, where

$$\begin{aligned}\Delta(a-1,b-1)&=ab(a+b)(2a+b)(3a+b)(3a+2b)\\&\quad\times[3a^2+b^2+3ab-7][3a^2+b^2+3ab-13].\end{aligned} \tag{4.45}$$

Therefore, to show $\tilde{a}_{(2^2)}(\lambda)\equiv a_{(2)}(\lambda) \bmod 6$ we must show $3\times8!|\Delta(a,b)$ for all integers a and b. We verify this by checking that the prime powers of $2^7\cdot3^3\cdot5\cdot7=3\times8!$ divide $\Delta(a,b)$. For example,

$$\begin{aligned}\Delta(a,b)\equiv&-5ab^7+2ab^9+6a^2b^8+6a^3b^7\\&-2a^7b+5a^7b^3+a^8b^2+a^9b \bmod 7.\end{aligned} \tag{4.46}$$

Upon using the Euler–Fermat theorem [i.e., $a^6\equiv1 \bmod 7$ if $(a,7)=1$] we then see $\Delta(a,b)\equiv0 \bmod 7$ for all integers a and b. The other divisors may be checked similarly, and thus

$$a_{(2)}(\lambda)\equiv\tilde{a}_{(2^2)}(\lambda)\bmod 6,\quad \text{for } G_2. \tag{4.47}$$

Finally for (1.9) we begin by observing the general $SO(2n+1)$ result:

$$\mathrm{Tr}_{\mathrm{spinor}}\, x^4=-2^{n-4}\,\mathrm{Tr}_{\mathrm{vector}}\, x^4+\text{lower traces}. \tag{4.48}$$

Thus for SO(7) the spinor representation is the appropriate representation to consider the generating representation. As we mentioned in Sec. II both the vector and spinor of SO(7) have the same second Dynkin index. Therefore, we have

$$\begin{aligned}&a_{(2)}(\lambda_{\mathrm{spinor}})=a_{(2)}(\lambda_{\mathrm{vector}})=a_{(4)}(\lambda_{\mathrm{spinor}})=1,\\&a_{(4)}(\lambda_{\mathrm{vector}})=-2.\end{aligned}$$

These two representations show the equivalence (4.9) holds at most modulo 3. Using Okubo's results for the eigenvalues of the fourth-order index one may show this holds in general using techniques similar to those above.

V. APPLICATIONS

We will now apply the congruences as we have described in the previous section. The congruences obviously serve as a useful check when calculating higher-order anomaly coefficients. A further application we will make is to show the absence of G_2, su(2), and su(3) global gauge anomalies in six dimensions.[21] For simplicity we will work with a $(d=2n)$-dimensional Euclidean space-time topologically equivalent to a sphere, $M\cong S^{2n}$. The gauge groups G and H in the ensuing discussion are taken to be compact, semisimple Lie groups.

Let us consider an H gauge theory free of perturbative anomalies. Witten's arguments[22] alert us to the possibility of global anomalies if $\pi_{2n}(H)\neq0$. Assume further that $\pi_{2n}(H)\cong Z_l$ with generator $\bar{h}$. For many cases of interest this assumption is not too restrictive and we note $\pi_6(\mathrm{SU}(2))\cong Z_{12}$, $\pi_6(\mathrm{SU}(3))\cong Z_6$, $\pi_6(G_2)\cong Z_3$ are the only nontrivial six-dimensional homotopy groups for the H we are considering. The question at issue with global anomalies is whether the fermionic determinants $[\det i\not{D}(A_\mu)]^{1/2}$ and $[\det i\not{D}(A_\mu^{\bar{h}})]^{1/2}$ are identical for an H gauge connection A_μ and its gauge transform $A_\mu^{\bar{h}}$. General arguments show these differ by, at most, a phase.[23]

To investigate these global anomalies we use a technique frequently employed when studying topological field configurations: we embed the theory under consideration into a larger theory that can interpolate between the various topological sectors of the original theory. Questions previously not amenable to a perturbative study can now be examined within the perturbative framework of the larger theory.[24] In the present context Elitzur and Nair made early use of this approach.[25] We wish then a larger group G into which we may embed H such that $\pi_{2n}(G)=0$, this condition allowing us to interpolate between $\bar{h}$ and the identity of $\pi_{2n}(H)$. The difference in phase between the above determinants is then given by the Wess–Zumino term of the G theory. In order that we may evaluate this most simply in terms of differential forms we further assume that $\pi_{2n+1}(G)\cong Z$ with generator $\bar{g}$.

Under a finite gauge transformation $g\in G$ we have the fermionic measure in the path integral transforming as

$$d\mu(g\Psi) = d\mu(\Psi)\exp[i\Gamma_\Lambda(g,A,F)], \tag{5.1}$$

where $\Gamma_\Lambda(g,A,F)$ is the Wess–Zumino action for fermions in a G-representation Λ. Expressing Γ_Λ in terms of an integral over a $(2n+1)$-dimensional disk D with boundary $\partial D = M$ we have[1]

$$\Gamma_\Lambda(g,A,F) = 2\pi \int_D \gamma^\Lambda(g,A,F), \tag{5.2}$$

where

$$\gamma^\Lambda(g,A,F) = \omega^\Lambda_{2n+1}(A^g,F^g) - \omega^\Lambda_{2n+1}(A,F)$$
$$= c_{2n}\,\mathrm{Tr}_\Lambda(g^{-1}dg)^{2n+1} + d\alpha_{2n}, \tag{5.3}$$

$$\mathrm{Tr}_\Lambda F^{n+1} = d\omega^\Lambda_{2n+1}(A,F). \tag{5.4}$$

We are now interested in evaluating $\Gamma_\Lambda(k,A,F)$ for fixed, H-valued A and F and for some interpolating field k taking the value $\bar h$ on ∂D and the identity at the origin of D. Our assumption of the absence of any perturbative H-anomalies means $\gamma^\Lambda(g,A,F) = 0$ whenever $g\in H$. In particular, this means $k\in\pi_{2n+1}(G/H)$. Further, consideration of the exact sequence

$$0 \to H \xrightarrow{i} G \xrightarrow{\pi} G/H \to 0$$

yields

$$\begin{array}{ccccccccc}
\to \pi_{2n+1}(G) & \xrightarrow{\pi_*} & \pi_{2n+1}(G/H) & \xrightarrow{\partial} & \pi_{2n}(H) & \to & \pi_{2n}(G) & \to & \\
||| & & & & ||| & & ||| & & . \\
\to Z & \to & & \to & Z_l & \to & 0 & \to & \\
\bar g & & & & \bar h & & & &
\end{array} \tag{5.5}$$

From (5.5) we see k is a generator of $\pi_{2n+1}(G/H)$ (whenever this has a single generator) and that the image of k^l in $\pi_{2n+1}(G)$ [which exists as $\partial k^l = 0\in\pi_{2n}(H)$] is $\bar g$. [Strictly speaking, the image of k^l in $\Pi_{2n+1}(G)$ has the form $\bar g^p$ for some integer p, not necessarily 1. However, when $\pi_{2n+1}(G/H) = Z$, then $p=1$ by injectivity. For the later su(2) example we calculate $p=2$.] As these two gauge fields on S^{2n+1} only possibly differ by terms in H we have

$$\gamma^\Lambda(\bar g,A,F) = \gamma^\Lambda(k^l,A,F). \tag{5.6}$$

To complete the evaluation of $\Gamma_\Lambda(k,A,F)$ we must now discuss the representation dependence Λ implicit in the above. By De Rham's theorem and our assumption $\pi_{2n+1}(G)\cong Z$ we have for the generating representation (defined in Sec. II B) that

$$1 = \int_{S^{2n+1}} \gamma^{\lambda_{\rm gen}}(\bar g,A,F). \tag{5.7}$$

In particular, the additive properties of the Wess–Zumino term yield[25]

$$\Gamma_\Lambda(k,A,F) = 2\pi A_{n+1(\Lambda)}/l. \tag{5.8}$$

Actually, when $\pi_{2n+1}(G/H)$ has more than one generator this discussion needs to be somewhat modified with l replaced by l/p. We will return to this point later in this section; for the moment we will assume this is not the case.

The existence of global anomalies depends, therefore, on the divisibility properties of the leading anomaly coefficient $A_{n+1}(\Lambda)$ by $l = |\pi_{2n}(H)|$. To check whether global anomalies exist, the procedure is as follows. Given an H-gauge theory with fermion content Λ^H we embed H in G so that a representation Λ^G branches to Λ^H up to possible singlets (which have vanishing global and local anomaly). Actually, a judicious use of vectorlike representations (with vanishing local anomaly) is also frequently required. We then calculate $A_{n+1}(\Lambda^G)$ and see if it is divisible by l; if it is not, we have a global anomaly. The approach of Elitzur and Nair has reduced the problem to an algebraic one in determining $A_{n+1}(\Lambda^G)$ and then a number-theoretic one in checking its divisors. Before using our earlier results on A_{n+1} and its properties, two comments on this procedure are in order. First, there are potential embeddings of H in G and choices of Λ^H for which no Λ^G exists such that it branches to Λ^H and singlets. If we are to show a general H theory free of global anomalies, we must ensure such a branching always exists. Second, because we only require Λ^G to branch to Λ^H up to singlets, the choice of Λ^G need not be unique. [For example, in $\mathrm{SO}(7)\supset \mathrm{G}_2$ we have $7=7$ and $8=7+1$.] Again, we must check that this ambiguity does not lead to differing results.

We may now use our earlier results to show the absence of su(3) or G_2 anomlies in six dimensions. First, recall the additive property (2.7) of the second Dynkin index. For the standard embedding of su(n) in su($n+1$) as well as $\mathrm{so}(7)\supset \mathrm{G}_2$ the constant p appearing in this equation is unity. For su(3) the fact that we have a perturbative anomaly free theory yields

$$0 = \sum \widetilde A_{(2^2)}(\lambda_i^{\mathrm{su}(3)}) \tag{5.9}$$

$$\equiv \sum A_{(2)}(\lambda_i^{\mathrm{su}(3)}) \mod 12 \tag{5.10}$$

$$\equiv \sum A_{(2)}(\lambda_i^{\mathrm{su}(4)}) \mod 12 \tag{5.11}$$

$$\equiv \sum A_{(4)}(\lambda_i^{\mathrm{su}(4)}) \mod 6. \tag{5.12}$$

Here $\widetilde A_{(2^2)}$ is the analog of $\tilde a_{(2^2)}$ [see the discussion after (4.10)]. The sum over su(3) representations in (5.9) includes a sum over fermion chiralities as well, both being needed in six dimensions to get a vanishing perturbative anomaly. In going from (5.9) to (5.10) we have used (4.8); the sum in (5.11) is over those su(4) representations that branch to the su(3) representations and we have used (2.7) in obtaining this. Finally, to obtain (5.12) we use (4.1). Now we are done, because, since $\pi_6(\mathrm{su}(3))\cong Z_6$, our analysis (5.8) shows there is no anomaly.

The absence of G_2 anomalies is proved analogously. We have

$$0 = \sum \widetilde A_{(2^2)}(\lambda_i^{\mathrm{G}_2}) \equiv \sum A_{(2)}(\lambda_i^{\mathrm{G}_2}) \mod 6 \tag{5.13}$$

$$\equiv \sum A_{(2)}(\lambda_i^{\mathrm{so}(7)}) \mod 6 \tag{5.14}$$

$$\equiv \sum A_{(4)}(\lambda_i^{\mathrm{so}(7)}) \mod 3. \tag{5.15}$$

This, together with $\pi_6(G_2) \cong Z_3$, shows there are no six-dimensional G_2 anomalies.

We will now consider SU(2) anomalies in six dimensions and argue that none such exist. Let us note $\pi_6(SU(2)) \equiv Z_{12}$ and we consider the standard embeddings $SU(4) \supset SU(3) \supset SU(2)$. This yields $SU(4)/SU(2) \equiv S^5 \times S^7$ and thus $\pi_7(SU(4)/SU(2)) \equiv Z \oplus Z_2$. In this case $\pi_7(SU(4)/SU(2))$ has more than one generator and we will need to modify our calculation as we earlier cautioned. To see why this arises the following example is illustrative of our discussion. Consider the two SU(4) representations

$$\Lambda_1 = 6_L + 4_R + \bar{4}_R = 3_L + \bar{3}_L + 3_R + \bar{3}_R + 2\times 1$$
$$= 2\times 2_L + 2\times 2_R + 6\times 1,$$
$$\Lambda_2 = 4_L + \bar{4}_L + 4_R + \bar{4}_R = 3_L + \bar{3}_L + 3_R + \bar{3}_R + 4\times 1$$
$$= 2\times 2_L + 2\times 2_R + 8\times 1.$$

These have SU(4) anomalies $a(\Lambda_1^{SU(4)}) = -4-1-1 = -6$ and $a(\Lambda_2^{SU(4)}) = 0$, the latter being vectorlike. This ambiguity on the choice of Λ does not affect the SU(3) theory as we are only interested in the anomaly coefficient modulo 6. Without modifying this treatment for SU(2), however, we would be considering these coefficients modulo 12. Clearly, if this were the case, we would be faced with a Z_2 ambiguity. (Note our analysis has shown there is at most a Z_2 anomaly and thus the ambiguity here is the only one possible.) The resolution of this dilemma is, in fact, the additional Z_2 piece of $\pi_7(SU(4)/SU(2))$. The exact sequence (5.5) shows the Z part of $\pi_7(SU(4)/SU(2))$ has a mod 6 periodicity. In particular we see (without loss of generality)

$$\begin{array}{ccccccc} \to \pi_7(su(4)) & \to & \pi_7(su(4)/su(2)) & \to & \pi_6(su(2)) & \to & \\ ||| & & ||| & & & & \\ Z & \to & Z\oplus Z_2 & \to & Z_{12} & \to & . \\ 1 & \to & (6,1) & \to & 0 & & \end{array} \tag{5.16}$$

In particular we have $\mathrm{Im}\, \pi_{*|Z} = 6Z$. As it is, the Z piece yields the form content of γ^Λ; it is this congruence that is important. This last statement may be verified by use of the Hurewicz homomorphism. Alternately we are interested in the divisibility properties of the anomaly coefficient mod l/p, which here is $12/2 = 6$. We argue then that there are no SU(2) anomalies in six dimensions contrary to Ref. 26.

VI. DISCUSSION

To conclude let us review some of the results obtained and the ground covered. We began by giving a new procedure for expanding the trace polynomials

$$\mathrm{Tr}_\lambda\, x^k = \sum_\pi A_\pi(\lambda) P_\pi(x),$$

where $P_\pi(x)$ is a Weyl-invariant polynomial expressed in a given basis of $P(H)^W$. For each partition $|\pi| = k$ we have the anomaly coefficient $A_\pi(\lambda)$; these we expressed in the form

$$A_\pi(\lambda) = \sum_{\mu\in\pi^+(\lambda)} m_\lambda(\mu) a_\pi(\mu).$$

For su(n) we gave a general expression (3.37) for $a_\pi(\mu)$ subject only to $|\pi| \leqslant n$. In particular, various closed expressions were given for $a_{(k)}(\mu)$ in Sec. III D. To illustrate these results we calculated both $A_{(k)}(\mu)$ and $a_{(k)}(\mu)$ for representations whose Young diagrams had five or fewer boxes in Sec. III E.

A consequence of having tractable expressions for $a_\pi(\mu)$ at hand led us to observe a variety of congruences amongst these coefficients. In particular, we noted

$$A_\pi(\mu) \equiv A_{\pi'}(\mu) \bmod q, \quad \text{for all } \mu\in\Lambda^+$$
$$\Leftrightarrow a_\pi(\mu) \equiv a_{\pi'}(\mu) \bmod q, \quad \text{for all } \mu\in\Lambda^+.$$

This meant we could prove congruences among the anomaly coefficients by working with the (often) simpler $a_\pi(\mu)$. In particular, for su(n) we showed (for $k,k' < n$)

$$a_{(k)}(\mu) \equiv a_{(k')}(\mu) \bmod q, \quad \text{for all } \mu\in\Lambda^+$$
$$\Leftrightarrow m^k \equiv m^{k'} \bmod q, \quad \text{for all } m\in Z,$$
$$p^{k-1} \equiv p^{k'-1} \bmod q, \quad \text{for } p<n.$$

Further, we showed how to construct the maximum q subject to these conditions.

The congruences we have discovered can often be quite helpful. We mentioned their use in checking anomaly calculations as well as utilizing them to prove that there were no global gauge anomalies in six dimensions. In doing this calculation we believe several assumptions often made when applying the work of Elitzur and Nair have been clarified. The case of su(2) in six dimensions was particularly interesting.

Several extensions of this work ought now to be considered. Obviously, the extension to arbitrary L and the general congruences for a_π are called for. Another question raised but left unanswered was the nature of the set $F(L)$ defined in Sec. II B. Further, one can use this approach to examine higher-dimensional global anomalies. In particular, it is well suited to studying su(2) anomalies; this line of investigation will be presented elsewhere.

ACKNOWLEDGMENTS

I am grateful for discussions with P. Frampton, A. Mewborn, and J. Stasheff.

This work was partially supported by the U. S. Department of Energy Grant No. DE-FG05-85ER-40219.

[1] See, for example, the fine review by L. Alvarez-Gaume, "An introduction to anomalies," Erice lectures, 1985.
[2] A. Borel and F. Hirzebruch, Am. J. Math. **80**, 458 (1958); **81**, 315 (1959); **82**, 491 (1960).
[3] J. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1980).
[4] H. Hiller, *Geometry of Coxeter Groups* (Pitman, London, 1982).
[5] J. Patera, R. T. Sharp, and P. Winternitz, J. Math. Phys. **17**, 1972 (1976); **18**, 1519(E) (1977).
[6] E. B. Dynkin, Am. Math. Soc. Transl. Ser. 2. **6**, 111 (1957).
[7] W. McKay, J. Patera, and R. T. Sharp, J. Math. Phys. **22**, 2770 (1981).
[8] W. G. McKay and J. Patera, *Tables of Dimensions, Indices, and Branching Rules for Representation of Simple Lie Algebras* (Marcel Dekker, New York, 1981).
[9] S. Okubo, J. Math. Phys. **23**, 8 (1982).

[10]S. Okubo and J. Patera, J. Math. Phys. **24**, 2722 (1983); **25**, 219 (1984); S. Okubo, *ibid.* **26**, 2127 (1986).
[11]S. Okubo, J. Math. Phys. **18**, 2382 (1977).
[12]M. R. Bremner, R. V. Moody, and J. Patera, *Tables of Dominant Weight Multiplicities for Representations of Simple Lie Algebras* (Marcel Dekker, New York, 1985).
[13]A. N. Schellekens and N. P. Warner, Nucl. Phys. B **287**, 317 (1987).
[14]I. G. MacDonald, *Symmetric Functions and Hall Polynomials* (Oxford U.P., Oxford, 1979).
[15]F. N. David, M. G. Kendall, and D. E. Barton, *Symmetric Functions and Allied Tables* (Cambridge U.P., Cambridge, 1966).
[16]J. Comtet, *Advanced Combinatorics* (Reidel, Dordrecht, 1974); J. Riordan, *Combinatorial Identities* (Krieger, New York, 1979); V. Krishnamurthy, *Combinatorics: Theory and Applications* (Wiley, New York, 1986).
[17]J. Riordan, *An Introduction to Combinatorial Analysis* (Wiley, New York, 1958); D. Foata and M. Schutzenberger, *Theorie Geometrique des Polynomes Euleriens, Springer Lecture Notes in Mathematics*, Vol. 138 (Springer, New York, 1970).
[18]P. H. Frampton and T. W. Kephart, Phys. Rev. D **28**, 1010 (1983); Phys. Rev. Lett. **50**, 1347 (1983).
[19]R. Holman and T. W. Kephart, Phys. Lett. B **167**, 417 (1986).
[20]T. Nagell, *Introduction to Number Theory* (Chelsea, New York, 1964); G. M. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers* (Oxford U.P., Oxford, 1958); I. M. Vinogradov, *An Introduction to the Theory of Numbers* (Pergamon, London, 1955).
[21]H. W. Braden, Phys. Lett. B **200**, 91 (1988).
[22]E. Witten, Phys. Lett. B **117**, 324 (1982).
[23]L. Alvarez-Gaume and E. Witten, Nucl. Phys. B **234**, 269 (1983); J. M. Bismut and D. S. Freed, Commun. Math. Phys. **106**, 159 (1986); **107**, 103 (1986); D. S. Freed, *ibid.* **107**, 483 (1986).
[24]F. Wilczek and A. Zee, Phys. Rev. Lett. **51**, 2250 (1983).
[25]S. Elitzur and V. P. Nair, Nucl. Phys. B **243**, 205 (1984).
[26]E. Kiritsis, Phys. Lett. B **178**, 53 (1986); **181**, 416(E) (1986).

Second-degree kinematical constraints associated with dynamical symmetries

Mircea Iosifescu and Horia Scutaru
Department of Theoretical Physics, Central Institute of Physics, Bucharest, P.O. Box MG-6, Romania

(Received 5 May 1987; accepted for publication 21 October 1987)

The tensor operators that transform under subrepresentations of the symmetrized Kronecker square of the adjoint representation for the classical semisimple Lie algebras have been determined. All irreducible representations that satisfy the identities obtained by the vanishing of these tensor operators have been deduced. The pairs of irreducible representations of nonexceptional semisimple Lie algebras, the Kronecker product of which decomposes into two irreducible components, have been pointed out. As a consequence of this result, a comparison of the identities, derived in the way indicated before, with those resulting by the application of a method due to Hannabuss, Kostant, and Okubo, has been obtained. Connections between these results and those obtained by Drinfeld, concerning the solutions of quantum Yang–Baxter equations, have been established.

I. INTRODUCTION

A classical (or Poisson bracket) realization of a Lie algebra L is a realization of L as an algebra of differentiable functions defined on a classical phase space M, i.e., on a symplectic manifold M, with the Poisson bracket on M playing the role of Lie bracket; a quantum realization is a realization of L as an algebra of linear operators on a vector space V (the space of quantum states); the Lie bracket on End V is the commutator.

The properties of many physical systems (quantum or classical) can be described by a dynamical Lie algebra L: in the quantum case, the Hamiltonian and other important observables are realizations of elements of the enveloping algebra $U(L)$; in the classical case, they are realizations of elements in the algebra $P(L^*)$ of all polynomial functions on the vector space L^* dual to L. [$P(L^*)$ is isomorphic with the symmetric algebra $S(L)$ of L.]

From the point of view of physical applications the most interesting realizations are those determined by the smallest number of observables, i.e., by the smallest number of degrees of freedom. Realizations of this type are characterized by relations between their generators, i.e., by polynomial identities; otherwise stated, elements of $U(L)$ [of $P(L^*)$] vanish when realized in End V [in $C^\infty(M)$].

It has indeed long been known (cf., e.g., Refs. 1–12) that the generators of classical and of quantum realizations of Lie algebras satisfy, besides the Lie relations, a set of specific polynomial identities. A number of papers have been devoted to the derivation of such identities and to the explanation of their origin[13–26]; in several cases the identities have been used to construct[6] or to classify[9] representations of particular Lie algebras.

The ideology of the present paper has its origin in the observation[22] that the polynomial identities (discovered by Györgyi[7]) satisfied by the classical realization of the conformal algebra $so(4,2) \sim su(2,2)$ as a dynamical algebra of the Kepler Hamiltonian have definite tensorial character with respect to the extension of the adjoint representation of $so(4,2)$ to $P(so(4,2)^*)$.

Simple arguments allowed to conclude[23,25] that this result is general: polynomial identities satisfied by classical (quantum) realizations can be generated by decomposing $P(L^*)$ [$U(L)$] in irreducible modules, called elementary tensors, and by solving the equations—"tensorial identities"—obtained by requiring that the elementary tensors vanish in a given but unknown realization. Thus candidates for polynomial identities have to be found out first, and only subsequently must the realizations which satisfy them be determined; these are the nontrivial solutions of the tensorial identities. In particular, candidates for homogeneous tensorial identities of degree k for a Lie algebra L are obtained by equating to zero the tensors transforming under subrepresentations of the symmetric component of the kth Kronecker power $(\mathrm{ad}^{\otimes k})_s$ of the adjoint representation of L.

In the present paper we determine the second-degree tensors which transform under subrepresentations of $(\mathrm{ad} \otimes \mathrm{ad})_s$ for the classical Lie algebras

$$A_n \;(n \geqslant 3), \quad B_n \;(n \geqslant 2), \quad C_n \;(n \geqslant 2), \quad D_n \;(n \geqslant 5), \tag{1.1}$$

and deduce all representations on which these tensors vanish. In other terms, we determine all second-degree tensorial identities for the algebras (1.1) and find their solutions.

These second-degree tensorial identities are listed in Sec. II. They can be transformed into equations for the Dynkin indices which characterize the highest weights of the finite-dimensional irreducible representations of semisimple Lie algebras. The equations are solved in Sec. III; their solutions are listed in Table I and rederived in Appendix B by using the Wigner–Eckart theorem. These solutions are representations with maximal degeneracy, i.e., with all Dynkin indices but one equal to zero.

The representations which satisfy the second-degree tensorial identities for the Lie algebras of types B_n, C_n, and D_n can be characterized in the following way: for a given simple root α_i the representation with the highest weight $m\Lambda_i$ occurs in Table I for all integers m if the coefficient c_i of α_i in the decomposition of the maximal root $\alpha_{\max}$ is equal to

 0022-2488/88/040742-16$02.50

TABLE I. Second-degree tensor operators $T_{L,\Lambda}$ that transform under irreducible components with highest weight Λ of the representation $(\mathrm{ad} \otimes \mathrm{ad})_s$ of the nondegenerate semisimple Lie algebra L (column 3). Irreducible representations on which these tensor operators vanish (column 4). For the Lie algebras of types B_2 and B_3 column 3 has to be replaced as follows: for B_2, read $\Lambda_1(4\Lambda_2)$ instead of $\Lambda_4(2\Lambda_2)$; for B_3, read $2\Lambda_3$ instead of Λ_4.

Lie algebra L	Proposition number	Highest weight Λ of the representation under which $T_{L,\Lambda}$ transforms	Highest weight of representation ρ for which $T_{L,\Lambda}(\rho)=0$
A_n $n>3$	1	$\Lambda_1+\Lambda_n$	$m\Lambda_{(n+1)/2}$ $(m=1,2,...)$
	2	$\Lambda_2+\Lambda_{n-1}$	$m\Lambda_1, m\Lambda_n$ $(m=1,2,...)$
	3	$2\Lambda_1+2\Lambda_n$	Λ_k $(k=1,...,n)$
B_n $n>2$	4	$2\Lambda_1$	Λ_n
	5	Λ_4	$m\Lambda_1$ $(m=1,2,...)$
	6	$2\Lambda_2$	Λ_1, Λ_n
C_n $n>2$	7	Λ_2	$m\Lambda_n$ $(m=1,2,...)$
	8	$4\Lambda_1$	Λ_k $(k=1,...,n)$
	9	$2\Lambda_2$	Λ_1
D_n $n>5$	10	$2\Lambda_1$	$m\Lambda_{n-1}, m\Lambda_n$ $(m=1,2,...)$
	11	Λ_4	$m\Lambda_1$ $(m=1,2,...)$
	12	$2\Lambda_2$	$\Lambda_1, \Lambda_{n-1}, \Lambda_n$

1; it occurs only for $m=1$ if $c_i=(\alpha_{\max},\alpha_{\max})/(\alpha_i,\alpha_i)$. This property shows that the representations which appear in Table I as solutions of the second-degree tensorial identities for the Lie algebras of types B_n, C_n, and D_n are precisely those for which there exist solutions for the quantum Yang–Baxter equations, as pointed out recently by Drinfeld[27] (cf. Appendix D).

A method due mainly to Hannabuss,[16] Kostant,[28] and Okubo,[19] and which we shall call in the following the HKO method, allows one to construct, in principle, for a given representation ρ, polynomial identities satisfied by ρ. This method is based on the determination of the Kronecker products of ρ with auxiliary representations π which are chosen such that the corresponding Clebsch–Gordan series are multiplicity-free. The degree of the polynomial identity associated with a given auxiliary representation π is then equal to the number of terms in the Clebsch–Gordan series of $\rho \otimes \pi$.

Hence in order to obtain, by the HKO method, the second-degree polynomial identities satisfied by a given representation ρ of the Lie algebra L we must find all auxiliary representations π of L such that $\rho \otimes \pi$ decomposes in precisely two terms $\rho \otimes \pi = \Theta \oplus \omega$. We shall call an auxiliary representation π with this property an Okubo partner. We determined the Okubo partners for the representations from Table I and found (Sec. IV) that for the Lie algebras of types B_n, C_n, and D_n these partners exist only for those representations with highest weights of types $k\Lambda_i$ $(k=1,2,...)$ which appear in Table I (see Table II).

These representations are those for which there exist classical limits, i.e., classical realizations which satisfy in $P(L^*)$ polynomial identities corresponding to the same tensor. This correspondence between the classical and quantum realizations has in fact been employed by Reshetikhin in his construction of the solutions for classical and quantum Yang–Baxter equations[29,30] (cf. also Ref. 31). These solutions are classified by the compact Hermitian symmetric spaces which are the coadjoint orbits described by the polynomial identities in $P(L^*)$.[30]

In Sec. IV we derive the second-degree polynomial identities satisfied by the representations ρ for which Okubo partners exist and observe that, for algebras of types B_n and D_n, they coincide with the identities considered in Secs. II and III. Complete proofs for this coincidence are provided for the algebras of type D_n.

TABLE II. Elements that characterize the second-degree minimal polynomials satisfied by HKO operators of nonexceptional semisimple Lie algebras (1.1). The highest weights of the fundamental representations are denoted by Λ_i $(i=1,2,...,n)$.

Lie algebra L	Pairs of weights λ,μ for which the product decomposes into a direct sum of two irreducible representations		The weights ω that appear in the decomposition of the product $\rho_\lambda \otimes \rho_\mu$		The roots of the second-degree minimal polynomial (4.6)	
	λ	μ	$\lambda+\mu$	ξ	(λ,μ)	$\frac{1}{2}[(\xi+2\delta,\xi)-(\lambda+2\delta,\lambda)-(\mu+2\delta,\mu)]$
A_n	$m\Lambda_1$	Λ_k	$m\Lambda_1+\Lambda_k$	$(m-1)\Lambda_1+\Lambda_{k+1}$	$m(1-k/(n+1))$	$-k(1+m/(n+1))$
	Λ_1	$m\Lambda_k$	$\Lambda_1+m\Lambda_k$	$(m-1)\Lambda_k+\Lambda_{k+1}$	$m(1-k/(n+1))$	$-k(1+m/(n+1))$
	$m\Lambda_n$	Λ_{n-k+1}	$m\Lambda_n+\Lambda_{n-k+1}$	$(m-1)\Lambda_n+\Lambda_{n-k}$	$m(1-k/(n+1))$	$-k(1+m/(n+1))$
	Λ_n	$m\Lambda_{n-k+1}$	$\Lambda_n+m\Lambda_{n-k+1}$	$(m-1)\Lambda_{n-k+1}+\Lambda_{n-k}$	$m(1-k/(n+1))$	$-k(1+m/(n+1))$
B_n	$m\Lambda_1$	Λ_n	$m\Lambda_1+\Lambda_n$	$(m-1)\Lambda_1+\Lambda_n$	$m/2$	$(-m-2n+1)/2$
C_n	Λ_1	$m\Lambda_n$	$\Lambda_1+m\Lambda_n$	$\Lambda_{n-1}+(m-1)\Lambda_n$	m	$-(m+n+1)$
D_n	$m\Lambda_1$	Λ_{n-1}	$m\Lambda_1+\Lambda_{n-1}$	$(m-1)\Lambda_1+\Lambda_n$	$m/2$	$(-m-2n+2)/2$
	$m\Lambda_1$	Λ_n	$m\Lambda_1+\Lambda_n$	$(m-1)\Lambda_1+\Lambda_{n-1}$	$m/2$	$(-m-2n+2)/2$
	Λ_1	$m\Lambda_{n-1}$	$\Lambda_1+m\Lambda_{n-1}$	$(m-1)\Lambda_{n-1}+\Lambda_n$	$m/2$	$(-m-2n+2)/2$
	Λ_1	$m\Lambda_n$	$\Lambda_1+m\Lambda_n$	$\Lambda_{n-1}+(m-1)\Lambda_n$	$m/2$	$(-m-2n-2)/2$

For the algebras of types C_n and A_n, several differences between our results and those obtained by the HKO method have to be noted. For the Lie algebras of type C_n, the representations $\Lambda_2,\Lambda_3,\ldots,\Lambda_{n-1}$ satisfy second-degree polynomial identities, which cannot be obtained by the HKO method because these representations have no Okubo partners. On the other side, for the Lie algebras of type A_n, there exist identities obtained by the HKO method which can be derived by our method only if we allow the adjoint tensors to be written as linear combinations between tensors of second and first degrees.

As physical applications are concerned, our results may offer a unification of many disparate results, casting in a purely algebraic framework physical systems with the same dynamical Lie algebra which satisfy the same kinematical constraints.

The kinematical constraints may be viewed as constraints imposed to a dynamical system by its symmetry properties. So far, kinematical constraints were obtained by direct inspection from the algebraic peculiarities of the various concrete realizations. The most notable compact Hermitian symmetric spaces which appear in several physical models in field theory or in nuclear physics are the spaces $SO(2n)/U(n)$, considered by Berezin[32] and Papanicolaou[33] in connection with the Gross–Neveu model and by Yamamura[34] and Nishiyama[35] in connection with the theory of nuclear collective motion. The noncompact Hermitian spaces $Sp(2n,R)/U(n)$ appear in the papers of Mlodinow and Papanicolaou[36] and of Deenen and Quesne[37] again in connection with the theory of collective motion in atoms and nuclei, respectively.

An immediate application of the second-degree kinematical constraints is the derivation of the Holstein–Primakoff realizations[33,37] by using a purely algebraic method that takes advantage of these constraints.[38]

A new field of possible applications of the tensorial identities is the characterization of "models" of representations, defined by Bernstein, Gelfand, and Gelfand[39] and discovered by Biedenharn and Flath for su(3).[40] Indeed, Bracken pointed out[41] that the infinite-dimensional representations of so(6,2), which he introduced as a model for su(3) representations, satisfies second-degree polynomial identities. It is easy to prove that these identities, which characterize this model of su(3), are equations of the type (2.15) derived and studied in the present paper.

It is evident that the second-degree tensorial identities do not exhaust the functionally independent identities satisfied by representations of the Lie algebras (1.1). The number of functionally independent tensors which vanish on a fixed representation is, however, finite.

The determination of all independent tensorial identities for a given semisimple Lie algebra is a problem complementary to the determination of the integrity basis in its degenerate enveloping algebras.[42–44] The set of independent tensorial identities satisfied by a given representation ρ may be viewed as a set of conditions which characterize ρ; their complete determination could provide an alternative for the identification of linear representations of semisimple Lie algebras.

II. TENSORIAL IDENTITIES

A. The existence of tensorial identities

Let us denote by A an associative algebra endowed also with a Lie product denoted by [,] and which has the following derivation property with respect to the associative product:

$$[a,bc] = [a,b]c + b[a,c]. \tag{2.1}$$

A realization of a Lie algebra L in A is a Lie algebra homomorphism $h: L \to A$. Any such realization can be extended in a unique way to a homomorphism of associative algebras $\tilde{h}$: $U(L) \to A$ with the property $\tilde{h}(1) = 1_A$, where 1_A is the unit element of A and $U(L)$ is the enveloping algebra of L.

If for an element $u \in U(L)$ we have $\tilde{h}(u) = 0$ then we shall say that the realization $h: L \to A$ satisfies the polynomial identity corresponding to the element $u \in U(L)$. The set

$$\ker \tilde{h} = \{u \in U(L);\ \tilde{h}(u) = 0\} \tag{2.2}$$

is then the set of all polynomial identities satisfied by the realization $h: L \to A$. Because $\tilde{h}(ab) = \tilde{h}(a)\tilde{h}(b)$ ($\tilde{h}$ is a homomorphism of associative algebras) it follows that $\ker \tilde{h}$ is a two-sided ideal of $U(L)$ and hence that it is an invariant subspace with respect to the unique extension of the adjoint representation of L on $U(L)$ defined by

$$\widetilde{\mathrm{ad}}(x)u \equiv xu - ux. \tag{2.3}$$

This representation is completely reducible (Ref. 45, §2.3.3).

The decomposition of this representation into irreducible components has been described by Kostant.[28] A given representation with the highest weight Λ ($\neq 0$) appears as a subrepresentation in $\widetilde{\mathrm{ad}}$ as many times as the multiplicity of the zero weight in the representation Λ. Also, among these subrepresentations of type Λ there exists one of the highest degree $\mathrm{hd}(\Lambda)$ equal with the sum of the coefficients of the decomposition of Λ as a linear combination of simple roots. All representations of type Λ of higher degrees are obtained from the representations of type Λ of degree at most equal with $\mathrm{hd}(\Lambda)$ by multiplication with elements from the subrepresentations of $\widetilde{\mathrm{ad}}$ with zero highest weight, i.e., with invariant elements or Casimir elements (Ref. 45, §8.3.11). The adjoint representation ad of L can be also extended in a unique way to a representation $\widetilde{\widetilde{\mathrm{ad}}}$ of L on the symmetric algebra $S(L)$ of L. The symmetrization map σ: $S(L) \to U(L)$ intertwines the two adjoint representations $\widetilde{\mathrm{ad}}$ and $\widetilde{\widetilde{\mathrm{ad}}}$, i.e., $S(L)$ and $U(L)$ are equivalent as L modules (Ref. 45, §2.4.10). From this equivalence it follows that the restriction of $\widetilde{\mathrm{ad}}$ to the subspace $U(L)_k$ of homogeneous elements of $U(L)$ of degree k is equivalent with the symmetric part of the Kronecker power k of ad, denoted by $(\mathrm{ad}^{\otimes k})_s$.

Let us denote by L^* the dual of the vectorial space L and by $P(L)$ the associative commutative algebra of all polynomials on L^*. The algebra $P(L^*)$ is generated by 1 and the first-degree polynomials $x(l)$ defined by $x(l) = l(x)$ for any $x \in L$ and any $l \in L^*$; as a vectorial space $P(L^*)$ is isomorphic with $S(L)$. It follows that the Lie bracket { , } defined for the zero and first-degree polynomials by $\{1,x(l)\} = 0$ and $\{x,y\} = l([x,y])$ can be extended to all polynomials from $P(L^*)$. The adjoint representation of L on $P(L^*)$ defined by

$(\mathrm{ad}(x)p)(l) = \{x,p\}(l)$ is also equivalent with the adjoint representation of L on $S(L)$; the intertwining operator is precisely the map $(x\otimes y\otimes\cdots\otimes z)_s \to x(l)y(l)\cdots z(l)$ which associates with any element of $S(L)$ a polynomial on L^*.

In the present work, we study the decomposition of the second-degree homogeneous component of the two-sided ideal $\ker\tilde{h}$. This amounts to determining the irreducible subrepresentations of $(\mathrm{ad}\otimes\mathrm{ad})_s$. The irreducible tensors which transform under these representations have been determined in Ref. 24 for the semisimple nonexceptional Lie algebras (1.1), using a projection method. The algebras (1.1) have in common the property that, for each of them, the Clebsch–Gordan series for $(\mathrm{ad}\otimes\mathrm{ad})_s$ is multiplicity-free and contains precisely four terms.[46,47] For reasons of convenience, the calculations have been performed in $P(L^*)$.

To each irreducible tensor $T_{L\Lambda}$ in $P(L^*)$ [in $U(L)$], transforming under a representation $\rho_\Lambda \subset (\mathrm{ad}\otimes\mathrm{ad})_s$ of highest weight Λ, a tensorial identity can be associated by equating $T_{L\Lambda}$ to zero in a classical (quantum) realization to be determined. Nontrivial realizations on which all components of $T_{L\Lambda}$ vanish will be called "solutions" of the tensorial identity $T_{L\Lambda}=0$. As already stated, our aim is to determine the quantum tensorial identities (Sec. II B) and to solve them (Sec. III).

B. The expressions of the second-degree tensorial identities

In this section we list the second-degree classical tensorial identities for all semisimple Lie algebras (1.1). Quantum tensorial identities are obtained from classical ones by replacing in Eqs. (2.8)–(2.10), (2.14)–(2.16), and (2.21)–(2.23) each product of two generators by their anticommutator, e.g.,

$$A_{pq}A_{rs} \to [A_{pq},A_{rs}]_+ \equiv A_{pq}A_{rs} + A_{rs}A_{pq}. \qquad (2.4)$$

Tensor components are identified by four labels since each component is projected from a generic second-degree monomial of the symmetric algebra. For particular tensors, however, two labels suffice. We denote by $\Lambda_1,\Lambda_2,\ldots,\Lambda_n$ the highest weights of the fundamental representations of a semisimple Lie algebra of rank n and by Λ_{ad} the highest weight of the adjoint representation.

1. Algebras of type A_n

The generators A_{ij} $(i,j=1,2,\ldots,n+1)$ of the algebra $\mathrm{sl}(n+1,C)$ satisfy the structure relations

$$[A_{ij},A_{kl}] = \delta_{jk}A_{il} - \delta_{il}A_{kj}. \qquad (2.5)$$

For algebras of type A_n, $\Lambda_{\mathrm{ad}} = \Lambda_1+\Lambda_n$ and we have $(\rho_{\Lambda_{\mathrm{ad}}}\equiv\mathrm{ad})$

$$(\mathrm{ad}\otimes\mathrm{ad})_s = \rho_{(0)}\oplus\rho_{\Lambda_1+\Lambda_n}\oplus\rho_{\Lambda_2+\Lambda_{n-1}}\oplus\rho_{2\Lambda_1+2\Lambda_n}, \qquad (2.6)$$

where ρ_Λ is a representation with highest weight Λ. Denoting

$$\mathscr{A}_{pq}(\lambda) \equiv \sum_{i=1}^{n+1} A_{pi}A_{iq} - \frac{\delta_{pq}}{\lambda}\sum_{i,j=1}^{n+1} A_{ij}A_{ji}, \qquad (2.7)$$

we obtain the classical tensorial identities

$$T_{\Lambda_1+\Lambda_n}(p,q) = \mathscr{A}_{pq}(n+1) = 0 \quad (p,q=1,2,\ldots,n+1), \qquad (2.8)$$

$$T_{\Lambda_2+\Lambda_{n-1}}(p,q,r,s) = A_{pq}A_{rs} - A_{ps}A_{rq} + [1/(n-1)]\{-\delta_{qr}A_{ps}(2n) - \delta_{ps}\mathscr{A}_{rq}(2n) + \delta_{pq}\mathscr{A}_{rs}(2n) + \delta_{rs}\mathscr{A}_{pq}(2n)\} = 0, \qquad (2.9)$$

$$T_{2\Lambda_1+2\Lambda_n}(p,q,r,s) = A_{pq}A_{rs} + A_{ps}A_{rq} - [1/(n+3)]\{\delta_{qr}\mathscr{A}_{ps}(2(n+2)) + \delta_{ps}\mathscr{A}_{rq}(2(n+2)) + \delta_{pq}\mathscr{A}_{rs}(2(n+2)) + \delta_{rs}\mathscr{A}_{pq}(2(n+2))\} = 0 \quad (p,q,r,s,=1,2,\ldots,n+1). \qquad (2.10)$$

Let us mention that $T_{(0)} = \sum_{i,j=1}^{n+1} A_{ij}A_{ji}$.

2. Algebras of types B_n and D_n

The generators M_{ij} $(i,j=1,2,\ldots,N)$ of the algebras $\mathrm{so}(N,C)$ ($N=2n+1$ for the type B_n and $N=2n$ for the type D_n) satisfy the structure relations

$$[M_{ij},M_{kl}] = \delta_{il}M_{jk} + \delta_{jk}M_{il} - \delta_{ik}M_{jl} - \delta_{jl}M_{ik} \qquad (2.11)$$

with $M_{ji} = -M_{ij}$ and $i,j,k,l=1,2,\ldots,N$.

For the algebras of type B_2, $\Lambda_{\mathrm{ad}} = 2\Lambda_2$; for the algebras of types B_n $(n\geqslant 3)$ and D_n $(n\geqslant 4)$, $\Lambda_{\mathrm{ad}} = \Lambda_2$. The symmetric part of the Kronecker square of the adjoint representation, $(\mathrm{ad}\otimes\mathrm{ad})_s$, decomposes into four irreducible components in the following cases:

type B_2: $$(\rho_{2\Lambda_2}\otimes\rho_{2\Lambda_2})_s = \rho_{(0)}\oplus\rho_{2\Lambda_1}\oplus\rho_{4\Lambda_2}\oplus\rho_{\Lambda_1}, \qquad (2.12a)$$

type B_3: $$(\rho_{\Lambda_2}\otimes\rho_{\Lambda_2})_s = \rho_{(0)}\oplus\rho_{2\Lambda_1}\oplus\rho_{2\Lambda_2}\oplus\rho_{2\Lambda_3}, \qquad (2.12b)$$

types B_n $(n\geqslant 3)$ and D_n $(n\geqslant 5)$:

$$(\rho_{\Lambda_2}\otimes\rho_{\Lambda_2})_s = \rho_{(0)}\oplus\rho_{2\Lambda_1}\oplus\rho_{2\Lambda_2}\oplus\rho_{\Lambda_4}. \qquad (2.12c)$$

We shall give the expressions of the tensorial identities associated with the nontrivial representations in the decomposition (2.12c). The results for the algebras of type B_2 will be obtained by replacing Λ_4 by Λ_1 and $2\Lambda_2$ by $4\Lambda_2$; the results for the type B_3 are obtained replacing in the generic case Λ_4 by $2\Lambda_3$.

Denoting

$$\mathscr{M}_{pq}(\lambda) \equiv \sum_{i=1}^{N} M_{pi}M_{iq} - \frac{\delta_{pq}}{\lambda}\sum_{i,j=1}^{N} M_{ij}M_{ji}, \qquad (2.13)$$

we obtain the following identities associated with irreducible subrepresentations of $(\rho_{\Lambda_{ad}} \otimes \rho_{\Lambda_{ad}})_s$:

$$T_{2\Lambda_1}(p,q) = \mathscr{M}_{pq}(N) = 0 \quad (p,q = 1,2,...,N), \tag{2.14}$$

$$T_{\Lambda_4}(p,q,r,s) = M_{pq}M_{rs} + M_{ps}M_{qr} + M_{pr}M_{sq} = 0 \quad (p,q,r,s, = 1,...,N), \tag{2.15}$$

$$\begin{aligned} T_{2\Lambda_2}(p,q,r,s) = {} & \tfrac{1}{3}(2M_{pq}M_{rs} - M_{ps}M_{qr} - M_{pr}M_{sq}) \\ & + [1/(N-2)][-\delta_{qr}\mathscr{M}_{ps}(2(N-1)) \\ & - \delta_{ps}\mathscr{M}_{qr}(2(N-1)) \\ & + \delta_{pr}\mathscr{M}_{qs}(2(N-1)) \\ & + \delta_{qs}\mathscr{M}_{pr}(2(N-1))] = 0. \end{aligned} \tag{2.16}$$

We have $T_{(0)} = \Sigma_{i,j=1}^{N} M_{ij}M_{ji}$. Let us obseve that the tensor (2.15) can be also expressed as

$$T_{\Lambda_4}(i,j) = \sum_{p,q,r,s} \epsilon_{ijpqrs} M_{pq}M_{rs}, \tag{2.15'}$$

where ϵ_{ijpqrs} is the six-dimensional antisymmetric tensor.[7,48]

3. Algebras of type C_n

Let $g_{ij} = \delta_{i,j+n} - \delta_{i+n,j}$ $(i,j = 1,...,2n)$ and let

$$S_{ij} = \sum_{k=1}^{2n} (g_{ik}e_{kj} - g_{kj}e_{ki}) \quad (i,j = 1,...,2n) \tag{2.17}$$

be the generators of the algebra sp$(2n,C)$ $(S_{ij} = S_{ji})$ with the structure relations[49]

$$[S_{ij}, S_{kl}] = g_{kj}S_{il} - g_{il}S_{kj} - g_{ik}S_{jl} + g_{lj}S_{ki}. \tag{2.18}$$

For algebras of type C_n $(n \geqslant 2)$ we have $\Lambda_{ad} = 2\Lambda_1$ and

$$(\text{ad} \otimes \text{ad})_s = \rho_{(0)} \oplus \rho_{\Lambda_2} \oplus \rho_{4\Lambda_1} \oplus \rho_{2\Lambda_2}. \tag{2.19}$$

Denoting

$$\mathscr{S}_{pq}(\lambda) \equiv \sum_{i,j=1}^{2n} g_{ji}S_{ip}S_{jq} - \frac{g_{pq}}{\lambda} \sum_{i,j,k,l=1}^{2n} g_{ji}g_{lk}S_{il}S_{jk} \tag{2.20}$$

we obtain the following identities associated with subrepresentations of $(\text{ad} \otimes \text{ad})_s$:

$$T_{\Lambda_2}(p,q) = \mathscr{S}_{pq}(2n) = 0 \quad (p,q = 1,2,...,2n), \tag{2.21}$$

$$T_{4\Lambda_1}(p,q,r,s) = S_{pq}S_{rs} + S_{ps}S_{rq} + S_{pr}S_{qs} = 0, \tag{2.22}$$

$$\begin{aligned} T_{2\Lambda_2}(p,q,r,s) = {} & \tfrac{1}{3}(2S_{pq}S_{rs} - S_{ps}S_{rq} - S_{pr}S_{qs}) \\ & + [1/2(n+1)][g_{ps}\mathscr{S}_{qr}(2(2n+1)) \\ & + g_{pr}\mathscr{S}_{qs}(2(2n+1)) \\ & + g_{qr}\mathscr{S}_{ps}(2(2n+1)) \\ & + g_{qs}\mathscr{S}_{pr}(2(2n+1))] = 0 \\ & (p,q,r,s, = 1,2,...,2n). \end{aligned} \tag{2.23}$$

We have also $T_{(0)} = \Sigma_{i,j,k,l=1}^{2n} g_{ij}g_{kl}S_{il}S_{jk}$.

The expressions of the tensors in Eqs. (2.10), (2.14), and (2.21) have been obtained also by Jarvis.[47]

III. DETERMINATION OF THE FINITE-DIMENSIONAL IRREDUCIBLE REPRESENTATIONS THAT SATISFY THE QUANTUM TENSORIAL IDENTITIES

As already mentioned, the quantum tensorial identities are obtained from the classical ones by symmetrizing each product of generators with respect to order [cf. (2.41)]. We shall refer, in the following, *only* to quantum identities, without further specification, quoting them by the same numbers as the corresponding classical ones.

The quantum tensorial identities (2.8)–(2.10), (2.14)–(2.16), and (2.21)–(2.23) contain information about the representations which satisfy them. To extract part of this information is the aim of the present section in which we shall determine the weights of the finite-dimensional irreducible representations for which the tensor operators vanish. The results of this analysis are summarized by the theorem that follows.

Theorem 1: Let L be one of the nonexceptional semisimple Lie algebras (1.1); let $T_{L,\Lambda}(x_1,...,x_{\dim L}) \in \mathrm{U}(L)$ (x_i = generators of L) be the second-degree tensor operator which transforms under the subrepresentation of $(\text{ad} \otimes \text{ad})_s$ with the highest weight Λ of the Lie algebra L, and let $\rho(x_i)$ $(i = 1,...,\dim L)$ be the generators of the representation ρ of L.

The finite-dimensional irreducible representations ρ of the Lie algebras (1.1) for which

$$T_{L,\Lambda}(\rho) = T_{L,\Lambda}(\rho(x_1), \rho(x_2), ..., \rho(x_{\dim L})) = 0$$

are those listed in Table I.

The theorem stated above is composed of 12 propositions (labeled from 1 to 12); the meaning of proposition number p is: "In line number p of Table I column 3 implies column 4."

Remark 1: We read from Table I that all second-degree tensorial identities of the Lie algebras (1.1) admit solutions which are representations with maximum degeneracy. This is not surprising: Okubo pointed out[18] that the degree of the identity satisfied by a representation ρ of su(n) decreases with the increase of the degeneracy of ρ.

Remark 2: A number of solutions of the tensorial identities present themselves as a series of representations, the highest weights of which are all integer multiples of the highest weight of a given fundamental representation (e.g., $m\Lambda_1$; $m = 1,2,...$). Solutions of this type of the quantum identities admit classical limits; the identity satisfied by the classical limit corresponds to the same tensor as the quantum identity.

Remark 3: A synthetic characterization of the solutions contained in column 4 of Table I can be obtained by giving an explicit formulation to a result obtained by Drinfeld.[27] For details, see Appendix D.

The proof of the theorem—which is based on Remark 5—consists in the straightforward determination of the highest weights of the solutions of the tensorial identities.

Remark 4: The solutions of the tensorial identities have been reobtained by a proof which makes use of the Wigner–Eckart theorem and which is to be found in Appendix B.

Remark 5: Let ρ be a finite-dimensional irreducible representation of the semisimple Lie algebra L, acting on the L module V_ρ. Let $T_{L,\Lambda}(\rho)$ be a tensor operator, the components

$$T_{L,\Lambda,j}(\rho(x_1),\rho(x_2),\ldots,\rho(x_{\dim L}))\quad (j=1,\ldots,\dim L) \tag{3.1}$$

of which are polynomials of the generators $\rho(x_i)$ of ρ and which, under the extension of the adjoint action, transform by a representation σ_Λ of the highest weight Λ (label L omitted):

$$[\rho(x_i),T_{\Lambda,j}(\rho(x_1),\ldots,\rho(x_{\dim L}))] = \sum_{k=1}^{\dim \sigma_\Lambda} (\sigma_\Lambda)^i_{jk} T_{\Lambda,k}(\rho(x_i),\ldots,\rho(x_{\dim L})) \tag{3.2}$$

where $(\sigma_\Lambda)_{jk}$ are matrix elements of representation σ_Λ. To prove

$$T_{\Lambda,j}(\rho(x_1),\ldots,\rho(x_{\dim L}))=0\quad (j=1,\ldots,\dim\sigma_\Lambda) \tag{3.3}$$

it is sufficient to prove that

$$T_{\Lambda,j}(\rho(x_1),\ldots,\rho(x_{\dim L}))v_\rho=0,\quad \text{for any } j=1,\ldots,\dim\sigma_\Lambda, \tag{3.4}$$

where $v_\rho\in V_\rho$ is the highest-weight vector of representation ρ. Indeed, if the operators (3.1) vanish on v_ρ they vanish on any vector of V_ρ; this results by induction using (3.2) and

$$T_{\Lambda,j}\rho(x_i)v_\rho = -[\rho(x_i),T_{\Lambda,j}]v_\rho+\rho(x_i)T_{\Lambda,j}v_\rho. \tag{3.5}$$

Let us now prove the 12 propositions which compose Theorem 1. The propositions will be proved in the following order: 10, 7, 11, 8, 12, 9, 4, 5, 6, 1, 2, 3. For the sake of concision we sketch only the proofs which give the highest weights.

Proof of the theorem: Algebras of types C_n and D_n.

Propositions 7 and 10 [$L=\mathrm{sp}(2n,C)$, $\Lambda=\Lambda_2$; $L=\mathrm{so}(2n,C)$, $\Lambda=2\Lambda_1$]: We shall treat these two cases simultaneously, using unifying expressions for the structure relations of the two algebras as well as for their tensor operators.

Let us characterize the algebras $\mathrm{so}(2n,C)$ and $\mathrm{sp}(2n,C)$ by means of the values taken by a parameter ϵ,

$$\epsilon=\begin{cases}+1, & \text{for } \mathrm{so}(2n,C),\\ -1, & \text{for } \mathrm{sp}(2n,C).\end{cases} \tag{3.6}$$

The structure relations of the two algebras, expressed in Cartan–Weyl bases, are then (cf. Appendix A)

$$[A_{ij},A_{kl}]=\delta_{jk}A_{il}-\delta_{il}A_{kj}, \tag{3.7}$$

$$[A_{ij},B_{kl}]=\delta_{jk}B_{il}-\varepsilon\delta_{jl}B_{ik}, \tag{3.8}$$

$$[A_{ij},C_{kl}]=\varepsilon\delta_{il}C_{jk}-\delta_{ik}C_{jl}, \tag{3.9}$$

$$[B_{ij},C_{kl}]=-\delta_{jk}A_{il}-\delta_{il}A_{jk}+\varepsilon\delta_{ik}A_{jl}+\varepsilon\delta_{jl}A_{ik}, \tag{3.10}$$

$$[B_{ij},B_{kl}]=[C_{ij},C_{kl}]=0. \tag{3.11}$$

The identities (2.14) for $\mathrm{so}(2n,C)$ and (2.21) for $\mathrm{sp}(2n,C)$, expressed in the Cartan–Weyl bases, become

$$\mathbf{AB}-\mathbf{BA}^t=-\varepsilon(\mathbf{AB}-\mathbf{BA}^t)^t, \tag{3.12}$$

$$\mathbf{CA}-\mathbf{A}^t\mathbf{C}=-\varepsilon(\mathbf{CA}-\mathbf{A}^t\mathbf{C})^t, \tag{3.13}$$

$$\mathbf{A}^2+((\mathbf{A}^t)^2)^t-\mathbf{BC}-(\mathbf{CB})^t = (1/n)\mathrm{Tr}\{\mathbf{A}^2+((\mathbf{A}^t)^2)^t-\mathbf{BC}-(\mathbf{CB})^t\mathbf{I}\}, \tag{3.14}$$

where $\mathbf{X}$ denotes the matrix (X_{ij}), $\mathbf{X}^t$ the transpose of $\mathbf{X}$, and $\mathbf{I}\equiv(\delta_{ij})$.

Let us make these remarks for both algebras under consideration.

(i) The operators A_{ii}, $i=1,\ldots,n$, are generators of the Cartan subalgebra; hence, denoting by v_ρ the highest-weight vector of representation ρ,

$$\rho(A_{ii})v_\rho=f_iv_\rho, \tag{3.15}$$

where f_i denotes the ith component of the highest weight.

(ii) The operators A_{ij}, with $i<j$ and B_{ij} with arbitrary i and j, are raising operators, i.e.,

$$\rho(A_{ij})v_\rho=0,\quad \text{for } i<j, \tag{3.16}$$

$$\rho(B_{kl})v_\rho=0,\quad \text{for any } k,l. \tag{3.17}$$

(iii) The operators A_{ij}, with $i>j$ and C_{kl} with arbitrary k and l, are lowering operators.

We shall examine successively the effect of applying the relations (3.12), (3.14), and (3.13) satisfied by the generators of a representation ρ to the highest-weight vector v_ρ of this representation. To avoid cumbersome expressions, we shall write A_{ij}, B_{ij}, C_{ij} instead of $\rho(A_{ij})$, $\rho(B_{ij})$, $\rho(C_{ij})$, respectively.

Both sides of Eq. (3.12) applied to v_ρ vanish as a consequence of Eqs. (3.8) and (3.17). Equation (3.12) thus gives no information.

Let us consider Eq. (3.14). The action of the (i,j) matrix element of the lhs of Eq. (3.14) upon v_ρ may be written

$$[\mathbf{A}^2+((\mathbf{A}^t)^2)^t-\mathbf{BC}-(\mathbf{CB})^t]_{i,j}v_\rho = 2[\mathbf{A}^2+(\mathrm{Tr}\,\mathbf{A})\mathbf{I}-\epsilon\mathbf{A}]_{ij}v_\rho. \tag{3.18}$$

Applying to v_ρ the operators in both members of Eq. (3.14) and taking (3.18) into account leads to the relation

$$(\mathbf{A}^2-\epsilon\mathbf{A})_{ij}v_\rho=(\delta_{ij}/n)\mathrm{Tr}(\mathbf{A}^2-\epsilon\mathbf{A})v_\rho \tag{3.19}$$

valid for any $i,j=1,\ldots,n$. We shall examine separately the cases $i=j$, $i<j$, and $i>j$.

Let $i=j$ and denote

$$(1/n)\mathrm{Tr}(\mathbf{A}^2-\epsilon\mathbf{A})v_\rho=cv_\rho. \tag{3.20}$$

Taking into account Eqs. (3.7), (3.15), and (3.16), the equality (3.19) becomes

$$\left(f_i^2+(n-i-\epsilon)f_i-\sum_{l=i+1}^{n}f_l\right)v_\rho=cv_\rho, \tag{3.21}$$

whence

$$(f_i-f_{i+1})(f_i+f_{i+1}+n-i-\epsilon)=0\quad (i=1,\ldots,n-1). \tag{3.22}$$

We shall solve the system (3.22) imposing the conditions

$$f_1\geqslant f_2\geqslant\cdots\geqslant f_n\geqslant 0,\quad \text{for } \mathrm{sp}(2n,C), \tag{3.23}$$

and

$$f_1\geqslant f_2\geqslant\cdots\geqslant f_{n-1}\geqslant|f_n|,\quad \text{for } \mathrm{so}(2n,C). \tag{3.24}$$

A solution of Eq. (3.22) admissible for both cases $\varepsilon=\pm1$ is

$$f_1 = f_2 = \cdots = f_n = k \quad (k = \text{integer}). \tag{3.25}$$

It results from the vanishing of the first factor in (3.22) for all values of $i = 1,...,n-1$.

Let us admit now that for a given i

$$f_i + f_{i+1} + n - i - \epsilon = 0. \tag{3.26}$$

Now we must distinguish between the two cases $\epsilon = \pm 1$.

(1) If $\epsilon = -1$, then for any $i = 1,2,...,n-1$, Eq. (3.26) leads to $f_i + f_{i+1} < 0$, in contradiction with condition (3.23): for $\mathrm{sp}(2n,C)$, the solution (3.25) is the only admissible.

(2) If $\epsilon = +1$, then for any $i = 1,2,...,n-2$, we get from (3.26) $f_i + f_{i+1} < 0$, a relation incompatible with (3.24). However, for $i = n-1$, we obtain $f_n + f_{n-1} = 0$ which is an admissible solution: for the algebra $\mathrm{so}(2n,C)$ a second solution

$$f_1 = f_2 = \cdots = f_{n-1} = -f_n = m$$
$$(m = \text{positive integer}) \tag{3.27}$$

exists.

Thus written in terms of fundamental weights Λ_i and Dynkin indices, the highest weights of the representations for which the components of the tensor operator $T_{\mathrm{sp}(2n,C),\Lambda_2}$ vanish have the expression $m\Lambda_n$ $(m = 1,2,...)$; the highest weights of the representations for which the components of $T_{\mathrm{so}(2n,C),2\Lambda_1}$ vanish have the expressions $m\Lambda_{n-1}$ and $m\Lambda_n$ $(m = 1,2,...)$.

Proposition 11 $[L = \mathrm{so}(2n,C);\ \Lambda = \Lambda_4]$: To get information about the components of the highest weight we shall consider the component $p = 2i$, $q = 2i-1$, $r = 2j$, $s = 2j-1$ of the tensor operator T_{Λ_4} and express it in the Cartan–Weyl basis

$$T_{\Lambda_4}(2i,2i-1,2j,2j-1) = -2A_{ii}A_{jj} - [B_{ij},C_{ij}]_+ + [A_{ij},A_{ji}]_+. \tag{3.28}$$

Recalling Eqs. (3.15)–(3.17) and using condition (3.4) we get

$$T_{\Lambda_4}(2i,2i-1,2j,2j-1)v_\rho = -2f_j(f_i+1)v_\rho = 0, \tag{3.29}$$

whence, assuming $1 = i < j$, the Cartan–Weyl components of the highest weight of representation ρ are

$$f_1 \neq 0, \quad f_2 = f_3 = \cdots = f_n = 0, \tag{3.30}$$

i.e., the highest weight of ρ is $m\Lambda_1$ (m = positive integer).

Proposition 8 $[L = \mathrm{sp}(2n,C);\ \Lambda = 4\Lambda_1]$: Expressed in the Cartan–Weyl basis, the components of the tensor operator $T_{4\Lambda_1}(p,q,r,s)$ have no longer a unique expression. We have to consider separately the cases

$$p,q,r,s < n; \quad p,q,r < n, \quad s > n; \quad p,q < n, \quad r,s > n, \tag{3.31}$$

$$p < n, \quad q,r,s, > n; \quad p,q,r,s > n. \tag{3.32}$$

The last two cases lead to operators which vanish if applied to the highest-weight vector v_ρ and are therefore irrelevant. The three tensor operators (3.31) lead to the equations

$$([C_{pq},C_{rs}]_+ + [C_{ps},C_{rq}]_+ + [C_{pr},C_{qs}]_+)v_\rho = 0, \tag{3.33}$$

$$([C_{pq},A_{s-n,r}]_+ + [A_{s-n,p},C_{rq}]_+ + [C_{pr},A_{s-n,q}]_+)v_\rho = 0, \tag{3.34}$$

$$([C_{pq},B_{r-n,s-n}]_+ + [A_{s-n,p},A_{r-n,q}]_+ + [A_{r-n,p},A_{s-n,q}]_+)v_\rho = 0. \tag{3.35}$$

The information about the highest weight of representation ρ is provided by Eq. (3.35) which becomes

$$\text{for } p = q = r-n = s-n = i\text{:} \quad f_i(f_i - 1)v_\rho = 0, \tag{3.36}$$

$$\text{for } p = q = r = i < j = s-n\text{:} \quad (f_i - 1)A_{ji}v_\rho = 0, \tag{3.37}$$

$$\text{for } p = i \neq j = q = r-n = s-n\text{:} \quad f_jA_{ji}v_\rho = 0. \tag{3.38}$$

Equation (3.36) implies that the components f_i $(i = 1,...,n)$ of the highest weight can take only the values 0 or 1. Hence recalling condition (3.23), only the fundamental representations ρ_{Λ_k} $(f_1 = \cdots = f_k = 1, f_{k+1} = \cdots = f_n = 0;$ $k = 1,2,...,n)$ are allowed as solutions. Let us prove that for any k Eqs. (3.37)–(3.38) are verified. The highest-weight vector of representation ρ_{Λ_k} is of the form $v_\rho = v_1 \wedge v_2 \wedge \cdots \wedge v_k$ [where $v_1,...,v_{2n}$ is a basis of the representation space of $\rho_{\Lambda 1}$ in which $\rho_{\Lambda 1}(A_{ji})v_\rho = \delta_{il}v_j - \delta_{j+n,l}v_i$]. We have $\rho_{\Lambda_k}(A_{ii})v_\rho = v_\rho$ if $i \leqslant k$; $\rho_{\Lambda_k}(A_{ii})v_\rho = 0$ if $i > k$ and $\rho_{\Lambda_k}(A_{ji})v_\rho = 0$ if $k < i \leqslant n$ and $1 \leqslant j \leqslant k$ or if $1 \leqslant i \leqslant k$ and $1 \leqslant j \leqslant k$. In the remaining case $(1 \leqslant i \leqslant k$ and $j > k)$, Eqs. (3.37)–(3.38) are verified as, in this case, $f_i = 1$ and $f_j = 0$.

Proposition 12 $[L = \mathrm{so}(2n,C),\ \Lambda = 2\Lambda_2]$: In the Cartan–Weyl basis

$$\begin{aligned} T_{2\Lambda_2}&(2i,2i-1,2i,2i-1)v_\rho \\ &= \Big[-A_{ii}^2 - \frac{1}{2(n-1)}\sum_{k,l=1}^{n}\Big(\delta_{il} + \frac{1}{2n-1}\Big)([B_{kl},C_{kl}]_+ + [A_{kl},A_{lk}]_+)\Big]v_\rho \\ &= \Big[-f_i^2 + \frac{1}{n-1}\Big\{f_i^2 + (n-i-1)f_i + \sum_{k=1}^{i} f_k - \frac{1}{2n-1}\Big(\sum_1^n f_l^2 + 2\sum_{l=1}^{n-1}(n-l)f_l\Big)\Big\}\Big]v_\rho = 0. \end{aligned} \tag{3.39}$$

Subtracting from each other the expressions (3.39) written for two consecutive values i and $i+1$ leads to the system

$$(f_{i+1} - f_i)[(f_{i+1} + f_i)(n-2) - (n-i+1)] = 0 \quad (i = 1,2,...,n-1), \tag{3.40}$$

which admits as solutions

$$f_1 = f_2 = \cdots = f_n = k, \tag{3.41}$$

$$f_1 = f_2 = \cdots = f_{n-1} = -f_n = k, \tag{3.42}$$

$$f_1 = 1, \quad f_2 = f_3 \cdots = f_n = 0. \tag{3.43}$$

The solution (3.42) is obtained from the vanishing of the second factor in the $(n-1)$th equation of the system

(3.40). [Solution (3.43) verifies also Eq. (3.39)]. We get $k=\frac{1}{2}$ by solving

$$T_{2\Lambda_2}(2,1,2,1)v_\rho=(n-1)k(-2k+1)v_\rho=0. \quad (3.44)$$

The highest weight of representation ρ is thus one of the fundamental weights Λ_1, Λ_{n-1}, and Λ_n.

Proposition 9 [$L=\mathrm{sp}(2n,C)$, $\Lambda=2\Lambda_2$]: We have

$$\begin{aligned}&[T_{2\Lambda_2}(p+1+n,p+1,p+1+n,p+1)\\&\quad -T_{2\Lambda_2}(p+n,p,p+n,p)]v_\rho\\&\quad=(f_{p+1}-f_p)[(n-2)(f_{p+1}+f_p)\\&\qquad-(n-3p+1)]v_\rho=0.\end{aligned} \quad (3.45)$$

The only nontrivial solution of the system (3.45) is

$$f_1=1,\quad f_2=f_3=\cdots=f_n=0, \quad (3.46)$$

i.e., the fundamental representation of highest weight Λ_1.

Algebras of type B_n.

Proposition 4 [$L=\mathrm{so}(2n+1,C)$, $\Lambda=2\Lambda_1$]: The condition $T_{2\Lambda_1}(2n+1,2n+1)v_\rho=0$, written in the Cartan–Weyl basis, becomes

$$\sum_{k=1}^{n}\left[f_k^2+f_k\left(n-2k+\frac{1}{2}\right)\right]=0. \quad (3.47)$$

Equation (3.47) admits the only solution $f_1=f_2=\cdots=f_n=\frac{1}{2}$, i.e., the representation ρ has the highest weight Λ_n.

Proposition 5 [$L=\mathrm{so}(2n+1,C)$, $\Lambda=\Lambda_4$]: In this case, the information given by the relations $T_{\Lambda_4}(p,q,r,s)=0$ with $1\leqslant p,q,r,s\leqslant 2n$ (cf. proof of Proposition 11) has to be completed with information provided by the relations in which the labels can take the value $2n+1$. The relations informing about the weight of representation ρ are those with $p=q=2n+1$. These relations are identically satisfied; hence for algebras of type B_n the identities $T_{\Lambda_4}(\rho)=0$ have the same solution as for algebras of type D_n: the highest weight of representation ρ is $m\Lambda_1$ ($m=1,2,\ldots$).

Proposition 6 [$L=\mathrm{so}(2n+1,C)$, $\Lambda=2\Lambda_2$]: We have

$$\begin{aligned}&T_{2\Lambda_2}(2i,2i-1,2i,2i-1)v_\rho\\&\quad=\left\{-f_i^2+\frac{1}{2n-1}\left(2f_i^2+(2n-2i-1)f_i\right.\right.\\&\qquad\left.+2\sum_{k=1}^{i}f_k\right)-\frac{1}{n(2n-1)}\\&\qquad\left.\times\sum_{k=1}^{n}[f_k^2+(2n-2k+1)f_k]\right\}v_\rho=0,\end{aligned} \quad (3.48)$$

whence the following system is obtained:

$$\begin{aligned}&[T_{2\Lambda_2}(2i+2,2i+1,2i+2,2i+1)\\&\quad-T_{2\Lambda_2}(2i,2i-1,2i,2i-1)]v_\rho\\&\quad=[1/(2n-1)](f_{i+1}-f_i)[(2n-3)(f_{i+1}+f_i)\\&\qquad-1]v_\rho=0,\end{aligned} \quad (3.49)$$

which admits the solutions $f_1=f_2=\cdots=f_n=k$ and $f_1=1$, $f_2=f_3=\cdots=f_n=0$. The second solution satisfies also Eq. (3.48). This equation gives also the value of k which is $k=\frac{1}{2}$. The highest weight of representation ρ is thus Λ_1 or Λ_n.

Algebras of type A_n.

Let us express in Eqs. (2.8)–(2.10) the generators of the algebra $\mathrm{sl}(n+1,C)$ in terms of the generators of $\mathrm{gl}(n+1,C)$,

$$A_{ij}=e_{ij}-\frac{\delta_{ij}}{n+1}\sum_{k=1}^{n+1}e_{kk}. \quad (3.50)$$

Let v_ρ be the highest-weight vector of the representation ρ of $\mathrm{gl}(n+1,C)$. We have

$$e_{ii}v_\rho=f_iv_\rho,\quad i=1,2,\ldots,n+1, \quad (3.51)$$

$$e_{ij}v_\rho=0,\quad\text{for any } i<j, \quad (3.52)$$

where f_i are the components of the highest weight, satisfying the inequalities

$$f_1\geqslant f_2\geqslant\cdots\geqslant f_n\geqslant f_{n+1}. \quad (3.53)$$

Let us denote

$$\sigma_1=\sum_{i=1}^{n+1}f_i$$

and (3.54)

$$\sigma_2=\sum_{i=1}^{n+1}[f_i^2+(n+2-2i)f_i],$$

where

$$\sum_{i,j=1}^{n+1}e_{ij}e_{ji}v_\rho=\sigma_2v_\rho.$$

Proposition 1 [$L=\mathrm{sl}(n+1,C)$, $\Lambda=\Lambda_1+\Lambda_n$]: From

$$\begin{aligned}T_{\Lambda_1+\Lambda_n}(p,p)v_\rho&\equiv\left(\sum_{i=1}^{n+1}[A_{pi},A_{ip}]_+-\frac{2}{n+1}\sum_{i,j=1}^{n+1}A_{ij}A_{ji}\right)v_\rho\\&=\left[2f_p^2+(n-2p+2)f_p\right.\\&\quad+\sum_{i=1}^{p-1}f_i-\sum_{i=p+1}^{n+1}f_i+\frac{4}{n+1}\\&\quad\left.\times\left(-f_p\sigma_1+\left(\frac{\sigma_1^2}{n+1}-\frac{\sigma_2}{2}\right)\right)\right]v_\rho=0,\end{aligned} \quad (3.55)$$

we obtain

$$\begin{aligned}&[T_{\Lambda_1+\Lambda_n}(p+1,p+1)-T_{\Lambda_1+\Lambda_n}(p,p)]v_\rho\\&\quad=(f_{p+1}-f_p)\left[2(f_{p+1}+f_p)\right.\\&\qquad\left.+(n-2p+1)-\frac{4\sigma_1}{n+1}\right]v_\rho=0.\end{aligned} \quad (3.56)$$

Assuming that the second factor in (3.56) vanishes for two distinct values $p>r$, we get $f_p+f_{p+1}>f_r+f_{r+1}$ in contradiction with Eq. (3.53): the highest weight has maximum degeneracy

$$f_1=f_2=\cdots=f_p>f_{p+1}=\cdots=f_{n+1}=f_p-k. \quad (3.57)$$

From the vanishing of the second factor in (3.56), we get

$$(1+2k/(n+1))(p-(n+1)/2)=0. \quad (3.58)$$

The weight (3.57) verifies Eq. (3.58) for any k iff $p=(n+1)/2$. This solution verifies also Eq. (3.55), which thus admits a solution only for odd values of n; this solution is the representation with highest weight $m\Lambda_{(n+1)/2}$ $(m=1,2,\ldots)$.

The apparition of this rather strange solution will obtain an explanation in Sec. IV.

Proposition 2 [$L=\mathrm{sl}(n+1,C)$, $\Lambda=\Lambda_2+\Lambda_{n-1}$]: It can be proved that the components of the highest weight satisfy the relations $f_1=\cdots=f_p>f_{p+1}=\cdots=f_{n+1}$. From

$$\begin{aligned}T_{\Lambda_2+\Lambda_{n-1}}&(p=q<p+1=r=s)v_p\\&=-2(p-1)(n-p)(f_{p+1}-f_p)(f_{p+1}\\&\quad-f_p+1)v_p=0\end{aligned}\tag{3.59}$$

and from the impossibility to vanish of the last parenthesis in Eq. (3.59) [cf. Eq. (3.53)] we get that the only admissible highest weights are

$$f_1>f_2=\cdots=f_{n+1}\tag{3.60}$$

and

$$f_1=f_2=\cdots=f_n>f_{n+1},\tag{3.61}$$

i.e., the representation ρ has the highest weight $m\Lambda_1$ or $m\Lambda_n$ $(m=1,2,\ldots)$.

Proposition 3 [$L=\mathrm{sl}(n+1,C)$, $\Lambda=2\Lambda_1+2\Lambda_n$]: Taking into account Eqs. (3.50), (3.52), and (3.54), the equality $T_{2\Lambda_1+2\Lambda_n}(p=q=r=s)v_p=0$ leads to the system

$$\begin{aligned}f_p^2-\frac{2}{n+3}\Bigg[&\sum_{i=1}^{p-1}f_i+f_p^2+\left(\sigma_1+\frac{n+3}{2}-p\right)f_p\\&-\frac{1}{2(n+2)}(\sigma_1^2+\sigma_2+(n+2)\sigma_1)\Bigg]=0,\\&\qquad(p=1,2,\ldots,n+1).\end{aligned}\tag{3.62}$$

The difference between the $(p+1)$th and the pth equations (3.62) is

$$\begin{aligned}(f_{p+1}-f_p)(&f_{p+1}+f_p-1\\&-[(2\sigma_1-2p)/(n+1)])=0.\end{aligned}\tag{3.63}$$

The second factor in Eq. (3.63) cannot vanish for two values of p simultaneously: suppose $f_p>f_{p+1}=\cdots=f_q>f_{q+1}$; we get from (3.63) $f_p-f_{q+1}<2$ which contradicts the assumption. Putting $f_{p+1}-f_p=k$, we get $k=1$, i.e., $T_{2\Lambda_1+2\Lambda_n}$ vanishes on any fundamental rep.

IV. COMPARISON WITH THE HANNABUSS–KOSTANT–OKUBO METHOD

In the present section we shall compare the results obtained so far with those which will be deduced by applying the HKO method.

A. The Hannabuss–Kostant–Okubo operator

Let ρ_λ be a finite-dimensional representation of highest weight λ of an n-dimensional semisimple Lie algebra L and let V_λ be the corresponding L module; let $c_2(\lambda)$ be the second-degree Casimir operator associated with representation ρ_λ:

$$c_2(\lambda)=\sum_{i=1}^{n}\rho_\lambda(e_i)\rho_\lambda(e^i).\tag{4.1}$$

In Eq. (4.1) $\{e_i,\ i=1,\ldots,n\}$ is a basis in L and $\{e^i,\ i=1,\ldots,n\}$ is the basis of L dual to $\{e_i\}$ with respect to the Cartan–Killing bilinear form: $(e^i,e_j)=\delta_{ij}$.

Definition: We shall call Hannabuss–Kostant–Okubo (HKO) operator associated with the pair of representations ρ_λ and ρ_μ of the semisimple Lie algebra L the operator $\mathscr{O}_{\lambda,\mu}$ defined by

$$\mathscr{O}_{\lambda,\mu}\equiv\sum_{i=1}^{n}\rho_\lambda(e_i)\otimes\rho_\mu(e^i).\tag{4.2}$$

The HKO operator $\mathscr{O}_{\lambda,\mu}$ has the following properties.

(1) It commutes with $\rho_\lambda\otimes\rho_\mu$:

$$[\mathscr{O}_{\lambda,\mu},\rho_\lambda\otimes\rho_\mu]=0.\tag{4.3}$$

(2) It is expressible as a function of the Casimir operators $c_2(\rho_\lambda\otimes\rho_\mu)$, $c_2(\rho_\lambda)$, and $c_2(\rho_\mu)$:

$$\mathscr{O}_{\lambda,\mu}=\tfrac{1}{2}[c_2(\rho_\lambda\otimes\rho_\mu)-c_2(\rho_\lambda)\otimes 1-1\otimes c_2(\rho_\mu)].\tag{4.4}$$

(3) The expression of the minimal polynomial satisfied by $\mathscr{O}_{\lambda,\mu}$ is

$$\begin{aligned}\prod_{\omega\in\Omega(\lambda,\mu)}&[\mathscr{O}_{\lambda,\mu}-\tfrac{1}{2}((\omega+2\delta,\omega)\\&-(\lambda+2\delta,\lambda)-(\mu+2\delta,\mu))I]=0,\end{aligned}\tag{4.5}$$

where $\Omega(\lambda,\mu)$ is the set of distinct weights in the Clebsch–Gordan series of the product $\rho_\lambda\otimes\rho_\mu$, 2δ is the sum of the positive roots of L, and the formula $c_2(\lambda)=(\lambda+2\delta,\lambda)$ for the Casimir operator of the representation ρ_λ has been used.

The HKO method[16,19,28] for the determination of the polynomial relations satisfied by a representation ρ_λ of L consists in taking the matrix elements of the polynomial relation (4.5) between basis vectors of the representation ρ_μ.

Thus in order to obtain second-degree polynomial relations satisfied by ρ_λ it is necessary to determine those representations ρ_μ for which the Kronecker product $\rho_\lambda\otimes\rho_\mu$ decomposes into only two terms, i.e., for which the set of weights $\Omega(\lambda,\mu)$ contains only two elements.

B. Representations for which the HKO operator satisfies an equation of second degree (Ref. 26)

The following theorem determines the pairs of representations $\{\rho_\lambda,\rho_\mu\}$ of nonexceptional semisimple Lie algebras for which relation (4.5) is of second degree, i.e., has the expression

$$\begin{aligned}(\mathscr{O}_{\lambda,\mu}-(\lambda,\mu))(&\mathscr{O}_{\lambda,\mu}-\tfrac{1}{2}((\xi+2\delta,\xi)\\&-(\lambda+2\delta,\lambda)-(\mu+2\delta,\mu))I)=0.\end{aligned}\tag{4.6}$$

Theorem 2: For the semisimple Lie algebras (1.1) the only Kronecker products $\rho_\lambda\otimes\rho_\mu$ of irreducible representations which decompose into a direct sum of two inequivalent irreducible representations are those in columns 2 and 3 of Table II. (The representations into which $\rho_\lambda\otimes\rho_\mu$ decomposes are indicated in columns 4 and 5. The roots of the corresponding minimal polynomials satisfied by the HKO operator are given in columns 6 and 7.)

The *proof* of this proposition is based on a dimensional calculation and on the following result due to Dynkin.[50]

Let L be a semisimple Lie algebra and let ρ_λ and ρ_μ be two irreducible representations of L labeled by their highest weights λ and μ. Let $\alpha_1,\alpha_2,...,\alpha_k$ be a minimal chain of simple roots connecting the weights λ and μ, i.e., a set of simple roots such that

$$(\lambda,\alpha_1)\neq 0 \quad (\alpha_k,\mu)\neq 0, \qquad (\alpha_i,\alpha_{i+1})\neq 0 \quad (i=1,2,...,k-1), \tag{4.7}$$

and such that no proper subset of $\{\alpha_1,...,\alpha_k\}$ having the same properties exists. Then

$$\xi(\alpha_1,\alpha_2,...,\alpha_k)\equiv\lambda+\mu-(\alpha_1+\alpha_2+\dots+\alpha_k) \tag{4.8}$$

is the maximum weight for one and only one of the irreducible components in the decomposition of the Kronecker product $\rho_\lambda\otimes\rho_\mu$.

Candidates for the factors ρ_λ and ρ_μ are the representations contained in column 4 of Table I. We know now two irreducible components of the product $\rho_\lambda\otimes\rho_\mu$, namely, $\rho_{\lambda+\mu}$ (of highest weight $\lambda+\mu$) and ρ_ξ [of highest weight (4.8)] and may thus write

$$\rho_\lambda\otimes\rho_\mu=\rho_{\lambda+\mu}\oplus\rho_\xi\oplus\cdots. \tag{4.9}$$

A dimensional calculation based on Weyl's dimension formula allows one to conclude that in all the cases included in Table II we have precisely

$$\rho_\lambda\otimes\rho_\mu=\rho_{\lambda+\mu}\oplus\rho_\xi. \tag{4.10}$$

Details of the calculation are given in Ref. 25.

The unicity of the decompositions (4.10) listed in Table II results from the Kronecker products pointed out in Appendix C.

C. Comparison for the Lie algebras of type D_n

Perfect coincidence between the results of the two methods is obtained for the algebras of types B_n and D_n. We shall give, in the following, complete proofs of this statement for the algebras of type D_n.

For these algebras, the only second-degree quantum tensorial identities are Eqs. (2.14)–(2.16) (symmetrized!). Theorem 1 points out that the irreducible representations labeled by the highest weights listed in column 4 of Table I are the only solutions of these identities.

What we have to prove is that, for algebras of type D_n, the identities obtained by applying the HKO method to the pairs of representations listed in columns 2 and 3 of Table II exhaust the quantum tensorial identities (2.14)–(2.16). More explicitly, we shall prove the following.

(i) Taking the matrix elements, with respect to ρ_{Λ_1}, of the equation satisfied by the HKO operator associated with $\rho_{m\Lambda_n}\otimes\rho_{\Lambda_1}$ we obtain the quantum tensorial identities (2.14),

$$\sum_{j=1}^{2n}[\rho_{m\Lambda_n}(M_{ij}),\rho_{m\Lambda_n}(M_{jl})]_+ = \frac{m(m+2n-2)}{2}I_{m\Lambda_n}\delta_{il}. \tag{4.11}$$

(ii) Similarly, applying the HKO method to the pair of representations $\rho_{m\Lambda_1}$ and $\rho_{\Lambda_{n-1}}\oplus\rho_{\Lambda_n}$ we obtain the quantum tensorial identities (2.15),

$$[\rho_{m\Lambda_1}(M_{ij}),\rho_{m\Lambda_1}(M_{kl})]_+ + [\rho_{m\Lambda_1}(M_{il}),\rho_{m\Lambda_1}(M_{jk})]_+ + [\rho_{m\Lambda_1}(M_{ik}),\rho_{m\Lambda_1}(M_{lj})]_+=0. \tag{4.12}$$

(iii) Applying the HKO method to the pair of representations $\rho\Lambda_n$ and $\rho_{2\Lambda_1}$ we obtain the quantum tensorial identities (2.16),

$$\tfrac{1}{3}\{2[\rho_{\Lambda_n}(M_{ij}),\rho_{\Lambda_n}(M_{kl})]_+ - [\rho_{\Lambda_n}(M_{il}),\rho_{\Lambda_n}(M_{jk})]_+ - [\rho_{\Lambda_n}(M_{ik}),\rho_{\Lambda_n}(M_{lj})]_+\} = \tfrac{1}{4}(\delta_{il}\delta_{jk}-\delta_{ik}\delta_{jl})I_{\Lambda_n}. \tag{4.13}$$

Proof of (i): The HKO operator $\mathscr{O}_{m\Lambda_n,\Lambda_1}$ has the expression

$$\mathscr{O}_{m\Lambda_n,\Lambda_1}\equiv\sum_{1=i<j}^{2n}\rho_{m\Lambda_n}(M_{ij})\otimes(e_{ji}-e_{ij}) = -\sum_{i,j=1}^{2n}\rho_{m\Lambda_n}(M_{ij})\otimes e_{ij}. \tag{4.14}$$

Introducing this expression in the equation

$$(\mathscr{O}_{m\Lambda_n,\Lambda_1}-(m/2)I_{m\Lambda_n}\otimes I_{\Lambda_1})(\mathscr{O}_{m\Lambda_n,\Lambda_1} + [(m+2n-2)/2]\,I_{m\Lambda_n}\otimes I_{\Lambda_1})=0 \tag{4.15}$$

satisfied by $\mathscr{O}_{m\Lambda_n,\Lambda_1}$, we get the equation

$$\sum_{i,l=1}^{2n}\sum_{j=1}^{2n}\rho_{m\Lambda_n}(M_{ij})\,\rho_{m\Lambda_n}(M_{jl})\otimes e_{il} - (n-1)\sum_{i,l=1}^{2n}\rho_{m\Lambda_n}(M_{il})\otimes e_{il} - \frac{m(m+2n-2)}{4}I_{m\Lambda_n}\otimes I_{\Lambda_1}=0. \tag{4.16}$$

Observing that

$$\sum_{j=1}^{2n}\rho_{m\Lambda_n}(M_{ij})\,\rho_{m\Lambda_n}(M_{jl}) = \frac{1}{2}\sum_{j=1}^{2n}[\rho_{m\Lambda_n}(M_{ij}),\rho_{m\Lambda_n}(M_{jl})]_+ + (n-1)\,\rho_{m\Lambda_n}(M_{il}), \tag{4.17}$$

we obtain

$$2\sum_{i,l,j=1}^{2n}[\rho_{m\Lambda_n}(M_{ij}),\rho_{m\Lambda_n}(M_{jl})]_+\otimes e_{il} = m(m+2n-2)\,I_{m\Lambda_n}\otimes I_{\Lambda_1}, \tag{4.18}$$

whence, taking the matrix elements of ρ_{Λ_1}, we obtain Eq. (4.11). To compare Eq. (4.11) with Eqs. (2.13) and (2.14) let us recall that

$$\sum_{i,j=1}^{2n}[\rho_{m\Lambda_n}(M_{ij}),\rho_{m\Lambda_n}(M_{ji})]_+ = mn(m+2n-2)\,I_{m\Lambda_n}. \tag{4.19}$$

Proof of (ii): From Table II we read that each of the products $\rho_{m\Lambda_1}\otimes\rho_{\Lambda_{n-1}}$ and $\rho_{m\Lambda_1}\otimes\rho_{\Lambda_n}$ decomposes into a direct sum of two irreducible representations and that the HKO operators $\mathscr{O}_{m\Lambda_1,\Lambda_{n-1}}$ and $\mathscr{O}_{m\Lambda_1,\Lambda_n}$ satisfy an equation similar to Eq. (4.15). This equation will be satisfied also by the HKO operator associated with the pair of representations $\{\rho_{m\Lambda_1}, \rho_{\Lambda_{n-1}} \oplus \rho_{\Lambda_n}\}$; we shall denote (abusively) this operator by $\mathscr{O}_{m\Lambda_1,\Lambda_{n-1}\oplus\Lambda_n}$.

Taking into account that the generators of the spinorial representation $\rho_{\Lambda_{n-1}} \oplus \rho_{\Lambda_n}$ of so$(2n,C)$ can be written as functions of the generators Γ_i, $i = 1,...,2n$,

$$[\Gamma_i,\Gamma_j]_+ = 2\delta_{ij}\, I_{2^n}, \tag{4.20}$$

of the Clifford algebra C_{2n},

$$\rho_{\Lambda_{n-1}}(M_{ij}) \oplus \rho_{\Lambda_n}(M_{ij}) = \tfrac{1}{4}[\Gamma_i,\Gamma_j]_- \tag{4.21}$$

we obtain, observing that $M_{ji} = -M_{ij}$, the expression of the HKO operator

$$\mathscr{O}_{m\Lambda_1,\Lambda_{n-1}\oplus\Lambda_n} = \frac{1}{4}\sum_{i,j=1}^{2n} \rho_{m\Lambda_1}(M_{ij}) \otimes \Gamma_j\Gamma_i\,. \tag{4.22}$$

Equation (4.15) satisfied by $\mathscr{O}_{m\Lambda_1,\Lambda_{n-1}\oplus\Lambda_n}$ becomes

$$\frac{1}{4}\sum_{i,j,k,l=1}^{2n} \rho_{m\Lambda_1}(M_{ij})\,\rho_{m\Lambda_1}(M_{kl}) \otimes \Gamma_j\Gamma_i\Gamma_l\Gamma_k + (n-1)\sum_{i,j=1}^{2n}\rho_{m\Lambda_1}(M_{ij})\otimes\Gamma_j\Gamma_i - m(m+2n-2)\, I_{m\Lambda_1}\otimes I_{2^n} = 0\,. \tag{4.23}$$

Taking into account Eq. (4.20), it can be proved that

$$\frac{1}{4}\sum_{i,j,k,l=1}^{2n} \rho_{m\Lambda_1}(M_{ij})\,\rho_{m\Lambda_1}(M_{kl}) \otimes \Gamma_j\Gamma_i\Gamma_l\Gamma_k = \frac{1}{8}\sum_{i,j,k,l=1}^{2n}[\rho_{m\Lambda_1}(M_{ij}),\rho_{m\Lambda_1}(M_{kl})]_+ \otimes \Gamma_j\Gamma_i\Gamma_l\Gamma_k - (n-1)\sum_{i,j=1}^{2n}\rho_{m\Lambda_1}(M_{ij})\otimes\Gamma_j\Gamma_i \tag{4.24}$$

and Eq. (4.23) becomes

$$\frac{1}{8}\sum_{i,j,k,l=1}^{2n}[\rho_{m\Lambda_1}(M_{ij}),\rho_{m\Lambda_1}(M_{kl})]_+ \otimes \Gamma_j\Gamma_i\Gamma_l\Gamma_k = m(m+2n-2)\, I_{m\Lambda_1}\otimes I_{2^n}\,. \tag{4.25}$$

In order to prove the equivalence of this equation with Eq. (2.15) we shall use the following slightly modified form of Lemma 6.4 from Ref. 51.

Lemma: Let $\{R_{ijkl},\ 1 \leqslant i,j,k,l \leqslant 2n\}$ be a set of operators with the properties

$$(1)\ R_{ijkl} = R_{klij}\,, \tag{4.26}$$

$$(2)\ R_{ijkl} = -R_{jikl}\,, \tag{4.27}$$

and let $\{\Gamma_i,\ i = 1,...,2n\}$ satisfying Eq. (4.20) be the generators of a Clifford algebra. Then

$$\sum_{i,j,k,l=1}^{2n} R_{ijkl}\otimes\Gamma_i\Gamma_j\Gamma_k\Gamma_l = 8\sum_{1=i<j<k<l}^{2n}(R_{ijkl}+R_{iljk}+R_{iklj})\otimes\Gamma_i\Gamma_j\Gamma_k\Gamma_l + 2\sum_{i,j=1}^{2n} R_{ijji}\otimes I_{2^n}\,. \tag{4.28}$$

Let now [cf. Table I and Eq. (2.15)]

$$R_{ijkl} \equiv [\rho_{m\Lambda_1}(M_{ij}),\rho_{m\Lambda_1}(M_{kl})]_+\,. \tag{4.29}$$

Taking into account Eqs. (4.28) and (4.29) and recalling that

$$\sum_{1=i<j}^{2n}\rho_{m\Lambda_1}(M_{ij})\,\rho_{m\Lambda_1}(M_{ji}) = m(m+2n-2)\, I_{m\Lambda_1}\,, \tag{4.30}$$

Eq. (4.25) becomes

$$\sum_{i<j<k<l}\{[\rho_{m\Lambda_1}(M_{ij}),\rho_{m\Lambda_1}(M_{kl})]_+ + [\rho_{m\Lambda_1}(M_{il}),\rho_{m\Lambda_1}(M_{jk})]_+ + [\rho_{m\Lambda_1}(M_{ik}),\rho_{m\Lambda_1}(M_{lj})]_+\}\otimes\Gamma_j\Gamma_i\Gamma_l\Gamma_k = 0\,. \tag{4.31}$$

The equivalence between the HKO equation and Eq. (2.15) is evident.

Proof of (iii): To prove that the HKO polynomial relations for the representations $\rho_{\Lambda_{n-1}}$ and ρ_{Λ_n} are precisely the quantum tensorial identities (2.16) we consider the products $\rho_{\Lambda_{n-1}}\otimes\rho_{2\Lambda_1}$ and $\rho_{\Lambda_n}\otimes\rho_{2\Lambda_1}$; the equation satisfied by the HKO operator is (cf. Table II)

$$(\mathscr{O}_{\Lambda_n,2\Lambda_1})^2 + (n-1)\mathscr{O}_{\Lambda_n,2\Lambda_1} - nI_{\Lambda_n}\otimes I_{2\Lambda_1} = 0\,. \tag{4.32}$$

Let us introduce in the representation space V_{Λ_1} of ρ_{Λ_1} an orthonormal basis $\{v_i,\ i = 1,...,2n\}$ on which the action of $\rho_{\Lambda_1}(M_{kl})$ is

$$\rho_{\Lambda_1}(M_{kl})v_i \equiv \delta_{li}v_k - \delta_{ki}v_l\,. \tag{4.33}$$

We have

$$\rho_{\Lambda_1}\otimes\rho_{\Lambda_1} \equiv \rho_{2\Lambda_1}\oplus\rho_{(0)}\,; \tag{4.34}$$

the vectors transforming under $\rho_{2\Lambda_1}$ are

$$v_i\vee v_j - \frac{\delta_{ij}}{2n}\sum_{s=1}^{2n} v_s\vee v_s\,, \tag{4.35}$$

where we denoted

$$v_i\vee v_j \equiv \tfrac{1}{2}(v_i\otimes v_j + v_j\otimes v_i)\,. \tag{4.36}$$

To determine the matrix elements of the HKO operator between states of the representation $\rho_{2\Lambda_1}$ we first calculate

$$\begin{aligned}
&\{(\mathscr{O}_{\Lambda_n,2\Lambda_1})^2 + (n-1)\mathscr{O}_{\Lambda_n,2\Lambda_1} - nI_{\Lambda_n}\otimes I_{2\Lambda_1}\}\Big(v_i\vee v_j - \frac{\delta_{ij}}{2n}\sum_{s=1}^{2n} v_s\vee v_s\Big)\\
&= \frac{1}{2}\bigg\{[\rho_{\Lambda_n}(M_{ki}),\rho_{\Lambda_n}(M_{lj})]_+ + [\rho_{\Lambda_n}(M_{kj}),\rho_{\Lambda_n}(M_{li})]_+ + \delta_{lj}\Big(\sum_{r=1}^{2n}\rho_{\Lambda_n}(M_{kr})\,\rho_{\Lambda_n}(M_{ri}) + (n-1)\,\rho_{\Lambda_n}(M_{ki})\Big)\\
&\quad + \delta_{li}\Big(\sum_{r=1}^{2n}\rho_{\Lambda_n}(M_{kr})\,\rho_{\Lambda_n}(M_{rj}) + (n-1)\,\rho_{\Lambda_n}(M_{kj})\Big) + \delta_{ki}\Big(\sum_{r=1}^{2n}\rho_{\Lambda_n}(M_{lr})\,\rho_{\Lambda_n}(M_{rj}) + (n-1)\,\rho_{\Lambda_n}(M_{lj})\Big)\\
&\quad + \delta_{kj}\Big(\sum_{r=1}^{2n}\rho_{\Lambda_n}(M_{lr})\,\rho_{\Lambda_n}(M_{ri}) + (n-1)\,\rho_{\Lambda_n}(M_{li})\Big)\bigg\}\otimes(v_k\vee v_l) - n\Big(v_i\vee v_j - \frac{\delta_{ij}}{2n}\sum_{s=1}^{2n} v_s\vee v_s\Big)I_{\Lambda_n} = 0\,.
\end{aligned}\tag{4.37}$$

Observing that, for $m = 1$, Eq. (4.11) becomes

$$\sum_{r=1}^{2n} \rho_{\Lambda_n}(M_{ir})\,\rho_{\Lambda_n}(M_{rl}) - (n-1)\,\rho_{\Lambda_n}(M_{il}) - \frac{2n-1}{4}\,\delta_{il}\,I_{\Lambda_n} = 0\,, \tag{4.38}$$

the scalar product of the lhs of Eq. (4.37) with $v_k \vee v_l$ becomes

$$[\rho_{\Lambda_n}(M_{ki}),\rho_{\Lambda_n}(M_{lj})]_+ - [\rho_{\Lambda_n}(M_{kl}),\rho_{\Lambda_n}(M_{ji})]_+ + \tfrac{1}{4}(\delta_{lj}\delta_{ki} + \delta_{kj}\delta_{li} - 2\delta_{ij}\delta_{kl})I_{\Lambda_n} = 0\,, \tag{4.39}$$

whence

$$\tfrac{1}{3}\{2[\rho_{\Lambda_n}(M_{ki}),\rho_{\Lambda_n}(M_{lj})]_+ - [\rho_{\Lambda_n}(M_{kj}),\rho_{\Lambda_n}(M_{il})]_+ - [\rho_{\Lambda_n}(M_{kl}),\rho_{\Lambda_n}(M_{ij})]_+\} = \tfrac{1}{4}(\delta_{kj}\delta_{il} - \delta_{kl}\delta_{ij})I_{\Lambda_n}\,, \tag{4.40}$$

which is precisely the result obtained from the symmetrized equations (2.16) and (2.14) (both satisfied by the representation ρ_{Λ_n}).

D. Comparison for the Lie algebras of type C_n

The inspection of Tables I and II shows that the derivation of the quantum polynomial identities, outlined in Sec. II, followed by the determination of their solutions leads, for the algebras of type C_n, to results which cannot be obtained directly using the HKO method. Indeed, the representations ρ_{Λ_k} $(k = 2,\ldots,n-1)$ which are solutions of the quantum tensorial identity (2.22) [cf. also Eq. (B10)] have no Okubo partners. The only HKO operators which satisfy equations of second degree are $\mathscr{O}_{m\Lambda_n,\Lambda_1}$.

Let us mention that the equation satisfied by this operator leads to polynomial identities for the representations $\rho_{m\Lambda_n}$ of sp$(2n,C)$ which coincide with the quantum tensorial identity (2.21).

Indeed, the HKO operator $\mathscr{O}_{m\Lambda_n,\Lambda_1}$ for the algebra sp$(2n,C)$ has the expression

$$\mathscr{O}_{m\Lambda_n,\Lambda_1} = \sum_{i,r,s=1}^{2n} g_{is}\,\rho_{m\Lambda_n}(S_{ir}) \otimes e_{rs} \tag{4.41}$$

(where $g_{ij} = \delta_{i,j+n} - \delta_{i+n,j}$) and satisfies the equation

$$(\mathscr{O}_{m\Lambda_n,\Lambda_1})^2 + (n+1)\mathscr{O}_{m\Lambda_n,\Lambda_1} - m(m+n+1)\,I_{m\Lambda_n} \otimes I_{\Lambda_1} = 0\,, \tag{4.42}$$

which, using the expression (4.41) of the operator, becomes

$$\sum_{i,r,s=1}^{2n} g_{is}\left\{\sum_{j,k=1}^{2n} g_{jk}\,\rho_{m\Lambda_n}(S_{jr})\,\rho_{m\Lambda_n}(S_{ik}) + (n+1)\,\rho_{m\Lambda_n}(S_{ir}) - m(m+n+1)\,g_{sr}\,I_{m\Lambda_n}\right\} \otimes e_{rs} = 0\,, \tag{4.43}$$

whence, by symmetrization, we get the quantum identities (2.21).

E. The comparison for the Lie algebras A_n

For Lie algebras of type A_n the correspondence between the tensorial identities derived using our method and those obtained by the HKO method becomes more involved, due to the fact that Lie algebras of type A_n admit second-degree adjoint tensors. This property allows the existence of inhomogeneous second-degree tensorial identities, which appear as linear combinations between adjoint tensors of second and first degrees.

Such identities can be obtained by the HKO method starting from the equation satisified by the HKO operator associated with the pairs of representations $\{\rho_{m\Lambda_k},\rho_{\Lambda_1}\}$, where $k = 1,\ldots,n$ (or with the conjugate pairs $\{\rho_{m\Lambda_{n-k+1}},\rho_{\Lambda_n}\}$). The HKO operator

$$\mathscr{O}_{m\Lambda_k,\Lambda_1} = \sum_{r,s=1}^{n+1} \rho_{m\Lambda_k}(A_{rs}) \otimes \rho_{\Lambda_1}(A_{sr}) \tag{4.44}$$

satisfies the identity (cf. Table II)

$$(\mathscr{O}_{m\Lambda_k,\Lambda_1} - m(1 - k/(n+1))I_{m\Lambda_k} \otimes I_{\Lambda_1}) \times (\mathscr{O}_{m\Lambda_k,\Lambda_1} + k(1 + m/(n+1))I_{m\Lambda_k} \otimes I_{\Lambda_1}) = 0\,. \tag{4.45}$$

Taking, in Eq. (4.45), the matrix elements of representation ρ_{Λ_1} we obtain the identity

$$\sum_{s=1}^{n+1} [\rho_{m\Lambda_k}(A_{js}),\rho_{m\Lambda_k}(A_{si})]_+ - \frac{2}{n+1}\,\delta_{ij} \sum_{r,s=1}^{n+1} \rho_{m\Lambda_k}(A_{rs})\,\rho_{m\Lambda_k}(A_{sr}) = 2\left(m - k + \frac{n+1}{2} - \frac{2mk}{n+1}\right)\rho_{m\Lambda_k}(A_{ji})\,. \tag{4.46}$$

The equality (4.46) expresses the following remarkable property: the second-degree adjoint tensors in the lhs are—in the representation $\rho_{m\Lambda_k}$—proportional with the first-degree adjoint tensors $\rho_{m\Lambda_k}(A_{ji})$.

Identities of this kind have not been obtained using our procedure because we restricted our search to *homogeneous* polynomials of second degree.

There is, however, one particular case in which the identities (4.46) coincide with those obtained by the method pointed out in Sec. II. Let us take, indeed, in Eq. (4.46)

$$k = (n+1)/2\,. \tag{4.47}$$

In this case, for any m (positive, integer), the numerical factor of the first-degree polynomial vanishes and the identity (4.46) obtained by the HKO method reduces to a homogeneous second-degree polynomial relation, which is precisely the quantum tensorial identity (2.8). As stated by Theorem 1 (Proposition 1), the representations which satisfy this identity are $\rho_{m\Lambda_{(n+1)/2}}$, $m = 1,2,\ldots$, in good agreement with the result just obtained.

The quantum tensorial identities (2.9) and (2.10) satisfied by the representations $m\Lambda_1$, $m\Lambda_n$, and Λ_k $(k = 1,\ldots,n)$ of the algebras of type A_n can be obtained by the HKO method, provided Eq. (4.46) is taken into account.

Note added in proof: After this paper was submitted, we realized[52] that our results can be systematized completely by using the coefficients c_i, d_i, $c_i^\vee$, $d_i^\vee$ $(i \in N \equiv \{1,2,\ldots,n\})$ in the expressions of the highest long root $(\alpha_{\max} = \alpha_{\rm hl} = \Sigma_{i=1}^n c_i\alpha_i)$, of the highest short root $(\alpha_{\rm hs} = \Sigma_{i=1}^n d_i\alpha_i)$, and of their duals $\alpha_{\rm hl}^\vee = \Sigma_{i=1}^n c_i^\vee \alpha_i^\vee$ and $\alpha_{\rm hs}^\vee = \Sigma_{i=1}^n d_i^\vee \alpha_i^\vee$

in terms of the simple roots α_i and coroots $\alpha_i^\vee$. Let us consider the sets of natural numbers $N_c = \{i\in N;\ c_i = 1\}$, $N_{d^\vee} = \{i\in N;\ d_i^\vee = 1\}$, and $N_{c^\vee} = \{i\in N;\ c_i^\vee = 1\}$. The pairs $\{\lambda, \mu\}$ of highest weights possessing the property defined in Theorem 2 are then of the form $\{\Lambda_i, m\Lambda_j\}$ with $i\in N_{d^\vee}$ and $j\in N_c$. The highest weights Λ_i with $i\in N_{d^\vee}$ are called minuscule weights, and the corresponding finite-dimensional irreducible representations have remarkable properties (cf. Ref. 53); the highest weights $m\Lambda_j$ with $j\in N_c$ have been obtained by Cavalli, D'Ariano, and Michel[54] from the condition that the orbits generated from the corresponding highest-weight vectors be Hermitian symmetric spaces. The fact that these spaces are classified by the set N_c is known.[55] The condition (D2) that determines the labels i of the Λ_i's that appear in Table I is equivalent with the condition $i\in N_{c^\vee}$. We remark that $N_{d^\vee} \subseteq N_{c^\vee}$. Let us recall that the representations of the Yangians that appear in Theorem 7 of Ref. 27 are classified by the set $N_c \cup N_{c^\vee}$. As the set N_c classifies the Hermitian symmetric spaces, it classifies also a part of the representations of the Yangians.[56]

APPENDIX A: TRANSFORMATIONS TO THE CARTAN-WEYL BASES

Let M_{ij} $(i,j = 1,\ldots,2n)$ be the generators of the algebra so$(2n,C)$, satisfying the structure relations (2.11). The transformation from the basis M_{ij} to the Cartan–Weyl basis, defined by the structure relations (3.7)–(3.11) with $\epsilon = +1$, is $(i = \sqrt{-1})$

$$M_{2k,2l-1} = -(i/2)(B_{kl} + C_{kl} - A_{kl} - A_{lk})\,, \quad \text{(A1)}$$

$$M_{2k-1,2l} = -(i/2)(B_{kl} + C_{kl} + A_{kl} + A_{lk})\,, \quad \text{(A2)}$$

$$M_{2k-1,2l-1} = -\tfrac{1}{2}(B_{kl} - C_{kl} - A_{kl} + A_{lk})\,, \quad \text{(A3)}$$

$$M_{2k,2l} = -\tfrac{1}{2}(B_{kl} - C_{kl} + A_{kl} - A_{lk})\,. \quad \text{(A4)}$$

In particular, $M_{2k-1,2k} = -iA_{kk}$; A_{kk} $(k = 1,\ldots,n)$ are generators of the Cartan subalgebra; A_{ij} $(i<j)$ and B_{ij} (any i,j) are raising operators; A_{ij} $(i>j)$ and C_{ij} (any i,j) are lowering operators.

For algebras so$(2n+1,C)$ one has to consider also the transformations

$$M_{2k,2n+1} = -(i/\sqrt{2})(a_k + b_k)\,, \quad \text{(A5)}$$

$$M_{2k-1,2n+1} = -(1/\sqrt{2})(a_k - b_k)\,, \quad \text{(A6)}$$

in which a_k (b_k) $(k = 1,\ldots,n)$ are raising (lowering) operators.

Let S_{ij} $(i,j = 1,\ldots,2n)$ be the generators of the algebra sp$(2n,C)$, which satisfy the structure relations (2.18). The transformations to the Cartan–Weyl basis, defined by the structure relations (3.7)–(3.11) with $\epsilon = -1$, are

$$A_{ij} = S_{i+n,j}\,, \quad B_{ij} = S_{i+n,j+n}\,, \quad C_{ij} = S_{ij}\,. \quad \text{(A7)}$$

APPENDIX B: DERIVATION OF THE SOLUTIONS OF THE SECOND-DEGREE TENSORIAL IDENTITIES USING THE WIGNER–ECKART THEOREM

The results of Table I can be rederived using the Wigner–Eckart theorem. To prove that a tensor operator T_Λ vanishes in a representation ρ_Ω [denoted, in the following, only by its highest weight (Ω)] it is sufficient to prove that in the Clebsch–Gordan series of the Kronecker product $(\Lambda)\otimes(\Omega)$ the representation (Ω) does not appear. This condition is, however, not necessary: the vanishing of the tensor $T_{\Lambda_1+\Lambda_n}$ in the representation $(m\Lambda_{n+1/2})$ (cf. Table I) is not a consequence of the Wigner–Eckart theorem; this phenomenon has been discussed in Sec. IV E.

In the following, we list, for each line of Table I, the Clebsch–Gordan series for the product of pairs of representations (Λ) and (Ω) belonging to columns 3 and 4 of Table I, respectively.

1. Algebras of type A_n

We have

$$\begin{aligned}(\Lambda_1+\Lambda_n)\otimes(m\Lambda_{(n+1)/2}) &= (\Lambda_1 + m\Lambda_{(n+1)/2} + \Lambda_n)\oplus(\Lambda_1+\Lambda_{(n-1)/2}+(m-1)\Lambda_{(n+1)/2})\\ &\oplus((m-1)\Lambda_{(n+1)/2}+\Lambda_{(n+3)/2}+\Lambda_n)\oplus(m\Lambda_{(n+1)/2})\\ &\oplus(\Lambda_{(n-1)/2}+(m-2)\Lambda_{(n+1)/2}+\Lambda_{(n+3)/2}),\end{aligned} \quad \text{(B1)}$$

$$\begin{aligned}(\Lambda_2+\Lambda_{n-1})\otimes(m\Lambda_1) &= (m\Lambda_1+\Lambda_2+\Lambda_{n-1})\oplus((m-1)\Lambda_1+\Lambda_3+\Lambda_{n-1})\\ &\oplus((m-1)\Lambda_1+\Lambda_2+\Lambda_n)\oplus((m-2)\Lambda_1+\Lambda_3+\Lambda_n),\end{aligned} \quad \text{(B2)}$$

$$\begin{aligned}(\Lambda_2+\Lambda_{n-1})\otimes(m\Lambda_n) &= (\Lambda_2+\Lambda_{n-1}+m\Lambda_n)\oplus(\Lambda_2+\Lambda_{n-2}+(m-1)\Lambda_n)\\ &\oplus(\Lambda_1+\Lambda_{n-1}+(m-1)\Lambda_n)\oplus(\Lambda_1+\Lambda_{n-2}+(m-2)\Lambda_n),\end{aligned} \quad \text{(B3)}$$

$$(2\Lambda_1+2\Lambda_n)\otimes(\Lambda_k) = (2\Lambda_1+\Lambda_k+2\Lambda_n)\oplus(\Lambda_1+\Lambda_{k+1}+2\Lambda_n)\oplus(2\Lambda_1+\Lambda_{k-1}+\Lambda_n)\oplus(\Lambda_1+\Lambda_k+\Lambda_n)\,. \quad \text{(B4)}$$

2. Algebras of type B_n

We have

$$(2\Lambda_1)\otimes(\Lambda_n) = (2\Lambda_1+\Lambda_n)\oplus(\Lambda_1+\Lambda_n)\,, \quad \text{(B5)}$$

$$(\Lambda_4)\otimes(m\Lambda_1) = (m\Lambda_1+\Lambda_4)\oplus((m-1)\Lambda_1+\Lambda_5)\oplus((m-1)\Lambda_1+\Lambda_3)\oplus((m-2)\Lambda_1+\Lambda_4)\,, \quad \text{(B6)}$$

$$(2\Lambda_2)\otimes(\Lambda_1) = (\Lambda_1+2\Lambda_2)\oplus(\Lambda_2+\Lambda_3)\oplus(\Lambda_1+\Lambda_2)\,, \quad \text{(B7)}$$

$$(2\Lambda_2)\otimes(\Lambda_n) = (2\Lambda_2+\Lambda_n)\oplus(\Lambda_1+\Lambda_2+\Lambda_n)\oplus(\Lambda_2+\Lambda_n)\,. \quad \text{(B8)}$$

3. Algebras of type C_n

We have

$$(\Lambda_2)\otimes(m\Lambda_n)=(\Lambda_2+m\Lambda_n)\oplus(\Lambda_1+\Lambda_{n-1}+(m-1)\Lambda_n)\oplus(\Lambda_{n-2}+(m-1)\Lambda_n),\tag{B9}$$

$$(4\Lambda_1)\otimes(\Lambda_k)=(4\Lambda_1+\Lambda_k)\oplus(3\Lambda_1+\Lambda_{k+1})\oplus(3\Lambda_1+\Lambda_{k-1})\oplus(2\Lambda_1+\Lambda_k),\tag{B10}$$

$$(2\Lambda_2)\otimes(\Lambda_1)=(\Lambda_1+2\Lambda_2)\oplus(\Lambda_2+\Lambda_3)\oplus(\Lambda_1+\Lambda_2).\tag{B11}$$

4. Algebras of type D_n

We have

$$(2\Lambda_1)\otimes(m\Lambda_{n-1})=(2\Lambda_1+m\Lambda_{n-1})\oplus(\Lambda_1+(m-1)\Lambda_{n-1}+\Lambda_n)\oplus((m-2)\Lambda_{n-1}+2\Lambda_n),\tag{B12}$$

$$(2\Lambda_1)\otimes(m\Lambda_n)=(2\Lambda_1+m\Lambda_n)\oplus(\Lambda_1+\Lambda_{n-1}+(m-1)\Lambda_n)\oplus(2\Lambda_{n-1}+(m-2)\Lambda_n),\tag{B13}$$

$$(\Lambda_4)\otimes(m\Lambda_1)=(m\Lambda_1+\Lambda_4)\oplus((m-1)\Lambda_1+\Lambda_5)\oplus((m-1)\Lambda_1+\Lambda_3)\oplus((m-2)\Lambda_1+\Lambda_4),\tag{B14}$$

$$(2\Lambda_2)\otimes(\Lambda_1)=(\Lambda_1+2\Lambda_2)\oplus(\Lambda_2+\Lambda_3)\oplus(\Lambda_1+\Lambda_3),\tag{B15}$$

$$(2\Lambda_2)\otimes(\Lambda_{n-1})=(2\Lambda_2+\Lambda_{n-1})\oplus(\Lambda_1+\Lambda_2+\Lambda_n)\oplus(\Lambda_2+\Lambda_{n-1}),\tag{B16}$$

$$(2\Lambda_2)\otimes(\Lambda_n)=(2\Lambda_2+\Lambda_n)\oplus(\Lambda_1+\Lambda_2+\Lambda_{n-1})\oplus(\Lambda_2+\Lambda_n).\tag{B17}$$

APPENDIX C: KRONECKER PRODUCTS OF REPRESENTATIONS THAT SATISFY THE SECOND-DEGREE QUANTUM TENSORIAL IDENTITIES

As pointed out by Theorem 2, the pairs of representations belonging to the same line of Table II and to columns 2 and 3 are the only pairs of which the Kronecker products decompose into precisely two irreducible components.

The proof of the unicity is provided by this Appendix, in which the products of the other representations contained in columns 2 and 3 of Table II are listed. The representations are labeled by their highest weights in parentheses. The symbol (Λ_0) denotes the zero weight.

For algebras of type A_n, the proof of the unicity results from the rules for Kronecker products of representations labeled by Young tableaux.

1. Algebras of type B_n

For $n=2$,

$$((m+p)\Lambda_1)\otimes(m\Lambda_1)=\bigoplus_{j=0}^{m}\bigoplus_{k=0}^{j}((2j-2k+p)\Lambda_1+2k\Lambda_2).\tag{C1a}$$

For $n\geqslant 3$,

$$((m+p)\Lambda_1)\otimes(m\Lambda_1)=\bigoplus_{j=0}^{m}\bigoplus_{k=0}^{j}((2j-2k+p)\Lambda_1+k\Lambda_2),\tag{C1b}$$

$$(\Lambda_n)\otimes(\Lambda_n)=\bigoplus_{k=1}^{n}(\Lambda_{n-k})\oplus(2\Lambda_n).\tag{C2}$$

2. Algebras of type C_n

We have

$$((m+p)\Lambda_n)\otimes(m\Lambda_n)=\bigoplus_{\Sigma_{i=1}^{n}k_i=0}^{m}\left(\sum_{i=1}^{n}2k_i\Lambda_i+p\Lambda_n\right),\tag{C3}$$

$$(\Lambda_k)\otimes(\Lambda_{n-q})=\bigoplus_{i=0}^{k}\bigoplus_{j=0}^{\min(q,i)}(\Lambda_{k-i}+\Lambda_{n-q-i+2j}),\tag{C4}$$

$$(\Lambda_k)\otimes(m\Lambda_n)=\bigoplus_{i=0}^{k}(\Lambda_i+\Lambda_{n-k+i}+(m-1)\Lambda_n).\tag{C5}$$

3. Algebras of type D_n

We have

$$(p\Lambda_1)\otimes(q\Lambda_1)=\bigoplus_{k=0}^{\min(p,q)}\bigoplus_{i=0}^{k}((p+q-2k)\Lambda_1+(k-i)\Lambda_2),\tag{C6}$$

$$(p\Lambda_1)\otimes(q\Lambda_n)=\bigoplus_{i=0}^{\min(p,q)}((p-i)\Lambda_1+i\Lambda_{n-1}+(q-i)\Lambda_n),\tag{C7}$$

$$(p\Lambda_n)\otimes(q\Lambda_n)=\bigoplus_{\Sigma_{k=1}^{[n/2]}i_k=0}^{\min(p,q)}\left(\sum_{k=1}^{[n/2]}i_k\Lambda_{n-2k}+\left(p+q-2\sum_{k=1}^{[n/2]}i_k\right)\Lambda_n\right),\tag{C8}$$

$$(p\Lambda_1)\otimes(q\Lambda_{n-1})=\bigoplus_{i=0}^{\min(p,q)}((p-i)\Lambda_1+(q-i)\Lambda_{n-1}+i\Lambda_n),\tag{C9}$$

$$(p\Lambda_{n-1})\otimes(q\Lambda_{n-1})=\bigoplus_{\Sigma_{k=1}^{[n/2]}i_k=0}^{\min(p,q)}\left(\sum_{k=1}^{[n/2]}i_k\Lambda_{n-2k}+\left(p+q-2\sum_{k=1}^{[n/2]}i_k\right)\Lambda_{n-1}\right),\tag{C10}$$

$$(p\Lambda_{n-1})\otimes(q\Lambda_n)=\bigoplus_{l=0}^{\min(p,q)}\bigoplus_{\Sigma_{k=0}^{[(n-3)/2]-1}i_k=0}^{\min(p-l,q-l)}\left(\sum_{k=0}^{[(n-3)/2]-1}i_k\Lambda_{n-3-2k}+\left(p-l-\sum_{k=0}^{[(n-3)/2]-1}i_k\right)\Lambda_{n-1}+\left(q-l-\sum_{k=0}^{[(n-3)/2]-1}i_k\right)\Lambda_n\right).\tag{C11}$$

TABLE III. Solutions (Λ) of Drinfeld's equations (column 8); the proofs (column 7) are based on formulas (D1)–(D3). Bourbaki's conventions are used for the simple roots α_i of the semisimple Lie algebras; their label i, ranging from 1 to n, is indicated in column 4 in order to label the coefficients c_i, the ratios r_i, and the fundamental highest weights Λ_i.

Lie algebra	$\alpha_{\max}=\sum_{i=1}^{n} c_i\alpha_i$	$(\alpha_{\max},\alpha_{\max})$	i	(α_i,α_i)	r_i	c_i	Proof	(Λ)
B_n	$\alpha_1+2\sum_{i=2}^{n}\alpha_i$	2	1	2	1	1	$c_1=1$ $c_1=r_1$	$m\Lambda_1$ Λ_1
			2 $\vdots$ $n-1$	2	1	2	$c_i\neq 1$ $c_i\neq r_i$	
			n	1	2	2	$c_n=r_n$	Λ_n
C_n	$2\sum_{i=1}^{n-1}\alpha_i+\alpha_n$	4	1 $\vdots$ $n-1$	2	2	2	$c_i=r_i$	Λ_i $(i=1,\ldots,n)$
			n	4	1	1	$c_n=r_n$ $c_n=1$	$m\Lambda_n$
D_n	$\alpha_1+2\sum_{i=2}^{n-2}\alpha_i+\alpha_{n-1}+\alpha_n$	2	1	2	1	1	$c_1=1$ $c_1=r_1$	$m\Lambda_1$ Λ_1
			2 $\vdots$ $n-2$	2	1	2	$c_i\neq 1$ $c_i\neq r_i$	
			$n-1$	2	1	1	$c_{n-1}=1$ $c_{n-1}=r_{n-1}$	$m\Lambda_{n-1}$ Λ_{n-1}
			n	2	1	1	$c_n=1$ $c_n=r_n$	$m\Lambda_n$ Λ_n
A_n	$\sum_{i=1}^{n}\alpha_i$	2	1 $\vdots$ n	2	1	1	$c_i=1$ $c_i=r_i$	$m\Lambda_i$ $(i=1,\ldots,n)$ Λ_i $(i=1,\ldots,n)$

APPENDIX D: COMPARISON BETWEEN DRINFELD'S SOLUTIONS AND THE SOLUTIONS OF THE SECOND-DEGREE QUANTUM TENSORIAL IDENTITIES

In this appendix we show that the representations obtained by Drinfeld (Ref. 27, Theorem 7) from the condition that a set of third-degree polynomials in the enveloping algebra vanish are exactly the representations listed in column 4 of Table I, deduced in Sec. III from the condition that second-degree polynomials in the enveloping algebra vanish.

The coincidence is perfect only for Lie algebras of types B_n, D_n, and C_n. For Lie algebras of type A_n, as explained in Sec. IV E and in Ref. 27 (Theorem 9), the existence of second-degree adjoint tensors gives rise to new identities in our case and to new solutions in Drinfeld's case.

The representation $(m\Lambda_i)$ is a solution of Drinfeld's equations if the coefficient c_i of the simple root α_i in the expression of the highest root $\alpha_{\max}$ as a linear combination of simple roots is

$$c_i=1\,. \tag{D1}$$

The representation (Λ_i) is a solution of Drinfeld's equations if

$$c_i=r_i\,, \tag{D2}$$

where

$$r_i=(\alpha_{\max},\alpha_{\max})/(\alpha_i,\alpha_i)\,. \tag{D3}$$

Table III presents the values of the coefficients c_i, r_i as well as (in column 7) Drinfeld's solutions; the arguments for the presence of these solutions [Eqs. (D2) or (D3)] are indicated in column 8 of this table.

[1]P. A. M. Dirac, Proc. R. Soc. London Ser. A **155**, 447 (1936).
[2]J. D. Louck, "Theory of angular momentum in N-dimensional space," Los Alamos Scientific Lab., 1960; Am. J. Phys. **31**, 378 (1963); J. Math. Phys. **6**, 1786 (1965).
[3]H. Bacry, Nuovo Cimento A **41**, 222 (1966).
[4]J. Rosen and P. Roman, J. Math. Phys. **7**, 2072 (1966).
[5]R. H. Pratt and T. F. Jordan, Phys. Rev. **148**, 1276 (1966).
[6]N. Mukunda, J. Math. Phys. **8**, 1069 (1967).
[7]G. Györgyi, Nuovo Cimento A **53**, 717 (1968); **62**, 449 (1969).
[8]G. Györgyi and S. Kövesi-Domokos, Nuovo Cimento B **58**, 191 (1968).
[9]A. O. Barut and A. Böhm, J. Math. Phys. **11**, 2938 (1970).
[10]I. Gyuk and H. Umezawa, Phys. Rev. D **3**, 899 (1971).
[11]L. P. Staunton, J. Math. Phys. **19**, 1471 (1978).
[12]I. V. Komarov, Teor. Mat. Fiz. **50**, 402 (1982).
[13]Y. Lehrer-Ylamed, Bull. Res. Counc. Isr. A **5**, 197 (1956).
[14]A. J. Bracken and H. S. Green, J. Math. Phys. **12**, 2099 (1971).
[15]H. S. Green, J. Math. Phys. **12**, 2106 (1971).
[16]K. C. Hannabuss, "Characteristic equations for semisimple Lie groups," Mathematical Institute preprint, Oxford, 1972.
[17]S. Okubo, "Symmetric representations of the U(n) group and exactly solvable nuclear model," University of Rochester preprint UR-47, 1974.
[18]S. Okubo, J. Math. Phys. **16**, 528 (1975).
[19]S. Okubo, J. Math. Phys. **18**, 2382 (1977).
[20]D. M. O'Brien, A. Cant, and A. L. Carey, Ann. Inst. H. Poincaré **26**, 405 (1977).
[21]M. D. Gould, J. Austral. Math. Soc. B **20**, 290 (1978); **26**, 257 (1985).

[22]M. Iosifescu and H. Scutaru, J. Math. Phys. **21**, 2033 (1980).
[23]M. Iosifescu and H. Scutaru, J. Math. Phys. **25**, 2856 (1984).
[24]M. Iosifescu and H. Scutaru, "Polynomial constraints associated to dynamical symmetries," Central Institute of Physics preprint FT-253, Bucharest, 1984; "Polynomial identities satisfied by realizations of Lie algebras," in *Geometrodynamics Proceedings*, edited by A. Pràstaro (World Scientific, Singapore, 1985).
[25]M. Iosifescu and H. Scutaru, "Representations which satisfy identities associated with tensor operators," Communications of the JINR, Dubna, USSR, E4-85-525,526, 1985; "Kronecker products which decompose into two irreducible representations," Communications of the JINR, Dubna, USSR, E4-85-527, 1985.
[26]M. Iosifescu and H. Scutaru, "Polynomial identities associated with dynamical symmetries," in *Conformal Groups and Related Symmetries. Physical Results and Mathematical Background*, edited by A. O. Barut and H. D. Doebner, *Lecture Notes in Physics*, Vol. 261 (Springer, Berlin, 1986).
[27]V. G. Drinfeld, Dokl. Akad. Nauk SSSR **283**, 1060 (1985).
[28]B. Kostant, J. Funct. Anal. **20**, 257 (1975).
[29]N. Yu. Reshetikhin and L. D. Faddeev, Teor. Mat. Fiz. **56**, 323 (1983).
[30]N. Yu. Reshetikhin, Teor. Mat. Fiz. **63**, 347 (1985).
[31]H. Scutaru, J. Phys. A **20**, L1053 (1987).
[32]F. A. Berezin, Commun. Math. Phys. **63**, 131 (1978).
[33]N. Papanicolaou, Ann. Phys. (NY) **136**, 210 (1981).
[34]M. Yamamura, Prog. Theor. Phys. **52**, 538 (1974).
[35]S. Nishiyama, Prog. Theor. Phys. **55**, 1146 (1976).
[36]L. D. Mlodinow and N. Papanicolaou, Ann. Phys. (NY) **131**, 1 (1981).
[37]J. Deenen and C. Quesne, J. Math. Phys. **23**, 878, 2004 (1982).
[38]M. Iosifescu and H. Scutaru, Phys. Lett. A **109**, 433 (1985); J. Math. Phys. **27**, 524 (1986).
[39]I. N. Bernstein, I. M. Gelfand, and S. I. Gelfand, Funkt. Anal. Prilozhen. **9**, 61 (1975).
[40]L. C. Biedenharn and D. E. Flath, "Tensor operators as an extension of the universal enveloping algebra," in *Group Theoretical Methods in Physics*, Proceedings, Trieste, 1983, edited by G. Denardo, G. Ghirardi, and T. Weber, *Lecture Notes in Physics*, Vol. 201 (Springer, Berlin, 1984); Commun. Math. Phys. **93**, 143 (1984).
[41]A. J. Bracken, Commun. Math. Phys. **94**, 371 (1984); A. J. Bracken and J. H. MacGibbon, J. Phys. A **17**, 2581 (1984).
[42]M. Couture and R. T. Sharp, J. Phys. A **13**, 1925 (1980).
[43]Y. Giroux, M. Couture, and R. T. Sharp, J. Phys. A **17**, 715 (1984).
[44]R. Gilmore and J. F. Draayer, J. Math. Phys. **26**, 3053 (1985).
[45]J. Dixmier, *Enveloping Algebras* (North-Holland, Amsterdam, 1977).
[46]S. Okubo, J. Math. Phys. **20**, 586 (1979); **23**, 8 (1982).
[47]P. D. Jarvis, J. Phys. A **12**, 1 (1979).
[48]M. Ikeda and T. Maekawa, Math. Japon. **16**, 173 (1971).
[49]S. A. Edwards, J. Math. Phys. **17**, 164 (1978).
[50]E. B. Dynkin, Trudy Moskov. Mat. Obšč. **1**, 37 (1952).
[51]A. Borel and N. Wallach, *Continuous Cohomology, Discrete Subgroups and Representations of Reductive Groups* (Princeton U. P., Princeton, NJ, 1980).
[52]M. Iosifescu and H. Scutaru, "Degenerate representations from quantum kinematical constraints," talk given at the 16th International Colloquium "Group theoretical methods in physics," Varna, 1987.
[53]N. Bourbaki, *Groupes et algèbres de Lie, Eléments de mathématique*, Vol. XXXVIII (Hermann, Paris, 1968), Chap. VIII, §7.3.
[54]A. Cavalli, G. D'Ariano, and L. Michel, Ann. Inst. H. Poincaré **44**, 173 (1986).
[55]J. A. Wolf, *Spaces of Constant Curvature* (University of California, Berkeley, 1972), Chap. 8, Theorem 8.10.9.
[56]V. G. Drinfeld, "Kvantovye gruppy," in *Zapiski Nauchnykh Seminarov* LOMI (Nauka, Leningrad, 1987) (extended version of "Quantum Groups," I.C.M. Berkeley Congress, 1986), final remark.

The matrix representations of g_2. I. Representations in an so(4) basis

R. Le Blanc
A. W. Wright Nuclear Structure Laboratory, Yale University, New Haven, Connecticut 06511
D. J. Rowe
Physics Department, University of Toronto, Toronto, Ontario, Canada M5S 1A7

(Received 4 June 1987; accepted for publication 28 October 1987)

Irreducible representations of the real compact Lie algebra g_2 are given in $g_2 \supset \mathrm{so}(4)$ bases. All missing labels are accounted for by the explicit construction of a $g_2 \supset \mathrm{so}(4)$ basis of vector holomorphic functions. The general problem of missing internal labels is also briefly discussed.

I. INTRODUCTION

The Lie algebra g_2 presents many interesting features. First, it has found many physical applications, briefly reviewed by Beckers, Hussin, and Winternitz[1] who also extensively studied its various and numerous real forms. It is also a rank-2 Lie algebra of order 14 and its algebraic structure is therefore relatively easy to handle.

Although the representation theory of g_2 has received much attention,[2–5] much remains to be done since any attempt to construct irreducible representations of g_2 has to confront a recurring problem of representation theory, namely the appearance of missing labels. More precisely, when bases for its irreducible representations are decomposed with respect to the subalgebra chains $g_2 \supset \mathrm{so}(4)$ and $g_2 \supset \mathrm{su}(3)$, one has to provide, respectively, 2 and 1 supplementary labels to obtain a complete basis specification.

A widely used and convenient approach to the resolution of the missing label problem is the introduction of an integrity basis[6] for tensorial operators belonging to the enveloping algebra of a Lie algebra **g** that are scalar in a subalgebra **h**, the eigenvalues of which will provide the missing labels for the desired $\mathbf{g} \supset \mathbf{h}$ basis. Alternative methods are the use of shift operators[7] or path labels.[2] These methods are usually applicable to the most general cases but their implementations are often complicated.

In this paper, we successfully address the missing label problem for g_2 by exploiting the versatility of vector coherent state (VCS) theory[8,11] and give an explicit basis construction for all irreducible ladder representations of g_2 when these are reduced with respect to $g_2 \supset \mathrm{so}(4)$. We show that VCS theory naturally resolves the missing labels problem in a manner that directly appeals to specific aspects of character theory. Finally, we discuss at some length the specific nature and group-theoretical significance of the various internal labels introduced by VCS theory.

II. THE g_2 LIE ALGEBRA

A. The Cartan basis

The complex extension g_2^c of the Lie algebra g_2 is a simple rank-2 Lie algebra of order 14.[1,2,12] Its root diagram (Fig. 1) has the form of a "star of David." A possible choice of positive roots is given by the set

$$\Delta_+ = \{\alpha_1, \alpha_2, \alpha_1+\alpha_2, \alpha_1+2\alpha_2, \alpha_1+3\alpha_2, 2\alpha_1+3\alpha_2\}, \tag{2.1}$$

where, using the Killing form, we give the following normalization:

$$(\alpha_1,\alpha_2)=2, \quad (\alpha_1,\alpha_2) = -1, \quad (\alpha_2,\alpha_2)=\tfrac{2}{3}, \tag{2.2}$$

for the two fundamental simple roots α_1 and α_2. The set of all nonzero roots is then given by

$$\Delta = \Delta_+ \oplus \Delta_-, \quad \Delta_- = -\Delta_+.$$

An alternative set of positive roots will also be used in this paper [cf. Eq. (3.6)].

Basis vectors for g_2^c are given by an orthonormal basis $\{h_1 h_2\}$ for the Cartan subalgebra

$$\mathscr{H} = \mathrm{span}\{h_1 h_2\}, \tag{2.3a}$$

corresponding to a pair of null root vectors, and a set of vectors

$$\{l_\rho;\rho\in\Delta\}. \tag{2.3b}$$

Elements of the Cartan subalgebra are put into one-to-one correspondence with the root vectors in the canonical way with

$$h_\rho = \sum \rho^i h_{\alpha_i} \sim \rho = \sum \rho^i \alpha_i.$$

The algebraic structure of g_2^c is then given by

$$\begin{aligned}&[h_\rho,h_\mu]=0, \quad [h_\rho,l_\mu]=(\rho,\mu)l_\mu,\\ &[l_\rho,l_{-\rho}]=h_\rho, \quad [l_\rho,l_\mu]=N_{\rho,\mu}l_{\rho+\mu}, \quad \rho+\mu\neq 0.\end{aligned} \tag{2.4}$$

We necessarily have

$$N_{\rho,\mu} = -N_{\mu,\rho} \tag{2.5a}$$

and also require

$$N_{\rho,\mu} = -N_{-\rho,-\mu} = N_{\rho,-\mu-\rho}. \tag{2.5b}$$

With the normalization (2.2), we can set

$$\begin{aligned}&N_{\alpha_1,\alpha_2}=1, \quad N_{\alpha_1+\alpha_2,\alpha_2} = -2\sqrt{3}/3, \quad N_{\alpha_1+2\alpha_2,\alpha_2} = -1,\\ &N_{\alpha_1+2\alpha_2,\alpha_1+\alpha_2} = -1, \quad N_{\alpha_1+3\alpha_2,\alpha_1} = -1.\end{aligned} \tag{2.6}$$

B. The $g_2 \supset \mathrm{su}(3) \supset \mathrm{u}(2)$ chain

Defining

$$\begin{aligned}&g_{12}=l_{\alpha_1}, \quad g_{21}=l_{-\alpha_1},\\ &g_{23}=l_{\alpha_1+3\alpha_2}, \quad g_{32}=l_{-\alpha_1-3\alpha_2},\\ &g_{13}=l_{2\alpha_1+3\alpha_2}, \quad g_{31}=l_{-2\alpha_1-3\alpha_2},\end{aligned} \tag{2.7a}$$

and

 0022-2488/88/040758-09$02.50

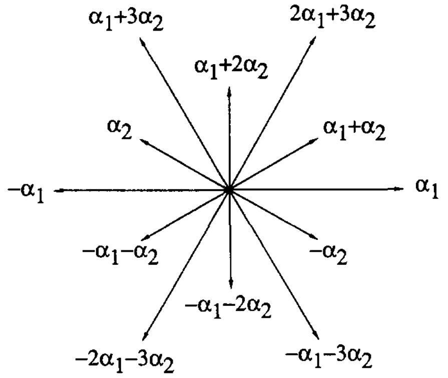


FIG. 1. Root diagram for g_2.

$$g_{11}-g_{22}=\sqrt{2}h_1,\quad g_{11}+g_{22}-2g_{33}=\sqrt{6}h_2, \tag{2.7b}$$

we easily verify, with restriction of the base field to $\mathbb{R}$, that the long roots of Δ and Cartan subalgebra $\mathscr{H}$ define an su(3) real subalgebra with commutation relations inferred from the u(3) relations

$$[g_{ij},g_{kl}]=\delta_{jk}g_{il}-\delta_{il}g_{kj}. \tag{2.8}$$

Defining the tensor operators **e** and **f** with the remaining roots (see Fig. 2),

$$\begin{aligned} &e_1=\sqrt{3}l_{\alpha_1+\alpha_2},\quad f_1=\sqrt{3}l_{-\alpha_1-\alpha_2},\\ &e_2=\sqrt{3}l_{\alpha_2},\quad f_2=\sqrt{3}l_{-\alpha_2},\\ &e_3=\sqrt{3}l_{-\alpha_1-2\alpha_2},\quad f_3=\sqrt{3}l_{\alpha_1+2\alpha_2},\end{aligned} \tag{2.9}$$

we then find with the help of Eqs. (2.4)–(2.6) the following commutation relations (see also Humphreys,[12] Sec. 19.3):

$$[g_{ij},e_k]=\delta_{jk}e_i,\quad [g_{ij},f_k]=-\delta_{ik}f_j, \tag{2.10a}$$

$$\begin{aligned}&[e_i,e_j]=-2\epsilon_{ijk}f_k,\quad [f_i,f_j]=2\epsilon_{ijk}e_k,\\ &[e_i,f_j]=3g_{ij}-\delta_{ij}g_{kk}.\end{aligned} \tag{2.10b}$$

According to (2.10a), **e** and **f** transform, respectively, as {10} and {11} tensor operators under su(3).

We require the Hermiticity conditions

$$g_{ij}=g_{ji}^{\dagger},\quad f_i=e_i^{\dagger}, \tag{2.11}$$

in order that the Lie algebra representations exponentiate to unitary representations of the group.

It will be useful for the following to consider the canonical (Gel'fand) u(2) subalgebra of su(3) to be given by the set of generators:

$$\mathrm{u}(2)=\mathrm{span}\{g_{12},g_{21},g_{11}-g_{33},g_{22}-g_{33}\}, \tag{2.12}$$

where the last two weight operators are linear combinations of the Cartan weight operators h_1 and h_2 [see Eq. (2.7b)].

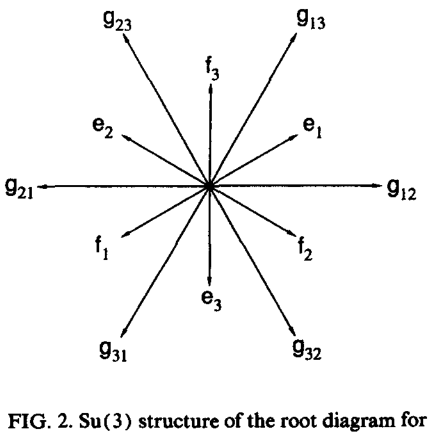


FIG. 2. Su(3) structure of the root diagram for g_2.

C. The $g_2\supset$so(4)$\supset$u(2) chain

It is well known that so(4) is a nonsimple Lie algebra isomorphic to the direct sum su(2) $\oplus$ su(2). Its root diagram is given by two orthogonal A_1 [the complexification of su(2)] root diagrams.

A Cartan basis for the complex extension of any so($2n$) $\supset$u(n) Lie algebra chain is given by the set[10–13]

$$\{c_{ij},a_{ij},b_{ij};\ 1\leqslant i,j\leqslant n\}, \tag{2.13a}$$

where $\{c_{ij}\}$ spans the u(n) subalgebra and

$$a_{ij}=-a_{ji},\quad b_{ij}=-b_{ji}. \tag{2.13b}$$

The commutations relations for so($2n$) are then given by

$$\begin{aligned}&[c_{ij},a_{kl}]=\delta_{jk}a_{il}+\delta_{jl}a_{ki},\quad [c_{ij},b_{kl}]=-\delta_{ik}b_{jl}-\delta_{il}b_{kj},\\ &[a_{ij},a_{kl}]=0,\quad [b_{ij},b_{kl}]=0,\\ &[b_{ij},a_{kl}]=\delta_{jk}c_{li}+\delta_{il}c_{kj}-\delta_{ik}c_{lj}-\delta_{jl}c_{ki},\\ &[c_{ij},c_{kl}]=\delta_{jk}c_{il}-\delta_{il}c_{kj}.\end{aligned} \tag{2.14}$$

According to the above commutation relations, the Cartan raising and lowering operators $\{a_{ij}\}$ and $\{b_{ij}\}$ transform as {11} and $\{-1,-1\}$ tensors, respectively, under transformations generated by the u(n) subalgebra $\{c_{ij}\}$.

It is easily verified that a possible embedding for the so(4)$\supset$u(2) subalgebra of g_2, when expressed in the above so($2n$) Cartan basis with $n=2$ and in terms of the $g_2\supset$su(3)$\supset$u(2) basis (2.7) and (2.9) is given by

$$\begin{aligned}&c_{12}=g_{12},\quad c_{21}=g_{21},\quad c_{11}=g_{11}-g_{33},\\ &c_{22}=g_{22}-g_{33},\quad a_{12}=-f_3,\quad b_{12}=-e_3.\end{aligned} \tag{2.15}$$

Note that the two chains have a u(2) subalgebra in common.

The above basis for the so(4) Lie algebra is isomorphic to the direct sum su(2) $\oplus$ su(2). Bases for the two commuting su(2) Lie algebras are given by

$$\{I_+,I_-,I_0\}=\{c_{12},c_{21},\tfrac{1}{2}(c_{11}-c_{22})\} \tag{2.16a}$$

and

$$\{S_+,S_-,S_0\}=\{a_{12},b_{12},\tfrac{1}{2}(c_{11}+c_{22})\}, \tag{2.16b}$$

respectively, thereby defining the so(4) $\sim$ su(2) $\oplus$ su(2) isomorphism. We shall distinguish these two su(2) subalgebras by denoting them su$_I$(2) and su$_S$(2). Irreps $[\lambda_1\lambda_2]$ of so(4) will be labeled by the eigenvalues λ_1 and λ_2 of the weight operators c_{11} and c_{22} on their highest-weight state. The corresponding su$_I$(2) $\oplus$ su$_S$(2) labels are then clearly

$$(I,S)=(\tfrac{1}{2}(\lambda_1-\lambda_2),\tfrac{1}{2}(\lambda_1+\lambda_2)). \tag{2.17}$$

The basis (2.15) will be preferred for the branching rule discussion (Sec. III) while (2.16) will be preferred in computations (Sec. IV).

The remaining 8 ($=14-6$) generators of g_2 complementary to the above so(4)$\supset$u(2) subalgebra are found to transform under so(4) as components of a single irreducible

[21] ($(I,S) = (\frac{1}{2},\frac{3}{2})$) tensor (Ref. 14) [see Eq. (2.23) and Fig. 3]. In both Gel'fand

$$t\begin{bmatrix} m_{12} & & m_{22} \\ & m_{11} & \end{bmatrix} \tag{2.18a}$$

and angular momentum notation

$$t^{1/2M_I}_{3/2M_S} = t^{M_I}_{M_S}, \tag{2.18b}$$

we find the following correspondences [see Sec. III B for the so(4)$\downarrow$u(2) branching rule]:

$$\begin{aligned} t\begin{bmatrix} 2 & & 1 \\ & 2 & \end{bmatrix} &= t^{+1/2}_{+3/2} = g_{13} = l_{2\alpha_1+3\alpha_2}, \\ t\begin{bmatrix} 1 & & 0 \\ & 1 & \end{bmatrix} &= t^{+1/2}_{+1/2} = e_1/\sqrt{3} = l_{\alpha_1+\alpha_2}, \\ t\begin{bmatrix} 0 & & -1 \\ & 0 & \end{bmatrix} &= t^{+1/2}_{-1/2} = f_2/\sqrt{3} = l_{-2\alpha_2}, \\ t\begin{bmatrix} -1 & & -2 \\ & -1 & \end{bmatrix} &= t^{+1/2}_{-3/2} = -g_{32} = l_{-\alpha_1-3\alpha_2}, \end{aligned} \tag{2.19a}$$

and

$$\begin{aligned} t\begin{bmatrix} 2 & & 1 \\ & 1 & \end{bmatrix} &= t^{-1/2}_{+3/2} = g_{23} = l_{\alpha_1+3\alpha_2}, \\ t\begin{bmatrix} 1 & & 0 \\ & 0 & \end{bmatrix} &= t^{-1/2}_{+1/2} = e_2/\sqrt{3} = l_{\alpha_2}, \\ t\begin{bmatrix} 0 & & -1 \\ & -1 & \end{bmatrix} &= t^{-1/2}_{-1/2} = -f_1/\sqrt{3} = -l_{-\alpha_1-\alpha_2}, \\ t\begin{bmatrix} -1 & & -2 \\ & -2 & \end{bmatrix} &= t^{-1/2}_{-3/2} = g_{31} = l_{-2\alpha_1-3\alpha_2}. \end{aligned} \tag{2.19b}$$

For unitarity, we require

$$(t^\mu_\nu)^\dagger = (-1)^{1/2-\mu}(-1)^{3/2-\nu} t^{-\mu}_{-\nu}. \tag{2.20}$$

The transformation properties of the tensor t under so(4)$\sim$su(2)$\oplus$su(2) are conveniently expressed in tensorial form by

$$\begin{aligned} [I_\alpha, t^\mu_\nu] &= [\tfrac{1}{2}(\tfrac{1}{2}+1)]^{1/2}\langle \tfrac{1}{2}\mu;1\alpha|\tfrac{1}{2}\mu+\alpha\rangle t^{\mu+\alpha}_\nu, \\ [S_\beta, t^\mu_\nu] &= [\tfrac{3}{2}(\tfrac{3}{2}+1)]^{1/2}\langle \tfrac{3}{2}\nu;1\beta|\tfrac{3}{2}\nu+\beta\rangle t^\mu_{\nu+\beta}. \end{aligned} \tag{2.21}$$

The commutation relations of its components are given by

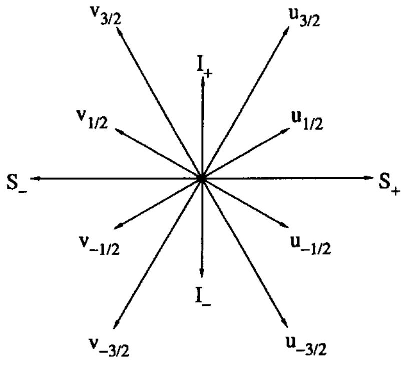


FIG. 3. So(4) structure of the root diagram for g_2.

$$\begin{aligned} [t^{+1/2}_\mu, t^{+1/2}_\nu] &= (-1)^{3/2-\nu}\delta_{\mu,-\nu} I_+, \\ [t^{-1/2}_\mu, t^{-1/2}_\nu] &= (-1)^{3/2-\mu}\delta_{\mu,-\nu} I_-, \\ [t^{+1/2}_\mu, t^{-1/2}_\nu] &= (2\sqrt{5}/3)\langle \tfrac{3}{2}\mu;\tfrac{3}{2}\nu|1\nu+\mu\rangle S_{\mu+\nu} \\ &\quad + \delta_{\mu,-\nu}(-1)^{3/2-\mu} I_0. \end{aligned} \tag{2.22}$$

For notational ease, it is useful to define the following $su_S(2)$ tensor operators $\{u\}$ and $\{v\}$:

$$u_\mu = t^{+1/2}_\mu, \tag{2.23a}$$

$$v_\mu = t^{-1/2}_\mu, \tag{2.23b}$$

both of which carry an irrep $S = \frac{3}{2}$ of the $su_S(2)$ Lie algebra. These tensors respectively raise and lower the I_0 weight of basis states.

III. THE STATE LABELING PROBLEM

A. State labeling for the $g_2 \supset$ su(3) $\supset$ u(2) chain

A unitary irrep of g_2 is labeled in the notation of King and Qubanchi[15] by $[\![\mu_1,\mu_2]\!]$, where μ_1 and μ_2 are the respective eigenvalues of the weight operators

$$h_{\alpha_1+3\alpha_2} = g_{22} - g_{33} \quad \text{and} \quad h_{\alpha_1} = g_{11} - g_{22} \tag{3.1a}$$

on the highest-weight state defined by

$$l_\rho |[\![\mu_1\mu_2]\!] hw\rangle = 0, \quad \forall \rho \in \Delta_+. \tag{3.1b}$$

The branching rule $g_2\downarrow$su(3) has been derived by King and Qubanchi[15] and by Perroud.[2] Curiously, it is related to the *weight* decomposition of an associated su(3) representation; more precisely, we have

$g_2\downarrow$su(3):

$$[\![\mu_1\mu_2]\!]\downarrow \sum_{\nu_1 > \nu_2 > 0} M^{\{\mu_1\mu_2\}}_{(\nu_1-\nu_2,\nu_2,\mu_1+\mu_2-\nu_1)}\{\nu_1\nu_2\}, \tag{3.2a}$$

where $M^{\{\mu_1\mu_2\}}_{(\nu_1-\nu_2,\nu_2,\mu_1+\mu_2-\nu_1)}$ is the multiplicity of the weight

$$(n_1,n_2,n_3) = (\nu_1-\nu_2, \nu_2, \mu_1+\mu_2-\nu_1)$$

in the u(3) irrep $\{\mu_1\mu_2 0\}$. The multiplicity M is easily obtained by using the well-known betweenness conditions of the associated Gel'fand pattern

$$\begin{vmatrix} \mu_1 & & \mu_2 & & 0 \\ & \nu_1-\vartheta & & \vartheta & \\ & & \nu_1-\nu_2 & & \end{matrix}\right\rangle \tag{3.2b}$$

and is given by the number of allowed values of ϑ in (3.2b). (For convenience, we reproduce the results of King and Qubanchi[15] for irreps of g_2 with $0 \leqslant \mu_1+\mu_2 \leqslant 6$ in Table I.)

According to Racah,[16] the number of internal labels required to specify the basis states of an irrep of a compact group is $\frac{1}{2}(l-r)$, where l is the order of the group (number of generators) and r its rank (number of commuting weight operator). In the case of interest to us, the number of internal labels is $6 = \frac{1}{2}(14-2)$. Since an su(3) basis provides us with five labels, one must therefore introduce an extra label (ϑ) to completely specify a $g_2 \supset$ su(3) basis corresponding to the branching rule (3.2a). It remains, however, to give an operational meaning to ϑ. As of now, the only satisfactory resolution to this missing label problem has been given by Perroud[2] with the introduction of a "path label."

TABLE I. Branching multiplicities $M^{[\mu_1\mu_2]}_{\{\nu_1\nu_2\}}$ of the representations $[\mu_1\mu_2]$ of g_2 into representations $\{\nu_1\nu_2\}$ of su(3) (King and Qubanchi[15]).

$\{\nu_1\nu_2\}$		11		22	32	33		43	44	53	54	55		64	65	66
$[\mu_1\mu_2]$	00	10	21	20	31	30	42	41	40	52	51	50	63	62	61	60
00	1															
10	1	1														
11		1	1													
20	1	1	1	1												
21		1	2	1	1											
30	1	1	1	1	1	1										
22			1	1	1		1									
31		1	2	1	2	1	1	1								
40	1	1	1	1	1	1	1	1	1							
32			1	1	2	1	2	1		1						
41		1	2	1	2	1	2	2	1	1	1					
50	1	1	1	1	1	1	1	1	1	1	1	1				
33					1	1	1	1		1			1			
42			1	1	2	1	3	2	1	2	1		1	1		
51		1	2	1	2	1	2	2	1	2	2	1	1	1	1	
60	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

It has been repeatedly demonstrated (see the review of Rowe et al.[10]) that vector VCS theory, when applicable, offers natural labeling schemes. Therefore since a VCS expansion exists[17] for the Lie algebra chain $g_2\supset su(3)$, it should provide an operational meaning for the label ϑ. However, the $g_2\supset su(3)$ Lie algebra chain has a raising operator algebra that is nilpotent of order 3, i.e,

$$\mathbf{n}_+ = \Delta_+/u(2) = \mathbf{n}^1_+ + \mathbf{n}^2_+ + \mathbf{n}^3_+ \tag{3.3a}$$

with

$$[\mathbf{n}_+,\mathbf{n}^i_+] \subset \sum_{j<i} \mathbf{n}^j_+, \tag{3.3b}$$

and such a case has not yet been covered in applications of VCS theory. We therefore postpone the uncovering of a meaning of ϑ to a subsequent publication.[17] It will nevertheless be shown in Sec. IV that VCS theory readily resolves the $g_2\supset so(4)$ basis construction since this Lie algebra chain has a nilpotent raising operator algebra n_+ of order 2, a case fully reviewed in Rowe *et al.*[10]

Because of the way we have embedded u(2) in su(3) [Eq. (2.12)], the su(3)$\downarrow$u(2) branching rule is given by

su(3)$\downarrow$u(2):

$$\{\nu_1\nu_2\}\downarrow \sum_{m_{12},m_{22}} (m_{12}-n_3, m_{22}-n_3) \equiv (m'_{12}m'_{22}), \tag{3.4a}$$

where m_{12} and m_{22} satisfy the usual betweenness conditions

$$\nu_1 \geqslant m_{12} \geqslant \nu_2, \quad \nu_2 \geqslant m_{22} \geqslant 0, \tag{3.4b}$$

and

$$n_3 = \nu_1 + \nu_2 - m_{12} - m_{22}. \tag{3.4c}$$

In Dirac notation, a $g_2\supset su(3)\supset u(2)$ basis will therefore be denoted by

$$|[\mu_1\mu_2]\vartheta;\ \{\nu_1,\nu_2\}(m'_{12}m'_{22})m'_{11}\rangle, \quad m'_{22} \leqslant m'_{11} \leqslant m'_{12}. \tag{3.5}$$

B. State labeling for the $g_2\supset so(4)\supset u(2)$ chain

To facilitate the construction of a $g_2\supset so(4)$ basis, it is convenient to now define a highest-weight state by

$$\begin{aligned} &u_\alpha|[\mu_1\mu_2]hw\rangle' = 0, \quad -\tfrac{3}{2} \leqslant \alpha \leqslant \tfrac{3}{2}, \\ &I_+|[\mu_1\mu_2]hw\rangle' = 0, \\ &S_+|[\mu_1\mu_2]hw\rangle' = 0. \end{aligned} \tag{3.6}$$

One sees that the positive roots corresponding to the raising operators in this equation are obtained from those of Eq. (2.1) by rotation through an angle of $-\pi/3$. Hence it becomes appropriate to define the representation labels μ_1 and μ_2 as the eigenvalues of the rotated weight operators

$$h_{2\alpha_1+3\alpha_2} = g_{11} - g_{33} = I_0 + S_0$$

and (3.7)

$$h_{-\alpha_1-3\alpha_2} = g_{33} - g_{22} = I_0 - S_0.$$

We conclude from (2.15)–(2.17), (3.6) and (3.7) that a g_2 unirrep $[\mu_1\mu_2]$ is characterized by the existence of subset of states with highest I angular momentum given by

$$I_{\max} = I_\mu = \tfrac{1}{2}(\mu_1 + \mu_2) \tag{3.8}$$

and with the corresponding S angular momentum given by

$$S = S_\mu = \tfrac{1}{2}(\mu_1 - \mu_2). \tag{3.9}$$

This subset will be referred to as the so(4) $[\mu_1, -\mu_2]$ [or $su_I(2)\oplus su_S(2)$ (I_μ,S_μ)] *maximal subrepresentation*. The existence of such a subrepresentation is of importance for the construction of a $g_2\supset so(4)$ basis. Note that we will use the labels I_μ and S_μ interchangeably with $[\mu_1\mu_2]$ in the following.

The branching rule

$$g_2\downarrow so(4):\ [\mu_1\mu_2]\downarrow[\lambda_1\lambda_2] \tag{3.10}$$

has been derived by Gaskell and Sharp.[6] Unfortunately, it has not be given in simple analytical form. It has been tabulated for low-dimensional representations by McKay and Patera[18] and some of their results are reproduced in Table II.

A basis for the so(4) irrep $[\lambda\ \lambda_2]$ is either labeled by the u(2) Gel'fand pattern

$$\left|\begin{matrix}\lambda_1-b & & \lambda_2-b \\ & m_{11}-b & \end{matrix}\right\rangle, \quad \begin{matrix} 0 \leqslant b \leqslant \lambda_1+\lambda_2, \\ \lambda_1 \leqslant m_{11} \leqslant \lambda_2, \end{matrix} \tag{3.11}$$

or by the angular momentum basis

$$\left|\begin{matrix} IM_I \\ SM_S \end{matrix}\right\rangle \tag{3.12a}$$

with I and S given by (2.18) and

$$M_S = \tfrac{1}{2}(\lambda_1+\lambda_2-2b), \quad M_I = \tfrac{1}{2}(2m_{11}-\lambda_1-\lambda_2). \tag{3.12b}$$

According to Racah,[16] one must complement the $g_2\downarrow so(4)$ branching rule with *two* supplementary labels in order to have a complete basis specification. In the following

TABLE II. Branching multiplicities $M_{I,S}^{(I_\mu,S_\mu)}$ (I,S) of the representations $[\mu_1\mu_2]$ of g_2 into representations (I,S) of $\mathrm{so}(4)\sim \mathrm{su}_I(2)\oplus \mathrm{su}_S(2)$ (McKay and Patera[18]).

$[\mu_1\mu_2]$	$M_{I,S}^{(I_\mu,S_\mu)}$ (I,S)
[10]	$(\frac{1}{2},\frac{1}{2}),(0,1)$
[11]	$(1,0),(\frac{1}{2},\frac{3}{2}),(0,1)$
[20]	$(1,1),(\frac{1}{2},\frac{3}{2}),(\frac{1}{2},\frac{1}{2}),(0,2),(0,0)$
[21]	$(\frac{3}{2},\frac{1}{2}),(1,2),(1,1),(\frac{1}{2},\frac{5}{2}),(\frac{1}{2},\frac{3}{2}),(\frac{1}{2},\frac{1}{2}),(0,2),(0,1)$
[30]	$(\frac{3}{2},\frac{3}{2}),(1,2),(1,1),(1,0),(\frac{1}{2},\frac{5}{2})(\frac{1}{2},\frac{3}{2}),(\frac{1}{2},\frac{1}{2}),(0,3),(0,1)$
[22]	$(2,0),(\frac{3}{2},\frac{3}{2}),(1,3),(1,1),(\frac{1}{2},\frac{5}{2}),(\frac{1}{2},\frac{3}{2}),(0,2)(0,0)$
[31]	$(2,1),(\frac{3}{2},\frac{5}{2}),(\frac{3}{2},\frac{3}{2}),(\frac{3}{2},\frac{1}{2}),(1,3),2(1,2),(1,1),(1,0),$ $(\frac{1}{2},\frac{7}{2}),(\frac{1}{2},\frac{5}{2}),2(\frac{1}{2},\frac{3}{2}),(\frac{1}{2},\frac{1}{2}),(0,3),(0,2),(0,1)$
[40]	$(2,2),(\frac{3}{2},\frac{5}{2}),(\frac{3}{2},\frac{3}{2}),(\frac{3}{2},\frac{1}{2}),(1,3),(1,2),2(1,1),$ $(\frac{1}{2},\frac{7}{2}),(\frac{1}{2},\frac{5}{2}),(\frac{1}{2},\frac{3}{2}),(\frac{1}{2},\frac{1}{2}),(0,4),(0,2),(0,0)$
[32]	$(\frac{5}{2},\frac{1}{2}),(2,2),(2,1),(\frac{3}{2},\frac{7}{2}),(\frac{3}{2},\frac{5}{2}),(\frac{3}{2},\frac{3}{2}),(\frac{3}{2},\frac{1}{2}),(1,4),(1,3),$ $2(1,2),(1,1),(\frac{1}{2},\frac{7}{2}),2(\frac{1}{2},\frac{5}{2}),(\frac{1}{2},\frac{3}{2}),(\frac{1}{2},\frac{1}{2}),(0,3),(0,2),(0,1)$
[41]	$(\frac{5}{2},\frac{3}{2}),(2,3),(2,2),(2,1),(2,0),(\frac{3}{2},\frac{7}{2}),2(\frac{3}{2},\frac{5}{2}),2(\frac{3}{2},\frac{3}{2}),(\frac{3}{2},\frac{1}{2}),$ $(1,4),2(1,3),2(1,2),2(1,1),(1,0),(\frac{1}{2},\frac{9}{2}),(\frac{1}{2},\frac{7}{2}),2(\frac{1}{2},\frac{5}{2}),2(\frac{1}{2},\frac{3}{2}),$ $(\frac{1}{2},\frac{1}{2}),(0,4),(0,3),(0,2),(0,1)$
[50]	$(\frac{5}{2},\frac{5}{2}),(2,3),(2,2),(2,1),(\frac{3}{2},\frac{7}{2}),(\frac{3}{2},\frac{5}{2}),2(\frac{3}{2},\frac{3}{2}),(\frac{3}{2},\frac{1}{2}),$ $(1,4),(1,3),2(1,2),(1,1),(1,0),(\frac{1}{2},\frac{9}{2}),(\frac{1}{2},\frac{7}{2}),(\frac{1}{2},\frac{5}{2}),2(\frac{1}{2},\frac{3}{2}),$ $(\frac{1}{2},\frac{1}{2}),(0,5),(0,3),(0,1)$

section, we will give a "canonical" resolution for these two missing labels (which we will denote τ and J) and provide an explicit construction of $g_2\supset \mathrm{su}_I(2)\oplus \mathrm{su}_S(2)$ basis of vector holomorphic functions. In Dirac notation, a $g_2\supset \mathrm{su}_I(2)\oplus \mathrm{su}_S(2)$ basis state will then be labeled

$$\left|\begin{matrix} I_\mu & & IM_I \\ S_\mu & \tau J & SM_S\end{matrix}\right\rangle . \tag{3.13}$$

IV. MATRIC REPRESENTATION OF g_2 IN AN so(4) BASIS

A. Vector coherent state theory

Vector coherent state (VCS) theory[8–11] and the associated K-matrix techniques for computing inner products[9,10] greatly simplify the construction of orthonormal bases for the ladder representations of a Lie algebra. In VCS theory, the familiar Cartan subalgebra $\mathbf{k}$ is extended to a larger non-Abelian stability algebra $\mathbf{h}$ that contains $\mathbf{k}$ as an Abelian subalgebra. The extension allows one to relate nontrivial sets of lowering (likewise raising) operators as components of irreducible tensors under the stability subalgebra. (For an Abelian stability algebra, the irreducible tensors are necessarily trivial; i.e., one dimensional.) As a result, one is able to tensorially couple polynomials in the lowering (raising) operators to highest- (lowest-) weight subrepresentations of the stability subalgebra $\mathbf{h}$ of a Lie algebra $\mathbf{g}$ and, thereby, construct an orthonormal basis for the irrep of $\mathbf{g}$ that reduces $\mathbf{h}$.

The construction is a refinement of the well-known Cartan techniques of building up basis states of ladder representations by the repeated actions of lowering (raising) operators on highest- (lowest-) states. The difference is that VCS theory exploits the strength of the Wigner–Echart theorem by making use of the existence of the tensor (Wigner–Racah) calculus for the stability subalgebra. A valuable feature of the construction is that the basis states automatically reflect the corresponding $\mathbf{g}\downarrow\mathbf{h}$ branching rule. Furthermore, a set of missing labels, having a meaningful group theoretical interpretation, is naturally introduced in the process.

The construction was originally introduced and applied to situations in which the algebras of raising operators are Abelian. Examples of such situations are the $\mathrm{u}(n+1)$, $\mathrm{sp}(2n,\mathbb{R})$, and $\mathrm{so}(2n)$ albegras with $\mathrm{u}(n)$ a stability subalgebra (see Rowe *et al.*[10] for references). In a recent development of K-matrix theory, Rowe *et al.*[10] demonstrated that the VCS construction can also be applied to situations with raising operator algebras that are nilpotent of order 2 [cf. Eq. (3.3)]. The generalized theory was used by them to compute the matrices for the $\mathrm{so}(2n+1)$ representations in an $\mathrm{so}(2n+1)\supset\mathrm{so}(2n)\supset\mathrm{u}(n)$ basis. We shall show that this generalization also allows us to construct an orthonormal basis and determine the explicit matrices for the ladder representations of g_2 in a $g_2\supset\mathrm{so}(4)\supset\mathrm{u}(2)$ basis.

B. Application to the $g_2\supset$so(4)$\supset$u(2) chain

The Cartan subalgebra $\mathrm{u}(1)\oplus\mathrm{u}(1)$ for g_2 is identified with the subalgebra $\mathrm{u}_I(1)\oplus\mathrm{u}_S(1)$ having generators I_0 and S_0 in the basis of Sec. II C. The highest-weight state for a g_2 irrep $[\mu_1\mu_2]$ is then the state with maximal (M_I,M_S), and M_I and M_S are, respectively, the eigenvalues of I_0 and S_0. For the highest-weight state, M_I and M_S take the values given by Eqs. (3.8) and (3.9); i.e.,

$$M_I=I_\mu=\tfrac{1}{2}(\mu_1+\mu_2),\quad M_S=S_\mu=\tfrac{1}{2}(\mu_1-\mu_2).$$

The highest-weight state for the g_2 irrep $[\mu_1\mu_2]$ is clearly also a highest-weight state for a highest-weight subrepresentation of the $\mathrm{u}(2)\sim\mathrm{u}_I(1)\oplus\mathrm{su}_S(2)$ subalgebra. Conversely, $\mathrm{u}(2)$ is the stability algebra for the subspace of states carrying its highest-weight representation. We therefore adopt $\mathrm{u}(2)$ as the stability algebra for the VCS construction.

The $\mathrm{u}(2)$ stability algebra contains the $\mathrm{su}_S(2)$ and lowering operators $S_\pm$ as well as the Cartan subalgebra $\mathrm{u}_I(1)\oplus\mathrm{u}_S(1)$. With respect to $\mathrm{u}(2)$ as stability algebra, the remaining generators of g_2 are the four components (u_α; $\alpha=\pm\frac{1}{2},\pm\frac{3}{2}$) of an ($M_I=\frac{1}{2}$, $S=\frac{3}{2}$) raising tensor, four components (v_α; $\alpha=\pm\frac{1}{2},\pm\frac{3}{2}$) of an ($M_I=-\frac{1}{2}$, $S=\frac{3}{2}$) lowering tensor, and the components $I_\pm$ of $M_I=\pm 1$, $S=0$) of the angular momentum algebra $\mathbf{I}$, respectively.

A set of raising operators of $g_2/\mathrm{u}(2)$ is therefore given by

$$n_+=\{I_+,u_\alpha;\ -\tfrac{3}{2}\leqslant\alpha\leqslant\tfrac{3}{2}\}. \tag{4.1}$$

According to the commutation relations (2.22), $\mathbf{n}_+$ is a non-Abelian subalgebra of g_2^c. In particular, we have the nonvanishing commutator

$$[u_\alpha,u_\beta]=\delta_{\alpha,-\beta}(-1)^{3/2-\beta}I_+$$

and we determine that $\mathbf{n}_+$ is nilpotent of order 2.

Let $\{|\eta\rangle\}$ be an arbitrary orthonormal basis for a $\mathrm{u}(2)$ highest-weight subspace of a g_2 irrep. The VCS representation of a state $|\psi\rangle$ is then defined by

$$(w,z|\psi\rangle = \sum_\eta |\eta\rangle\langle\eta|e^{Z(w,z)}|\psi\rangle,$$

where

$$Z(w,z) = wI_+ + z\cdot u, \quad z\cdot u = \sum_\alpha (-1)^{3/2-\alpha} z_{-\alpha} u_\alpha, \tag{4.2}$$

and w and z_α are a set of five complex (Bargman) variables.

The vector coherent state representation $\Gamma(X)$ of an arbitrary generator $X \in g_2$ is defined by

$$\begin{aligned}\Gamma(X)(w,z|\psi\rangle &= \sum_\eta |\eta\rangle\langle\eta|\exp Z(w,z)X|\psi\rangle \\ &= |\eta\rangle\langle\eta|(X + [Z,X] + \tfrac{1}{2}[Z,[Z,X]] \\ &\quad + \cdots)\exp Z(w,z)|\psi\rangle.\end{aligned} \tag{4.3}$$

With

$$\nabla = \frac{\partial}{\partial w}, \quad \partial_\alpha = \frac{\partial}{\partial z_\alpha}, \tag{4.4}$$

and the use of the identity

$$\begin{aligned}&\partial_{\tilde\alpha} \exp(z\cdot u) = (u_\alpha - \tfrac{1}{2}z_\alpha I_+)\exp(z\cdot u), \\ &\partial_{\tilde\alpha} = (-1)^{3/2-\alpha}\partial_{-\alpha},\end{aligned} \tag{4.5}$$

we find the following coherent state expansion for $g_2 \supset so(4)$:

$$\Gamma(I_+) = \nabla, \tag{4.6a}$$

$$\Gamma(u_\alpha) = \partial_{\tilde\alpha} - \tfrac{1}{2}z_\alpha\nabla, \tag{4.6b}$$

$$\Gamma(S_m) = \vec{\mathbb{S}}_m + \mathscr{S}_m, \quad \mathscr{S}_m = -\sqrt{5}[z\times\partial]^1_m, \tag{4.6c}$$

$$\Gamma(I_0) = \mathbb{I}_0 - \tfrac{1}{2}z_\alpha\,\partial_\alpha - w\nabla, \tag{4.6d}$$

$$\begin{aligned}\Gamma(v_\alpha) &= w\,\partial_{\tilde\alpha} + z_\alpha(\mathbb{I}_0 - w\nabla - \tfrac{1}{4}z_\beta\nabla_\beta) \\ &\quad - (\tfrac{5}{3})^{1/2}[z\times S]^{3/2}_\alpha + (5/2\sqrt{3})\,[z\times[z\times\partial]^1]^{3/2}_\alpha \\ &\quad - (5/12\sqrt{3})\,[z\times[z\times z]^1]^{3/2}_\alpha\nabla,\end{aligned} \tag{4.6e}$$

$$\begin{aligned}\Gamma(I_-) &= 2w\mathbb{I}_0 - w^2\nabla - wz_\alpha\,\partial_\alpha \\ &\quad - (\tfrac{5}{3})^{1/2}[[z\times z]^1\times\mathscr{S}]^0 \\ &\quad + (5/3\sqrt{3})[[z\times z]^1\times[z\times\partial]^1]^0 \\ &\quad + (5/12\sqrt{3})[[z\times z]^1\times[z\times z]^1]^0\nabla,\end{aligned} \tag{4.6f}$$

where $\mathbb{I}_0$ and $\vec{\mathbb{S}}$ are intrinsic operators acting only on a basis of states carrying an *intrinsic* (highest-weight) representation $\{I_\mu\}\times(S_\mu)$ of the stability algebra $u_I(1)\oplus su_S(2)$ and the various square brackets stand for coupled operators, e.g.,

$$[z\times z]^j_m = \sum_{\alpha\beta}\langle\tfrac{3}{2}\alpha;\,\tfrac{3}{2}\beta\,|\,jm\rangle z_\alpha z_\beta. \tag{4.7}$$

From its definition, it is clear that a coherent state wave function $(w,z|\psi\rangle$ is a vector-valued holomorphic function of the variables (w,z_α). The inner product for the VCS Hilbert space of holomorphic functions has been given in integral form by Rowe *et al.*[11] For practical purposes, however, it is generally much simpler and more useful to determine the inner product, and hence calculate Lie algebra matrix elements in an orthonormal Bargmann basis, by the K-matrix technique.[9,10]

C. Construction of an orthonormal $g_2 \supset$ so(4) $\supset$ u(2) basis

We start by defining a basis of states that is orthonormal with respect to a Bargmann inner product. The transformation K that maps this basis onto a basis that is orthonormal with respect to the VCS inner product is then determined by K-matrix theory and used to map the VCS representation Γ of g_2 to an equivalent representation $\gamma = K^{-1}\,\Gamma K$ that is unitary with respect to the Bargmann inner product. Matrix elements for this unirrep are then easily evaluated.

By Proposition 2 of Rowe *et al.*[10] we known that the (z_α) variables transform under u(2) as components of an $I_0 = \frac{1}{2}$, $S = \frac{3}{2}$ tensor. We can therefore construct a set of orthogonal polynominals $\{P^l_{\tau J}(z)\}$, where $P^l_{\tau J}(z)$ is a polynomial of degree l in the (z_α), of $su_S(2)$ spin J, and where τ is a multiplicity index to distinguish distinct polynomials having the same values of l and J. We require the polynomials to be orthonormal with respect to the Bargmann measure so that

$$[P^l_{\tau J}(\partial)(P^{l'}_{\tau' J'}(z))]_{z=0} = \delta_{ll'}\delta_{JJ'}\delta_{\tau\tau'}. \tag{4.8}$$

Since the $l=1$ polynomials carry the fundamental four-dimensional spinor representation[10] of a suitably defined so(5) Lie algebra (in the notation of McKay and Patera[18]), it follows that the polynomials of degree l carry a symmetric spinor irrep $[l\,0]$ of so(5). The possible values of J and their multiplicity for a polynomial of degree l are therefore given by the branching of the so(5) representation $[l\,0]$ on restriction to su(2),

$$\mathrm{so}(5)\downarrow\mathrm{su}(2){:}[l\,0]\downarrow\sum_J M^l_J(J). \tag{4.9}$$

For convenience, we reproduce part of the tabulation of McKay and Patera[18] in Table III. We also relate the multiplicity of su(2) irreps to the concept of permissible diagrams[6] in the Appendix.

Given a set of orthogonal polynomials, we can use them to construct the orthonormal Bargmann basis of holomorphic vector-valued wave functions

TABLE III. Branching multiplicities $M^l_J(J)$ of the symmetric spinor representations $(l0)$ of so(5) into representations (J) of su(2) (McKay and Patera[18]).

l	$M^l_J(J)(J)$
1	$(\frac{3}{2})$
2	(3,),(1)
3	$(\frac{9}{2}),(\frac{5}{2}),(\frac{3}{2})$
4	(6),(4),(3),(2),(0)
5	$(\frac{15}{2}),(\frac{11}{2}),(\frac{9}{2}),(\frac{7}{2}),(\frac{5}{2}),(\frac{3}{2})$
6	(9),(7),(6),(5),(4),2(3),(1)
7	$(\frac{21}{2}),(\frac{17}{2}),(\frac{15}{2}),(\frac{13}{2}),(\frac{11}{2}),2(\frac{9}{2}),(\frac{7}{2}),(\frac{5}{2}),(\frac{3}{2})$
8	(12),(10),(9),(8),(7),2(6),(5),2(4),(3),(2),(0)
9	$(\frac{27}{2}),(\frac{23}{2}),(\frac{21}{2}),(\frac{19}{2}),(\frac{17}{2}),2(\frac{15}{2}),(\frac{13}{2}),2(\frac{11}{2}),2(\frac{9}{2}),(\frac{7}{2}),(\frac{5}{2}),(\frac{3}{2})$
10	(15),(13),(12),(11),(10),2(9),(8),2(7),2(6),2(5),(4),2(3),(1)

$$\left(w,z\left|\begin{matrix} I_\mu & & IM_I \\ S_\mu & \tau J & SM_S\end{matrix}\right.\right)$$
$$= \frac{w^{I-M_I}}{\sqrt{(I-M_I)!}}\left[P^l_{\tau J}(z)\times\left|\begin{matrix} I_\mu \\ S_\mu\end{matrix}\right\rangle\right]^S_{M_S}, \tag{4.10a}$$

with l taking the value

$$l = 2(I_\mu - I). \tag{4.10b}$$

We note that, with l fixed in this way, this basis is labeled by a total of six internal labels in correspondence with the number required to label a basis for the chain $g_2 \supset \mathrm{so}(4)$.

Now, a transformation K can be defined that maps the Bargmann basis onto a basis of states that are orthonormal with respect to the VCS inner product

$$\left(w,z\left|\begin{matrix} I_\mu & & IM_I \\ S_\mu & \tau J & SM_S\end{matrix}\right.\right)_{\mathrm{VCS}}$$
$$= \left(w,z\left|K\right|\begin{matrix} I_\mu & & IM_I \\ S_\mu & \tau J & SM_S\end{matrix}\right)$$

conversely, the transformation K defines a representation

$$\gamma(X) = K^{-1}\Gamma(X)K, \quad X\in g_2,$$

which is equivalent to the VCS representation Γ but which, unlike the VCS representation, is unitary with respect to the Bargmann inner product.

As shown by Rowe *et al.*[10] the K transformation commutes with the stability subalgebra. Thus K is diagonal in M_I and S independent of M_S. Furthermore, it was shown that, since the matrix elements of the

$$\mathrm{so}(4) \sim \mathrm{su}_I(2) \oplus \mathrm{su}_S(2)$$

subalgebra are already well known, one needs only the submatrices,

$$\mathscr{K}(IS)_{\tau_i J_i, \tau_j J_j} \equiv \left\langle\begin{matrix} I_\mu & & I,I \\ S_\mu & \tau_i J_i & S\end{matrix}\right|K\left|\begin{matrix} I_\mu & & I,I \\ S_\mu & \tau_j J_j & S\end{matrix}\right\rangle \tag{4.11}$$

to be able to infer all the matrix elements of the g_2 algebra.

The following linear recursion relation for $\mathscr{K}(IS)^2$, obtained using Hermiticity arguments [cf. Eq. (2.20)] and the VCS expansion Γ [i.e. (1.6)], is given by the adaptation to $g_2 \supset \mathrm{so}(4)$ of the general relationship given by Rowe *et al.*[10] (see also the Appendix of Ref. 17):

$$\begin{aligned}\mathscr{K}(IS)^2_{\tau_i J_i, \tau_j J_j} = {} & \frac{1}{2(I_\mu - I)}\sum_{\substack{\tau_k J_k S_k \\ \tau_l J_l}}\left\langle\begin{matrix} I_\mu & & I,I \\ S_\mu & \tau_j J_j & S\end{matrix}\right\|z\left\|\begin{matrix} I_\mu & & I+\frac12, I+\frac12 \\ S_\mu & \tau_k J_k & S_k\end{matrix}\right\rangle \\ & \times\left\{\left[\Omega(IJ_iS) - \Omega\left(I+\frac{1}{2}, J_l S_k\right)\right]\times\left\langle\begin{matrix} I_\mu & & I,I \\ S_\mu & \tau_i J_i & S\end{matrix}\right\|z\left\|\begin{matrix} I_\mu & & I+\frac12, I+\frac12 \\ S_\mu & \tau_k J_k & S_k\end{matrix}\right\rangle\right. \\ & + \frac{1}{(2I+2)}\left(\frac{5}{3}\right)^{1/2}\sum_{\tau_m J_m}\left\langle\begin{matrix} I_\mu & & I+1, I+1 \\ S_\mu & \tau_m J_m & S\end{matrix}\right\|\partial\left\|\begin{matrix} I_\mu & & I+\frac12, I+\frac12 \\ S_\mu & \tau_l J_l & S_k\end{matrix}\right\rangle \\ & \left.\times\left\langle\begin{matrix} I_\mu & & I,I \\ S_\mu & \tau_i J_i & S\end{matrix}\right\|[[z\times z]^1\times S]^0 - \tfrac{3}{2}[[z\times z]^1\times\mathscr{S}]^0\left\|\begin{matrix} I_\mu & & I+1, I+1 \\ S_\mu & \tau_m J_m & S\end{matrix}\right\rangle\right\} \\ & \times\mathscr{K}(I+\tfrac12, S_k)^2_{\tau_l J_l, \tau_k J_k},\end{aligned} \tag{4.12}$$

where

$$\begin{aligned}\Omega(IJS) = {} & ((I_\mu + \tfrac34) - \tfrac14(I_\mu - I))2(I_\mu - I) \\ & - \tfrac13 S(S+1) + \tfrac16 J(J+1)\end{aligned} \tag{4.13}$$

and where, e.g.,

$$\left\langle\begin{matrix} I_\mu & & I,I \\ S_\mu & \tau_j J_j & S\end{matrix}\right\|z\left\|\begin{matrix} I_\mu & & I+\frac12, I+\frac12 \\ S_\mu & \tau_k J_k & S_k\end{matrix}\right\rangle$$

is an $\mathrm{su}_S(2)$-reduced matrix element. The latter is easily evaluated using standard su(2) recoupling techniques

$$\left\langle\begin{matrix} I_\mu & & I,I \\ S_\mu & \tau_j J_j & S\end{matrix}\right\|z\left\|\begin{matrix} I_\mu & & I+\frac12, I+\frac12 \\ S_\mu & \tau_k J_k & S_k\end{matrix}\right\rangle$$
$$= U(S_\mu, J_k, S, \tfrac32; S_k, J_j)\times\left(\begin{matrix} l \\ \tau_j J_j\end{matrix}\right\|z\left\|\begin{matrix} l-1 \\ \tau_k J_k\end{matrix}\right), \tag{4.14}$$

where the last term is an $\mathrm{su}_S(2)$-reduced matrix element of z between the Bargmann basis polynomials

$$\left(z\left|\begin{matrix} l \\ \tau J M_J\end{matrix}\right.\right) \equiv P^l_{\tau J M_J}(z). \tag{4.15}$$

As an example, consider the g_2 irrep [21] corresponding to $(I_\mu, S_\mu) = (\frac32, \frac12)$. By Table II, one determines that this irrep decomposes into the sum of irreps

$$(\tfrac32,\tfrac12),\ (1,2),\ (1,1),\ (\tfrac12,\tfrac52),\ (\tfrac12,\tfrac32),\ (\tfrac12,\tfrac12),\ (0,2),\ (0,1)$$

under $\mathrm{so}(4) \sim \mathrm{su}_I(2) \oplus \mathrm{su}_S(2)$. From Table III, we determine that we have the Bargmann basis states

l	(J)	(I,S)
0	(0)	$(\frac32,\frac12)$
1	$(\frac32)$	(1,2),(1,1)
2	(3),(1)	$(\frac12,\frac72),(\frac12,\frac52),(\frac12,\frac32),(\frac12,\frac12)$
3	$(\frac92),(\frac52),(\frac32)$	(0,5),(0,4),(0,3),2(0,2),(0,1).

Observe that this Bargmann space is larger than the required VCS irrep space, which means that the K operator must annihilate some Bargmann states.

Solving the recursion relation (4.12), we readily compute the following values for the one-dimensional $\mathscr{K}^2$ matrices:

$$\mathscr{K}(\tfrac{3}{2},\tfrac{1}{2})^2 = 1, \quad \mathscr{K}(1,2)^2 = 1, \quad \mathscr{K}(1,1)^2 = \tfrac{7}{3},$$
$$\mathscr{K}(\tfrac{1}{2},\tfrac{7}{2})^2 = 0, \quad \mathscr{K}(\tfrac{1}{2},\tfrac{5}{2})^2 = \tfrac{7}{3},$$
$$\mathscr{K}(\tfrac{1}{2},\tfrac{3}{2})^2 = \tfrac{7}{3}, \quad \mathscr{K}(\tfrac{1}{2},\tfrac{1}{2})^2 = \tfrac{35}{9},$$
$$\mathscr{K}(0,5)^2 = 0, \quad \mathscr{K}(0,4)^2 = 0, \quad \mathscr{K}(0,3)^2 = 0.$$

For the single two-dimensional $\mathscr{K}(0,2)^2$ matrix, we find

$$\mathscr{K}(0,2)^2 = \begin{pmatrix} \dfrac{2^3\cdot 7}{3\cdot 5} & \dfrac{-2.7\sqrt{3\cdot 7}}{3^2\cdot 5} \\[2ex] \dfrac{-2\cdot 7\sqrt{3\cdot 7}}{3^2\cdot 5} & \dfrac{7^2}{2\cdot 3^2\cdot 5} \end{pmatrix},$$

which is observed to have one zero eigenvalue. It follows that the K operators maps both Bargmann (0,2) states onto a single VCS state and annihilates the Bargmann states $(\frac{1}{2},\frac{7}{2})$, (0,5), (0,4), and (0,3). Thus the K operator is seen to define a "physical subspace" of Bargmann space in which the extraneous states in the null space of the K operator are excluded.

D. Calculation of matrix elements

Having determined the K matrices, the matrix elements of the g_2 generators may be calculated in the unirrep γ. Since matrix elements of the so(4) $\sim \mathrm{su}_I(2) \oplus \mathrm{su}_S(2)$ Lie algebra are already known, it remains to compute the reduced matrix elements for the so(4) tensor t of Eq. (2.19).

We easily determine the particular $M_I = I$ $\mathrm{su}_S(2)$-reduced matrix elements

$$\begin{aligned}
&\left\langle \begin{matrix} I_\mu & & I-\frac{1}{2}, I-\frac{1}{2} \\ S_\mu & \tau'J' & S' \end{matrix} \right\| \gamma(v) \left\| \begin{matrix} I_\mu & & I,I \\ S_\mu & \tau J & S \end{matrix} \right\rangle \\
&\quad = \left\langle \begin{matrix} I_\mu & & I-\frac{1}{2}, I-\frac{1}{2} \\ S_\mu & \tau'J' & S' \end{matrix} \right\| KzK^{-1} \left\| \begin{matrix} I_\mu & & I,I \\ S_\mu & \tau J & S \end{matrix} \right\rangle \\
&\quad = \sum_{\substack{\tau_k J_k \\ \tau_l J_l}} \mathscr{K}\left(I-\frac{1}{2}, S'\right)_{\tau'J',\tau_l J_l} \\
&\qquad \times \left\langle \begin{matrix} I_\mu & & I-\frac{1}{2}, I-\frac{1}{2} \\ S_\mu & \tau_l J_l & S' \end{matrix} \right\| z \left\| \begin{matrix} I_\mu & & I,I \\ S_\mu & \tau_k J_k & S \end{matrix} \right\rangle \\
&\qquad \times \mathscr{K}(I,S)^{-1}_{\tau_k J_k, \tau J}.
\end{aligned} \tag{4.16}$$

Hence we determine the $\mathrm{su}_I(2) \oplus \mathrm{su}_S(2)$- (triple bar-) reduced elements for t

$$\begin{aligned}
&\left\langle \begin{matrix} I_\mu & & I-\frac{1}{2} \\ S_\mu & \tau'J' & S' \end{matrix} \right|\!\left\| \gamma(t) \right\|\!\left| \begin{matrix} I_\mu & & I \\ S_\mu & \tau J & S \end{matrix} \right\rangle \\
&\quad = \left\langle \begin{matrix} I_\mu & & I-\frac{1}{2}, I-\frac{1}{2} \\ S_\mu & \tau'J' & S' \end{matrix} \right\| \gamma(v) \left\| \begin{matrix} I_\mu & & I,I \\ S_\mu & \tau J & S \end{matrix} \right\rangle \\
&\qquad \times \left(\langle I,I;\tfrac{1}{2},-\tfrac{1}{2} | I-\tfrac{1}{2}, I-\tfrac{1}{2} \rangle \right)^{-1}
\end{aligned} \tag{4.17}$$

from which all other $\mathrm{su}(2)_I \oplus \mathrm{su}(2)_S$ can be obtained through Hermitian conjugation,

$$\begin{aligned}
&\left\langle \begin{matrix} I_\mu & & I' \\ S_\mu & \tau'J' & S' \end{matrix} \right|\!\left\| \gamma(t) \right\|\!\left| \begin{matrix} I_\mu & & I \\ S_\mu & \tau J & S \end{matrix} \right\rangle \\
&\quad = (-1)^{I+\frac{1}{2}-I'}(-1)^{S+3/2-S'} \\
&\qquad \times \left[\left(\frac{2I+1}{2I'+1} \right) \left(\frac{2S+1}{2S'+1} \right) \right]^{1/2} \\
&\qquad \times \left\langle \begin{matrix} I_\mu & & I \\ S_\mu & \tau J & S \end{matrix} \right|\!\left\| \gamma(t) \right\|\!\left| \begin{matrix} I_\mu & & I' \\ S_\mu & \tau'J' & S' \end{matrix} \right\rangle.
\end{aligned} \tag{4.18}$$

V. DISCUSSION

A major object for giving an explicit construction of the representations of g_2 in a $g_2 \supset \mathrm{so}(4) \supset \mathrm{u}(2)$ basis was to obtain a simple solution to the problem of missing labels for this chain. As we have shown, VCS theory automatically provides both a complete set of labels and a group theoretical interpretation of their significance. The construction of basis states by tensorially coupling polynomials in tensor lowering operators to a highest-weight irrep gives rise to three distinct kinds of missing labels.

(1) Labels like J that define the tensor rank of a polynomial under the stability algebra. The polynomial tensors arise in the reduction of multiple products of elementary lowering operator tensors. For g_2, the elementary lowering operators (v_α) comprise a $(j=\frac{3}{2})$ tensor under $\mathrm{su}_S(2)$. The $\mathrm{su}_S(2)$ irreps (J) of interest are therefore those that arise in the reduction of the symmetric su(2) plethysms

$$(\tfrac{3}{2})\circ\{1\} = \tfrac{3}{2}, \quad (\tfrac{3}{2})\circ\{2\} = (3) + (1),$$
$$(\tfrac{3}{2})\circ\{3\} = (\tfrac{9}{2}) + (\tfrac{5}{2}) + (\tfrac{3}{2}), \text{ etc.},$$

cf. Table III. These plethysms can also be expressed in terms of branching rules. Thus, for g_2, the plethysms are related to the so(5)$\downarrow$su(2) branching

$$\mathrm{so}(5)\downarrow\mathrm{su}(2){:}\ \ [l0]\downarrow(\tfrac{3}{2})\circ\{l\}.$$

The series of irreps that occur in such plethysms are often well known in character theory (King[19]).

(2) Labels like τ that distinguish multiply occurring irreps in the plethysm decomposition. By relating the plethysm to a branching rule, it is clear that these labels are identified with the missing labels of some other algebraic chain reduction.

(3) Labels that arise from multiplicity in the coupling of the polynomial tensors to a highest-weight irrep. [These are not needed for g_2 because the su(2) Kronecker products are multiplicity-free.] To date, all applications of VCS theory have chosen a stability algebra of the $\mathrm{u}(n)$ type and we note that a canonical resolution of the $\mathrm{u}(n)$ outer product has been given by Biedenharn and collaborators in terms of operator patterns.[20] Thus all the missing labels have a meaningful group theoretical significance.

Elsewhere,[10] we have remarked on the wide range of problems in Lie algebra representation theory and its associated Wigner–Racah calculus that can be tackled with VCS theory. The application to $\mathrm{so}(2n+1) \supset \mathrm{so}(2n) \supset \mathrm{u}(n)$ given there and the present application to $g_2 \supset \mathrm{so}(4) \supset \mathrm{u}(2)$ adds a further dimension to the versatility of the theory. In a following publication,[17] we shall investigate its application with stability subalgebras for which the complementary lowering and raising operator algebras are nilpotent of order 3 or more. This will enable us to treat, for example, g_2 in an su(3) basis.

APPENDIX: THE POLYNOMIALS $P^{l}_{\tau JM_J}(z)$

In this appendix, we show how to construct the Bargmann polynomials $P^{l}_{\tau JM_J}(z)$ introduced in Sec. IV C. We restrict the presentation to its main features and thus refer the reader to Sharp and Lam,[21] Gaskell *et al.*[6] and especially Dumitrescu[14] for details

A *nonorthonormal* basis (lowercase p)

$$p^{l}_{\tau JM_J} = J^{(z)} p^{l}_{\tau JM_J = J}(z)$$

for the $M_J = J$ components of the polynomials is given by the stretched coupling of the $M_J = J$ component of the four elementary polynomials

$$\begin{aligned} &p^{1}_{0(3/2)\alpha}(z) = z_\alpha, \quad p^{2}_{01\alpha}(z) = [z\times z]^{1}_{\alpha}, \\ &p^{3}_{0(3/2)\alpha}(z) = [z\times[z\times z]^{1}]^{3/2}_{\alpha}, \qquad \text{(A1)} \\ &p^{4}_{100}(z) = [[z\times z]^{1}\times[z\times z]^{1}]^{0}_{0}, \end{aligned}$$

which compose an integrity basis (Gaskell *et al.*[6]). More precisely,

$$p^{l}_{\tau JJ} = [p^{1}_{0(3/2)(3/2)}]^{a}[p^{2}_{011}]^{b}[p^{3}_{0(3/2)(3/2)}]^{\epsilon}[p^{4}_{100}]^{\tau}, \qquad \text{(A2a)}$$

with

$$\begin{aligned} &\epsilon = 0, \quad (3l-2J)/2 \text{ even}, \\ &\epsilon = 1, \quad (3l-2J)/2 \text{ odd}, \\ &0 \leqslant 4\tau \leqslant l - 3\epsilon, \qquad \text{(A2b)} \\ &2a = (2J - l + 4\tau) \geqslant 0, \\ &2b = (\tfrac{3}{2}l - J - 3\epsilon - 6\tau) \geqslant 0. \end{aligned}$$

Thus for given l, J, and consequently ϵ, the nonorthonormal basis is parametrized by τ and M^{l}_{J} [Eq. (4.9)] is equal to the number of allowed values of τ in (A2b).

The basis (A2a) is nonorthonormal with respect to the Bargmann scalar product

$$[p^{l}_{\tau JJ}(\partial)(p^{l'}_{\tau' J'J'}(z))]_{z=0}.$$

Defining the overlap matrix

$$O^{2}_{\tau\tau'} = [p^{l}_{\tau JJ}(\partial)(p^{l'}_{\tau' J'J'}(z))]_{z=0},$$

we can define an orthonormal basis[22]

$$P^{l}_{\tau JJ}(z) = \sum_{\tau'} O^{-1}_{\tau\tau'} p^{l}_{\tau' JJ}(z),$$

which has the advantage of retaining the usefulness (if not the direct meaning) of the τ labeling scheme introduced in (A3).

[1] J. Becker, V. Hussin, and P. Winternitz, J. Math. Phys. **27**, 2217 (1986).
[2] M. Perroud, J. Math. Phys. **17**, 1998 (1976).
[3] J. Patera, J. Math. Phys. **11**, 3027 (1970).
[4] J. W. B. Hughes and J. Van der Jeugt, J. Math. Phys. **26**, 894 (1985).
[5] M. Günaydin and F. Gürsey, J. Math. Phys. **14**, 1651 (1973).
[6] R. Gaskell and R. T. Sharp, J. Math. Phys. **22**, 2736 (1981); R. Gaskell, A. Peccia, and R. T. Sharp, *ibid* **19**, 727 (1978).
[7] J. W. B. Hughes, in *Group Theoretical Methods in Physics, Lecture Notes in Physics*, Vol. 180, edited by M. Seradoğlu and E. Inönü (Springer, Berlin, 1983), p. 38.
[8] J. Deenen and C. Quesne, J. Math. Phys. **25**, 2354 (1984); **26**, 2705 (1985); C. Quesne, *ibid.* **27**, 428, 869 (1986).
[9] D. J. Rowe, J. Math. Phys. **25**, 2662 (1984); "Coherent states, contractions and classical limits of the non-compact symplectic groups," in *Proceedings of the XIII International Colloquium on Group Theoretical Methods in Physics*, College Park, Maryland, edited by W. W. Zachary (World Scientific, Singapore, 1984); *Phase Space Approach to Nuclear Dynamics*, edited by M. di Toro, W. Nörenberg, M. Rosina, and S. Stringari (World Scientific, Singapore, 1986), p. 546.
[10] D. J. Rowe, R. Le Blanc, and K. T. Hecht, J. Math. Phys. **29**, 287 (1988).
[11] D. J. Rowe, G. Rosensteel, and R. Gilmore, J. Math. Phys. **26**, 2787 (1985).
[12] J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972).
[13] D. J. Rowe and J. Carvalho, Phys. Lett. B **175**, 243 (1986).
[14] T. S. Dimitrescu, Rev. Roum. Phys. **22**, 145 (1977).
[15] R. C. King and A. H. A. Qubanchi J. Phys. A: Math. Gen **11**, 1491 (1978).
[16] G. Racah, Lectures notes on *Group Theory and Spectroscopy* (Institute of Advanced Study, Princeton, NJ, 1951).
[17] R. Le Blanc and D. J. Rowe, J. Math. Phys. **29**, 766 (1988).
[18] W. G. McKay and J. Patera, *Tables of Dimensions, Indices, and Branching Rules for Representations of Simple Lie Algebra* (Dekker, New York, 1981).
[19] R. C. King, J. Phys. A: Math. Gen. **8**, 429 (1975).
[20] J. D. Louck, Am. J. Phys. **38**, 3 (1970); R. Le Blanc and K. T. Hecht, J. Phys. A: Math. Gen. **20**, 4613 (1987).
[21] R. T. Sharp and C. S. Lam, J. Math. Phys. **11**, 2033 (1969).
[22] D. J. Rowe, J. Math. Phys. **9**, 1774 (1969).

The matrix representations of g_2. II. Representations in an su(3) basis

R. Le Blanc
A. W. Wright Nuclear Structure Laboratory, Yale University, New Haven, Connecticut 06511
D. J. Rowe
Physics Department, University of Toronto, Toronto, Ontario M5S 1A7, Canada

(Received 21 July 1987; accepted for publication 28 October 1987)

Irreducible representations of the real compact Lie algebra g_2 are given in $g_2 \supset$ su(3) bases. A missing label is accounted for by the explicit construction of a $g_2 \supset$ su(3) basis of vector holomorphic functions. Analytical results are given for two multiplicity-free classes of irreps. It is also shown how vector coherent state (VCS) theory accommodates the decomposition of the nilpotent raising operator subalgebra of an arbitrary Lie algebra into a finite but arbitrary number of irreducible tensorial sets under transformations generated by a stability algebra.

I. INTRODUCTION

It has recently been shown in considerable detail[1–3] how vector coherent state (VCS) theory provides, under certain conditions, a means to construct the ladder representations of a Lie algebra in a basis that is symmetry adapted to a subalgebra chain $\mathbf{g}\supset\mathbf{h}$. The conditions are that the subalgebra $\mathbf{h}$ contains a Cartan subalgebra and that the complex extension $\mathbf{g}^c$ of $\mathbf{g}$ can be decomposed

$$\mathbf{g} = \mathbf{n}_- + \mathbf{h}^c + \mathbf{n}_+,$$

where $\mathbf{n}_\pm$ are, respectively, nilpotent subalgebras of raising and lowering operators carrying (generally reducible) representations of $\mathbf{h}^c$, i.e., $[\mathbf{h}^c,\mathbf{n}_\pm] = \mathbf{n}_\pm$.

Now, while VCS theory easily gives functional representations of g, for practical applications one usually needs the explicit matrix forms of the representations in an orthonormal basis and, to this end, K-matrix theory[2,3] was developed. As originally presented, K-matrix theory was restricted to situations in which $\mathbf{n}_\pm$ are Abelian. This condition covered a wide range of situations of interest but it did not include, for example, the cases $\mathrm{so}(2n+1)\supset \mathrm{u}(n)$ or $g_2\supset\mathrm{so}(4)$, the latter the subject of part I of this series.[4] The theory was therefore developed and extended so that it would apply to these cases by Rowe *et al.*[3]

The situations in which $\mathbf{n}_\pm$ are Abelian are conveniently summarized by means of the following corollary, which follows immediately from Propositions 6 and 7 of Rowe *et al.*[3]

Corollary: The subalgebras $\mathbf{n}_\pm$ of $\mathbf{g}^c$ are Abelian if and only if $\mathbf{h}$ is a maximal subalgebra of $\mathbf{g}$. (Recall that a subalgebra $\mathbf{h}\subset\mathbf{g}$ is said to be maximal if it is a proper subalgebra and if there exists no other proper subalgebra $\mathbf{l}$ such that $\mathbf{g}\supset\mathbf{l}\supset\mathbf{h}$.)

Thus the first applications of VCS theory and K-matrix theory were to $\mathbf{g}\supset\mathbf{h}$ situations in which $\mathbf{h}$ was a maximal subalgebra such as, for example, $\mathrm{sp}(n)\supset\mathrm{u}(n)$, $\mathrm{su}(n+1)\supset\mathrm{u}(n)$, and $\mathrm{so}(2n)\supset\mathrm{u}(n)$. The developments of Rowe *et al.*[3] made it possible to apply the theory also to $\mathbf{g}\supset\mathbf{l}\supset\mathbf{h}$ situations such as $\mathrm{so}(2n+1)\supset\mathrm{o}(2n)\supset\mathrm{u}(n)$ and $g_2\supset\mathrm{so}(4)\supset\mathrm{u}(2)$. However, these are still special cases in that the raising operator algebra $\mathbf{n}_+$ is nilpotent of order 2; i.e., all double commutators vanish,

$$[x,[y,z]] = 0 \quad \forall x,y,z\in\mathbf{n}_+.$$

It was not obvious therefore that the theory would work in higher-order situations. In particular, it was not obvious that the techniques would apply, for example, to the case $g_2\supset\mathrm{su}(3)\supset\mathrm{u}(2)$ for which the raising operator algebra is of order 3. This example is therefore of considerable interest both for its own sake and as a prototype of situations with an arbitrary order n of nilpotency.

Thus after having successfully addressed the missing label problem for $g_2\supset\mathrm{so}(4)$ in the first part[4] of this series, we now address the parallel problem in a $g_2\supset\mathrm{su}(3)$ basis by exploiting once again the versatility of VCS theory.[1–3] We give an explicit basis construction for all irreducible ladder representations of g_2 when these are reduced with respect to $g_2\supset\mathrm{su}(3)$. We also give analytical results for the two multiplicity-free classes of $g_2\supset\mathrm{su}(3)$ classes. We demonstrate that the VCS construction of unirreps of $g_2\supset\mathrm{su}(3)\supset\mathrm{u}(2)$, which is a Lie algebra with a raising operator subalgebra nilpotent of order 3, does not differ substantially from the construction for Lie algebras with order of nilpotency $n=2$ developed by Rowe *et al.*[3]

We give in the Appendix a formal review of the most important structural properties of the K matrices. We show there that the details of application of VCS theory are independent of the order n of nilpotency of the Lie algebra under study (except for the $n=1$ case for which drastic simplifications occur). These developments seem to lead the way toward a final and fully generalized formulation of VCS theory that should apply to all semisimple Lie algebras, classical and exceptional, with an arbitrary order of nilpotency for the raising operator algebra.

II. THE $g_2\supset$su(3)$\supset$u(2) LIE ALGEBRA CHAIN

We recall[4] that the g_2 Lie algebra can be given as

$$g_2 = \mathrm{span}\{g_{ij} - \tfrac{1}{3}\delta_{ij}g_{kk}, e_i, f_j;\ 1\leqslant i,j\leqslant 3\} \tag{2.1}$$

with the commutation relations

$$\begin{aligned}
&[g_{ij},g_{kl}] = \delta_{jk}g_{il} - \delta_{il}g_{kj}, \quad [e_i,e_j] = -2\epsilon_{ijk}f_k,\\
&[g_{ij} - \tfrac{1}{3}\delta_{ij}g_{ll},e_k] = \delta_{jk}e_i - \tfrac{1}{3}\delta_{ij}e_k, \quad [f_i,f_j] = 2\epsilon_{ijk}e_k,\\
&[g_{ij} - \tfrac{1}{3}\delta_{ij}g_{ll},f_k] = -\delta_{ik}f_j + \tfrac{1}{3}\delta_{ij}f_k,\\
&[e_i,f_j] = 3g_{ij} - \delta_{ij}g_{kk}.
\end{aligned} \tag{2.2}$$

 0022-2488/88/040767-10$02.50

Note that, since $u(3) \not\subset g_2$ and g_2 is a Lie algebra of rank 2, only two linear combinations of the three weight operators $\{g_{11},g_{22},g_{33}\}$ of u(3) are needed to define a $u(1) \oplus u(1)$ Cartan subalgebra for g_2. A careful determination of the weight structure of g_2 [see Eqs. (2.3), (2.4), and (3.2) below] shows that the pair of weight operators

$$(g_{11} - g_{33}) = [g_{13},g_{31}], \quad (g_{22} - g_{33}) = [g_{23},g_{32}]$$

is, for our purposes, the appropriate basis for the Cartan subalgebra.

Under $\mathrm{su}(3) = \mathrm{span}\{g_{ij} - \frac{1}{3}\delta_{ij}g_{kk}\}$, the vectors $\mathbf{e}(=e^{\{1\}})$ and $\mathbf{f}(=f^{\{11\}})$ transform, respectively, as $\{1\}$ and $\{11\}$ tensors. We shall require the Hermiticity conditions

$$g_{ij} = g_{ji}^\dagger, \quad f_i = e_i^\dagger,$$

in order that the Lie algebra representations exponentiate to unitary representations of the group.

It is convenient for the following to consider the canonical (Gel'fand) u(2) subalgebra of su(3) to be given by

$$\mathrm{u}(2) = \mathrm{span}\{g_{\alpha\beta} - \delta_{\alpha\beta}g_{33};\ 1\leqslant\alpha,\beta\leqslant 2\}. \tag{2.3}$$

(We use Einstein's summation convention throughout. Furthermore, Roman indices run from 1 to 3 while Greek indices run from 1 to 2 only.) The Cartan subalgebra is then contained in the subalgebra chain

$$g_2 \supset \mathrm{su}(3) \supset \mathrm{u}(2) \supset \mathrm{u}(1) \oplus \mathrm{u}(1)$$

and the conditions for VCS theory are satisfied. Anticipating the terminology of VCS theory in the next section, we will refer to this u(2) subalgebra as the stability subalgebra.

Under transformations generated by the stability algebra, the various generators $t^{(\alpha_1\alpha_2)} \in g_2$ have the following u(2) tensorial rank $(\alpha_1\alpha_2)$:

$$g_2 \bmod \mathrm{u}(2) = \mathrm{span}\{g_{\sigma 3}^{(21)}, f_3^{(11)}, e_\sigma^{(10)}, f_\sigma^{(0,-1)}, e_3^{(-1,-1)}, g_{3\sigma}^{(-1,-2)};\ 1\leqslant\sigma\leqslant 2\}. \tag{2.4}$$

A unitary irrep of g_2 is labeled in the notation of King and Qubanchi[5] by $[\mu_1\mu_2]$, where μ_1 and μ_2 are the respective eigenvalues of the weight operators $g_{22} - g_{33}$ and $g_{11} - g_{22}$ on the g_2 highest-weight state defined by

$$(g_{11} - g_{33})|[\mu_1\mu_2]hw\rangle = (\mu_1 + \mu_2)|[\mu_1\mu_2]hw\rangle, \tag{2.5a}$$

$$(g_{22} - g_{33})|[\mu_1\mu_2]hw\rangle = \mu_1|[\mu_1\mu_2]hw\rangle, \tag{2.5b}$$

$$g_{ij}|[\mu_1\mu_2]hw\rangle = 0, \quad 1\leqslant i<j\leqslant 3, \tag{2.5c}$$

$$e_\sigma|[\mu_1\mu_2]hw\rangle = 0, \quad 1\leqslant\sigma\leqslant 2, \tag{2.5d}$$

$$f_3|[\mu_1\mu_2]hw\rangle = 0, \tag{2.5e}$$

where the generators g_{ij} $(1\leqslant i<j\leqslant 3)$, e_σ $(1\leqslant\sigma\leqslant 2)$, and f_3 span a set of positive roots[4] for g_2.

The branching rule $g_2 \downarrow \mathrm{su}(3)$, derived by King and Qubanchi[5] and by Perroud,[6] is given by

$$g_2\downarrow\mathrm{su}(3):\ [\mu_1\mu_2]\downarrow \sum_{\nu_1\geqslant\nu_2\geqslant 0} M^{\{\mu_1\mu_2\}}_{(\nu_1-\nu_2,\nu_2,\mu_1+\mu_2-\nu_1)}\{\nu_1\nu_2\}, \tag{2.6}$$

where $M^{\{\mu_1\mu_2\}}_{(\nu_1-\nu_2,\nu_2,\mu_1+\mu_2-\nu_1)}$ is the multiplicity of the weight

$$(n_1,n_2,n_3) = (\nu_1-\nu_2,\nu_2,\mu_1+\mu_2-\nu_1)$$

in the u(3)irrep$\{\mu_1\mu_2 0\}$ and where an su(3) irrep$\{\nu_1\nu_2\}$ is labeled by the highest weights of the weight operators $g_{11}-g_{33}, g_{22}-g_{33}$ [see Eq. (2.9)].

With $\mu(2)$ embedded in su(3) as in Eq. (2.3), the $\mathrm{su}(3)\downarrow\mathrm{u}(2)$ branching rule is given by

$$\mathrm{su}(3)\downarrow\mathrm{u}(2):\{\nu_1\nu_2\}\downarrow \sum_{\substack{\nu_1\geqslant m_{12}\geqslant\nu_2\\ \nu_2\geqslant m_{22}\geqslant 0}} (m_{12}-n_3, m_{22}-n_3) \equiv (h_1,h_2), \tag{2.7a}$$

where

$$n_3 = \nu_1 + \nu_2 - m_{12} - m_{22}. \tag{2.7b}$$

According to Racah,[7] the number of internal labels required to specify the basis states of an irrep of a compact group is $\frac{1}{2}(l-r)$, where l is the order of the group (number of generators) and r its rank (number of commuting weight operators). In the case of interest to us, the number of internal labels is $6 = \frac{1}{2}(14-2)$. Since an su(3) basis provides us with five labels, we need one extra label (b) to completely specify a $g_2\supset\mathrm{su}(3)$ basis corresponding to the branching rule (2.6). It will be shown in Sec. III how VCS theory assigns an operational meaning to this label. A $g_2\supset\mathrm{su}(3)\supset\mathrm{u}(2)\supset\mathrm{u}(1)$ basis will therefore be denoted, in Dirac notation, by

$$|[\mu_1\mu_2]b\{\nu_1\nu_2\}(h_1h_2)\eta\rangle, \quad h_1\geqslant\eta\geqslant h_2. \tag{2.8}$$

Of particular importance for our purpose is the existence, within a given g_2 unirrep, of a subset of su(3) highest-weight states defined by the conditions

$$(g_{11}-g_{33})|[\mu_1\mu_2]b\{\nu_1\nu_1\}(\nu_1\nu_2)\nu_1\rangle = \nu_1|[\mu_1\mu_2]b\{\nu_1\nu_2\}(\nu_1\nu_2)\nu_1\rangle, \tag{2.9a}$$

$$(g_{22}-g_{33})|[\mu_1\mu_2]b\{\nu_1\nu_2\}(\nu_1\nu_2)\nu_1\rangle = \nu_2|[\mu_1\mu_2]b\{\nu_1\nu_2\}(\nu_1\nu_2)\nu_1\rangle, \tag{2.9b}$$

$$g_{ij}|[\mu_1\mu_2]b\{\nu_1\nu_2\}(\nu_1\nu_2)\nu_1\rangle = 0, \quad 1\leqslant i<j\leqslant 3, \tag{2.9c}$$

for each $\{\nu_1\nu_2\}$ contained in the g_2 irrep$[\mu_1\mu_2]$. Our goal is the construction of this subset of states and the determination of the u(2)-reduced matrix elements of the tensors $e^{\{1\}}$ and $f^{\{11\}}$ between these states. The full matrix representations of $g_2\supset\mathrm{su}(3)$ are then obtained by means of the celebrated Wigner–Eckart theorem.

III. MATRIX REPRESENTATION OF g_2 IN AN su(3) BASIS

A. Vector coherent state theory

Vector coherent state (VCS) theory[1,2] and the associated K-matrix techniques for computing inner products[2,3] greatly simplify the construction of orthonormal bases for the ladder representations of a Lie algebra. In VCS theory, the familiar Cartan subalgebra $\mathbf{k}$ is extended to a larger non-Abelian stability algebra $\mathbf{h}$ that contains $\mathbf{k}$ as an Abelian subalgebra. The construction is a refinement of the well-known Cartan technique of building up basis states for ladder representations by the repeated actions of lowering (raising) operators on highest (lowest) states. The difference is that VCS theory exploits the strength of the Wigner–Eckart theorem by making use of the fact that one can regroup in a very convenient way the numerous generators of a given Lie algebra into a much smaller set of irreducible tensors under the stability algebra and, thus, take advantage of the existence of the tensor (Wigner–Racah) calculus for the stability subalgebra.

B. Application to the $g_2 \supset su(3) \supset u(2)$ chain

A Cartan subalgebra $u(1) \oplus u(1)$ for g_2 is spanned by the set of weight operators $\{g_{11} - g_{33}, g_{22} - g_{33}\}$. The highest-weight state for a g_2 irrep $[\mu_1 \mu_2]$ is then the state with maximal eigenvalues $\mu_1 + \mu_2$ and μ_1 for $g_{11} - g_{33}$ and $g_{22} - g_{33}$, respectively. This state is clearly also a highest-weight state for a highest-weight subrepresentation of the $u(2)$ subalgebra (2.3), and hence, by definition, $u(2)$ is the stability algebra for the corresponding subspace of states.

A set of raising operators for g_2 mod $u(2)$ is given by

$$\mathbf{n}_+ = \mathrm{span}\{g^{(21)}, f^{(11)}, e^{(10)}\}. \tag{3.1}$$

According to the commutation relations (2.2), $\mathbf{n}_+$ is a non-Abelian nilpotent subalgebra of order 3, i.e.,

$$\mathbf{n}_+ = \mathbf{n}^1_+ + \mathbf{n}^2_+ + \mathbf{n}^3_+, \tag{3.2a}$$

with

$$\mathbf{n}^3_+ = g^{(21)}, \quad \mathbf{n}^2_+ = f^{(11)}, \quad \mathbf{n}^1_+ = e^{(10)}, \tag{3.2b}$$

and

$$[\mathbf{n}_+, \mathbf{n}^i_+] \subset \sum_{j>i} \mathbf{n}^j_+ \tag{3.2c}$$

[see Eqs. (3.7a)–(3.7c)].

Let $\{|\eta\rangle\}$ be an arbitrary orthonormal basis for a $u(2)$ highest-weight subspace of a g_2 irrep. The VCS representation of a state $|\psi\rangle$ is then defined by

$$(y,z|\psi\rangle = \sum_\eta |\eta\rangle\langle\eta| \exp Z(y,z)|\psi\rangle, \tag{3.3a}$$

where

$$Z(y,z) = z^1_\alpha e_\alpha + z^2_\beta g_{\beta 3} + y f_3 \tag{3.3b}$$

and where y and z^α_β, $1 \leqslant \alpha,\beta \leqslant 2$, is a set of five complex (Bargmann) variables.

The vector coherent state representation $\Gamma(t)$ of an arbitrary generator $t \in g_2$ is defined by

$$\begin{aligned}\Gamma(t)(y,z|\psi\rangle &= \sum_n |\eta\rangle\langle\eta| \exp Z(y,z) t\,|\psi\rangle \\ &= |\eta\rangle\langle\eta|(t + [Z,t] + (1/2!)[Z,[Z,t]] \\ &\quad + \cdots)\exp Z(y,z)|\psi\rangle.\end{aligned} \tag{3.4}$$

With

$$\nabla = \frac{\partial}{\partial y}, \quad \partial^\alpha_\beta = \frac{\partial}{\partial z^\alpha_\beta}, \tag{3.5}$$

and with $\epsilon_{\alpha\beta}$ the totally antisymmetric tensor in two dimensions

$$\epsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{3.6}$$

we determine the following commutators:

$$[Z, g_{\sigma 3}] = 0, \tag{3.7a}$$

$$[Z, f_3] = 3 z^1_\alpha g_{\alpha 3}, \tag{3.7b}$$

$$[Z, e_\sigma] = 2\epsilon_{\sigma\mu} z^1_\mu f_3 - 3y g_{\sigma 3}, \tag{3.7c}$$

$$[Z, g_{\mu\nu} - \delta_{\mu\nu} g_{33}] = -z^1_\nu e_\mu - z^2_\nu g_{\mu 3} - \delta_{\mu\nu}(z^2_\beta g_{\beta 3} + y f_3), \tag{3.7d}$$

$$\begin{aligned}[Z, f_\sigma] &= 3 z^1_\alpha (g_{\alpha\sigma} - \delta_{\alpha\sigma} g_{33}) \\ &\quad - z^1_\sigma (g_{\mu\mu} - 2 g_{33}) - z^2_\sigma f_3 + 2y\epsilon_{\sigma\mu}\epsilon_\mu,\end{aligned} \tag{3.7e}$$

$$[Z, e_3] = 2 z^1_\sigma \epsilon_{\alpha\beta} f_\beta + z^2_\beta e_\beta + y(g_{\sigma\sigma} - 2g_{33}), \tag{3.7f}$$

$$[Z, g_{3\sigma}] = -z^1_\sigma e_3 + z^2_\beta (g_{\beta\sigma} - \delta_{\beta\sigma} g_{33}) + y f_\sigma, \tag{3.7g}$$

useful for the evaluation of the expansion (3.4).

With the help of (3.7) and the identities

$$\begin{aligned} g_{\sigma 3} \exp Z &= \partial^2_\sigma \exp Z, \\ f_3 \exp Z &= (\nabla - \tfrac{3}{2} c^{12}) \exp Z, \\ e_\sigma \exp Z &= (\partial^1_\sigma - \epsilon_{\sigma\mu} z^1_\mu \nabla + \tfrac{1}{2}\epsilon_{\sigma\mu} z^1_\mu c^{12} + 3y\,\partial^2_\sigma)\exp Z,\end{aligned} \tag{3.8}$$

where we have introduced the two mutually commuting $u(2)$ algebras

$$c_{\alpha\beta} = -z^\sigma_\beta\, \partial^\sigma_\alpha, \tag{3.9a}$$

$$c^{\alpha\beta} = z^\alpha_\sigma\, \partial^\beta_\sigma, \tag{3.9b}$$

we find the following VCS expansions for the various $u(2)$ tensors $t^{(\alpha_1 \alpha_2)} \in g_2$:

$$\Gamma(g^{(21)}_\sigma) = \partial^2_\sigma, \tag{3.10a}$$

$$\Gamma(f^{(11)}) = \nabla + \tfrac{3}{2} c^{12}, \tag{3.10b}$$

$$\Gamma(e^{(10)}_\sigma) = \partial^1_\sigma + \epsilon_{\sigma\mu} z^1_\mu \nabla + \tfrac{1}{2}\epsilon_{\sigma\mu} z^1_\mu c^{12} - \tfrac{3}{2} y\, \partial^2_\sigma, \tag{3.10c}$$

$$\begin{aligned}\Gamma(g_{\alpha\beta} - \delta_{\alpha\beta} g_{33}) &= (g^0_{\alpha\beta} - \delta_{\alpha\beta} g^0_{33}) \\ &\quad + c_{\alpha\beta} - \delta_{\alpha\beta}(c^{22} + y\nabla),\end{aligned} \tag{3.10d}$$

$$\begin{aligned}\Gamma(f^{(0-1)}_\sigma) &= 3 z^1_\mu (g^0_{\mu\sigma} - \delta_{\mu\sigma} g^0_{33}) \\ &\quad - z^1_\sigma (g^0_{\mu\mu} - 2 g^0_{33}) - z^1_\sigma(\tfrac{1}{2} y\nabla + c^{11}) \\ &\quad + 2y\epsilon_{\sigma\mu}\, \partial^1_\mu - z^2_\sigma \nabla - (\tfrac{3}{2} z^2_\sigma + \tfrac{1}{4} y z^1_\sigma) c^{12},\end{aligned} \tag{3.10e}$$

$$\begin{aligned}\Gamma(e^{(-1-1)}) &= y(g^0_{\mu\mu} - 2g^0_{33}) - y(y\nabla + \tfrac{3}{2} c^{11} + \tfrac{3}{2} c^{22}) \\ &\quad + 3\epsilon_{\mu\sigma} z^1_\mu\, z^1_\nu (g^0_{\nu\sigma} - \delta_{\nu\sigma} g^0_{33}) \\ &\quad + c^{21} - z^{12}_{12}\, \nabla - (\tfrac{1}{2} z^{12}_{12} + \tfrac{3}{4} y^2) c^{12},\end{aligned} \tag{3.10f}$$

$$\begin{aligned}\Gamma(g^{(-1-2)}_\sigma) &= -y z^1_\sigma (g^0_{\mu\mu} - 2g^0_{33}) + \tfrac{3}{2} y z^1_\mu (g^0_{\mu\sigma} - \delta_{\mu\sigma} g^0_{33}) \\ &\quad + \tfrac{1}{2} y z^1_\sigma (y\nabla + c^{11}) + y^2 \epsilon_{\sigma\mu}\, \partial^1_\mu \\ &\quad + z^1_\sigma \epsilon_{\alpha\mu} z^1_\mu z^2_\beta (g^0_{\beta\alpha} - \delta_{\beta\alpha} g^0_{33}) \\ &\quad + z^2_\mu (g^0_{\mu\sigma} - \delta_{\mu\sigma} g^0_{33}) - z^2_\sigma (c^{22} + y\nabla) \\ &\quad - z^1_\sigma c^{21} + \tfrac{1}{4} y^2 z^1_\sigma c^{12} + \tfrac{1}{2} y^3 \epsilon_{\sigma\mu}\, \partial^2_\mu\,.\end{aligned} \tag{3.10g}$$

In (3.10),

$$z^{12}_{12} = \epsilon_{\alpha\beta} z^1_\alpha\, z^2_\beta = \begin{vmatrix} z^1_1 & z^1_2 \\ z^2_1 & z^2_2 \end{vmatrix},$$

and $\{g^0_{\alpha\beta} - \delta_{\alpha\beta} g^0_{33}\}$ are intrinsic operators acting specifically on a basis of states carrying the *intrinsic* (highest-weight) representation of the $u(2)$ stability algebra.

From its definition, it is clear that a coherent state function $(y,z|\psi\rangle$ is a vector-valued holomorphic function of the variables (y, z^α_β). The inner product for the VCS Hilbert space of holomorphic functions has been given in integral form by Rowe *et al.*[8] For a practical purposes, however, it is generally much simpler and more useful to determine the inner product, and hence calculate Lie algebra matrix elements in an orthonormal Bargmann basis, by the K-matrix technique.[2,3]

C. Construction of an orthonormal $g_2 \supset \mathrm{su}(3) \supset \mathrm{u}(2)$ basis

We start by defining [Eq. (3.11)] a basis of states that is orthonormal with respect to a Bargmann inner product. The transformation K that maps this basis onto a basis that is orthonormal with respect to the VCS inner product is then determined by K-matrix theory and used to map the VCS representation Γ of g_2 to an equivalent representation $\gamma = K^{-1}\Gamma K$ that is unitary with respect to the Bargmann inner product. Matrix elements for this unirrep are then easily evaluated.

By Proposition of 2 of Rowe *et al.*[3] The Bargmann variables z^2, y, and z^1 transform, respectively, as the components of $(-1,-2)$, $(-1,-1)$, and $(0,-1)$ tensors under the Bargmann realization $\Gamma(g_{\alpha\beta} - \delta_{\alpha\beta}\, g_{33})$ of the u(2) stability algebra. An orthonormal basis of u(2)-coupled Bargmann states is therefore conveniently defined by the u(2)-coupled polynomial states

$$
\begin{aligned}
&(y,z^1,z^2|[\mu_1\mu_2]b\{\nu_1\nu_2\}(h_1h_2)\eta\rangle \\
&\quad = [P^{(-w_2,-2w_2)}(z^2)\times[(y^b/\sqrt{b!})\times[P^{(0,-w_1)}(z^1) \\
&\qquad \times|(\mu_1+\mu_2,\mu_1)\rangle]^{(\nu_1+b,\nu_2+b)}]^{(\nu_1\nu_2)}]_\eta^{(h_1h_2)}, \qquad (3.11)
\end{aligned}
$$

where $P^{(0,-w_1)}(z^1)$, e.g., is a polynomial of symmetric rank w_1 in the Bargmann spinor z^1 having highest-weight component

$$
P_0^{(0,-w_1)}(z^1) = (z_2^1)^{w_1}/\sqrt{w_1!}, \qquad (3.12)
$$

where the set of states $\{|(\mu_1+\mu_2,\mu_1)\eta\rangle\}$ span an intrinsic $\mathrm{u}(2)\subset g_2$ highest-weight subrepresentation acted upon specifically by the intrinsic operators $\{g^0_{\alpha\beta} - \delta_{\alpha\beta}g^0_{33}\}$, and where the various square brackets represent u(2) coupling with the convention that all such couplings are sequentially ordered from right to left.

A transformation K can now be defined that maps the Bargmann basis (3.11) onto a basis of states that are orthonormal with respect to the VCS inner product

$$
\begin{aligned}
&(y,z|\varphi([\mu_1\mu_2]b\{\nu_1\nu_2\}(h_1h_2)\eta)\rangle \\
&\quad = (y,z|K|[\mu_1\mu_2]b\{\nu_1\nu_2\}(h_1h_2)\eta\rangle.
\end{aligned}
$$

The VCS representation Γ of g_2, which is unitary with respect to the VCS inner product, correspondingly maps to an equivalent representation γ, which is unitary with respect to the Bargmann inner product

$$
\Gamma(t)\to\gamma(t) = K^{-1}\Gamma(t)K, \quad t\in g_2, \qquad (3.13a)
$$

$$
\gamma^\dagger(t) = \gamma(t^\dagger) = K^\dagger\Gamma^\dagger(t)(K^{-1})^\dagger. \qquad (3.13b)
$$

As shown in Rowe *et al.*[3] (Proposition 5), a choice for the transformation K can be made such that the VCS and Bargmann representations of g_2 are identical when restricted to the stability algebra [in this case u(2)], i.e.,

$$
\gamma(g_{\alpha\beta} - \delta_{\alpha\beta}g_{33}) = \Gamma(g_{\alpha\beta} - \delta_{\alpha\beta}g_{33}).
$$

That this is possible in the present case is easily verified by noting that the VCS realization Γ of $\mathrm{u}(2)\subset g_2$ is already unitary with respect to the Bargmann inner product. Thus in order that $\gamma(t) = \Gamma(t)$ for $t\in\mathrm{u}(2)$, we simply require K to be diagonal in the u(2) representation labels (h_1h_2) and independent of η,

$$
\begin{aligned}
&\langle[\mu]i\{\nu'\}(h')\eta'|K|[\mu]j\{\nu\}(h)\eta\rangle \\
&\quad = \delta_{hh'}\delta_{\eta\eta'}\langle[\mu]i\{\nu'\}(h)|K|[\mu]j\{\nu\}(h)\rangle. \qquad (3.14)
\end{aligned}
$$

As shown in Rowe *et al.*,[3] we can use the fact that the matrix elements for the su(3) subalgebra are already known[9] to restrict the construction of a VCS basis to that of a basis of su(3) highest-weight states,

$$
(y,z^1,z^2|K|[\mu_1\mu_2]b\{\nu_1\nu_2\}(\nu_1\nu_2)\eta\rangle. \qquad (3.15a)
$$

Since, by Eq. (3.10a) a su(3) VCS highest-weight state satisfies

$$
\begin{aligned}
&\Gamma(g_{\sigma 3}^{(21)})(y,z^1,z^2|K|[\mu_1\mu_2]b\{\nu_1\nu_2\}(\nu_1\nu_2)\eta\rangle \\
&\quad = \partial_\sigma^2(y,z^1,z^2|K|[\mu_1\mu_2]b\{\nu_1\nu_2\}(\nu_1\nu_2)\eta\rangle = 0, \quad (3.15b)
\end{aligned}
$$

it follows that the states (3.15a) must consist of a superposition of z^2-independent $(w_2=0)$ states (3.11),

$$
\begin{aligned}
&(y,z^1|[\mu_1\mu_2]b\{\nu_1\nu_2\}(\nu_1\nu_2)\eta\rangle \\
&\quad = [y^b/\sqrt{b!}\times[P^{(0,-w_1)}(z^1) \\
&\qquad \times|(\mu_1+\mu_2,\mu_1)\rangle]^{(\nu_1+b,\nu_2+b)}]_\eta^{(\nu_1\nu_2)} \qquad (3.16)
\end{aligned}
$$

and this implies (cf. Proposition 9 of Ref. 3) that

$$
\begin{aligned}
&\langle[\mu_1\mu_2]\bar b\{\bar\nu_1\bar\nu_2\}(\nu_1\nu_2)|K|[\mu_1\mu_2]b\{\nu_1\nu_2\}(\nu_1\nu_2)\rangle \\
&\quad = \delta_{\bar\nu_1\nu_1}\delta_{\bar\nu_2\nu_2}\langle[\mu_1\mu_2]\bar b\{\nu_1\nu_2\}(\nu_1\nu_2) \\
&\qquad \times|K|[\mu_1\mu_2]b\{\nu_1\nu_2\}(\nu_1\nu_2)\rangle. \qquad (3.17a)
\end{aligned}
$$

Furthermore, the VCS basis (3.15a) can be defined uniquely by choosing the submatrices

$$
\begin{aligned}
&\mathscr{K}(\nu_1\nu_2)_{ij} \\
&\quad = \langle[\mu_1\mu_2]i\{\nu_1\nu_2\}(\nu_1\nu_2)|K|[\mu_1\mu_2]j\{\nu_1\nu_2\}(\nu_1\nu_2)\rangle \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (3.17b)
\end{aligned}
$$

of the full (but generally non-Hermitian) K operator to be Hermitian; i.e.,

$$
\mathscr{K}(\nu_1\nu_2)^\dagger_{ij} = \mathscr{K}(\nu_1\nu_2)^*_{ji}.
$$

Note that a multiplicity space is defined, for the su(3) highest-weight states (3.16) and for a given partition $(\nu_1\nu_2)$, by the condition

$$
2b + w_1 = 2\mu_1 + \mu_2 - \nu_1 - \nu_2, \qquad (3.16')
$$

also taking into account of the usual selection (triangle) rules for the u(2) coupling.

Once we have determined the $\mathscr{K}(\nu)$ submatrices, we can infer all the relevant g_2 mod su(3) matrix elements. Indeed, as shown in the Appendix, upon using the definition (3.13) for the unitary representation γ, the Hermiticity condition

$$
\gamma(t^{(-\alpha_2,-\alpha_1)}) = \gamma^\dagger(t^{(\alpha_1\alpha_2)}), \quad (\alpha_1\alpha_2)>0 \qquad (3.18)
$$

and the Bargmann expansion (3.16) for the su(3) highest-weight states, one finds the following expressions for matrix elements between su(3) highest-weight states of the various su(2) tensors $t^{(\alpha_1\alpha_2)}$ belonging to g_2 mod su(3): (1) for the u(2) scalars $f^{(11)}$ and $e^{(-1,-1)}=e_3=f_3^\dagger=(f^{(11)})^\dagger$,

$$\langle[\mu]i(\nu')\|\gamma(f^{(11)}\|[\mu]j\{\nu\}(\nu)\rangle$$
$$=\langle[\mu]i\{\nu'\}(\nu')\|\mathscr{K}(\nu')^{-1}\nabla\mathscr{K}(\nu)\|[\mu]j\{\nu\}(\nu)\rangle, \tag{3.19a}$$

$$\langle[\mu]i\{\nu'\}(\nu')\|\gamma(e^{(-1,-1)})\|[\mu]j\{\nu\}(\nu)\rangle$$
$$=\langle[\mu]i\{\nu'\}(\nu')\|\mathscr{K}(\nu')y\mathscr{K}(\nu)^{-1}\|[\mu]j\{\nu\}(\nu)\rangle, \tag{3.19b}$$

(2) for the u(2) spinors $e^{(10)}$ and $f^{(0,-1)}=(e^{(10)})^\dagger$,

$$\langle[\mu]i\{\nu'\}(\nu')\|\gamma(e^{(10)})\|[\mu]j\{\nu\}(\nu)\rangle$$
$$=\langle[\mu]i\{\nu'\}(\nu')\|\mathscr{K}(\nu')^{-1}(\partial^1+z^1\nabla)\mathscr{K}(\nu)\|\times[\mu]j\{\nu\}(\nu)\rangle, \tag{3.20a}$$

$$\langle[\mu]i\{\nu'\}(\nu')\|\gamma(f^{(0,-1)})\|[\mu]j\{\nu\}(\nu)\rangle$$
$$=\langle[\mu]i\{\nu'\}(\nu')\|\mathscr{K}(\nu')(z^1+y\partial^1)\mathscr{K}(\nu)^{-1}\|\times[\mu]j\{\nu\}(\nu)\rangle\,. \tag{3.20b}$$

When divided by the appropriate su(3)$\supset$u(2)-reduced Wigner coefficients,[9] Eqs. (3.19) and (3.20) yield all the $g_2\supset$su(3)-reduced matrix elements needed to construct the full matrix representations of g_2.

Recursion formulas for the various $\mathscr{K}(\nu)$ matrices are derived in the Appendix. One finds the following pair of equations:

$$\langle[\mu]i\{\nu'\}(\nu')\|\mathscr{K}(\nu')^2[\Gamma^{(0)}(f^{(11)})]^\dagger\|[\mu]j\{\nu\}(\nu)\rangle$$
$$=\langle[\mu]i\{\nu'\}(\nu')\|\Gamma^{(0)}(e^{-1,-1)})\mathscr{K}(\nu)^2\|[\mu]j\{\nu\}(\nu)\rangle$$
$$-\sum_{l,\bar\nu>\nu}\frac{\langle[\mu]i\{\nu'\}(\nu')\|\Gamma^{(0)}(g^{(-1,-2)})\|[\mu]l\{\bar\nu\}(\bar\nu)\rangle}{\langle[\mu]l\{\bar\nu\}(\nu')\|\Gamma^{(1)}(g^{(-1,-2)})\|[\mu]l\{\bar\nu\}(\bar\nu)\rangle}$$
$$\times\langle[\mu]l\{\bar\nu\}(\nu')\|\Gamma^{(1)}(e^{(-1,-1)})\mathscr{K}(\nu)^2\|\times[\mu]j\{\nu\}(\nu)\rangle, \tag{3.21a}$$

$$\langle[\mu]i\{\nu'\}(\nu')\|\mathscr{K}(\nu')^2[\Gamma^{(0)}(e^{(10)}]^\dagger\|[\mu]j\{\nu\}(\nu)\rangle$$
$$=\langle[\mu]i\{\nu'\}(\nu')\|\Gamma^{(0)}(f^{(0,-1)})\mathscr{K}(\nu)^2\|[\mu]j\{\nu\}(\nu)\rangle$$
$$-\sum_{l,\bar\nu>\nu}\frac{\langle[\mu]i\{\nu'\}(\nu')\|\Gamma^{(0)}(g^{-1,-2)})\|[\mu]l\{\bar\nu\}(\bar\nu)\rangle}{\langle[\mu]l\{\bar\nu\}(\nu')\|\Gamma^{(1)}(g^{(-1,-2)})\|[\mu]l\{\bar\nu\}(\bar\nu)\rangle}$$
$$\times\langle[\mu]l\{\bar\nu\}(\nu')\|\Gamma^{(1)}(f^{(0,-1)})\mathscr{K}(\nu)^2\|\times[\mu]j\{\nu\}(\nu)\rangle, \tag{3.21b}$$

where the various restrictions $\Gamma^{(0)}(t)$ and $\Gamma^{(1)}(t)$, $t\in g_2$, are given by

$$[\Gamma^{(0)}(f^{(11)})]^\dagger=y, \tag{3.22}$$
$$[\Gamma^{(0)}(e_\sigma^{(10)})]^\dagger=z_\sigma^1+\epsilon_{\sigma\mu}y\,\partial_\mu^1, \tag{3.23}$$
$$\Gamma^{(0)}(f_\sigma^{(0,-1)})=3z_\mu^1(g_{\mu\sigma}^0-\delta_{\mu\sigma}g_{33}^0)-z_\sigma^1(g_{\mu\mu}^0-2g_{33}^0)-z_\sigma^1(\tfrac12y\nabla+c^{11})+2y\epsilon_{\sigma\mu}\,\partial_\mu^1, \tag{3.24a}$$
$$\Gamma^{(1)}(f_\sigma^{(0,-1)})=-z_\sigma^2\,\nabla, \tag{3.24b}$$
$$\Gamma^{(0)}(e^{(-1,-1)})=y(g_{\mu\mu}^0-2g_{33}^0)-y(y\nabla+\tfrac52c^{11})+3\epsilon_{\mu\sigma}z_\mu^1z_\nu^1(g_{\nu\sigma}^0-\delta_{\nu\sigma}g_{33}^0), \tag{3.25a}$$
$$\Gamma^{(1)}(e^{(-1,-1)})=c^{21}-z_{12}^{12}\,\nabla, \tag{3.25b}$$
$$\Gamma^{(0)}(g_\sigma^{(-1,-2)})=-yz_\sigma^1(g_{\mu\mu}^0-2g_{33}^0)+\tfrac32yz_\mu^1(g_{\mu\sigma}^0-\delta_{\mu\sigma}g_{33}^0)+\tfrac12yz_\sigma^1(y\nabla+c^{11})+y^2\epsilon_{\sigma\mu}\,\partial_\mu^1+z_\sigma^1\epsilon_{\alpha\mu}z_\mu^1z_\beta^1(g_{\beta\alpha}^0-\delta_{\beta\alpha}g_{33}^0), \tag{3.26a}$$
$$\Gamma^{(1)}(g_\sigma^{(-1,-2)})=z_\mu^2(g_{\mu\sigma}^0-\delta_{\mu\sigma}g_{33}^0)-z_\sigma^2y\nabla-z_\sigma^1\,c^{21}. \tag{3.26b}$$

Note that $\Gamma^{(0)}(t)$ maps an su(3) highest-weight state to another such state while $\Gamma^{(1)}(t)$ maps an su(3) highest-weight state to a z^2-dependent state with $w_2=1$ in (3.11). Together with $\mathscr{K}(\mu_1+\mu_2,\mu_1)=1$, Eqs. (3.21) and (3.21b) are sufficient to determine all the needed $\mathscr{K}$ submatrices.

The Bargmann basis (3.16) of VCS su(3) highest-weight irreps is, in general, overcomplete and this results in the appearance of zero eigenvalues for $\mathscr{K}(\nu)^2$. The eigenvectors associated with the zero eigenvalues clearly correspond to state vectors that are nonvanishing in the Bargmann space but that map into null VCS vectors. We therefore retain only the subset of eigenvectors with nonvanishing $\mathscr{K}(\nu)^2$ eigenvalues. The span of these non-null vectors defines the VCS space as a subspace of Bargmann space, sometimes referred to as the physical subspace.

IV. EXAMPLES

A. The g_2 irrep $[\mu_1,\mu_2=0]$

The g_2 irrep $[\mu_1 0]$ is multiplicity-free, i.e., the branching rule $g_2\downarrow$su(3) [Eq. (2.6)] gives

$$[\mu_1 0]\downarrow\sum_{0\leqslant\nu_2\leqslant\nu_1\leqslant\mu_1}\{\nu_1\nu_2\} \tag{4.1}$$

with each su(3) irrep appearing only once. We parametrize the various su(3) partitions $\{\nu_1\nu_2\}$ appearing (4.1) by (see Table I)

$$\{\nu_1\nu_2\}=\{\mu_1-\theta,\mu_1-\theta-\lambda\},\quad 0\leqslant\theta\leqslant\mu_1,\quad 0\leqslant\lambda\leqslant\mu_1-\theta. \tag{4.2}$$

From a VCS point of view, the irrep $[\mu_1 0]$ is particularly

TABLE I. Parametrization $\{\nu_1\nu_2\}=\{\mu_1-\theta,\mu_1-\theta-\lambda\}$ for the su(3) irreps belonging to the g_2 irrep $[\mu_1 0]$.

θ strings↓ \ λ string→				
$\{\mu_1,\mu_1\}$	$\{\mu_1,\mu_1-1\}$	$\{\mu_1,\mu_1-2\}$	$\cdots$	$\{\mu_1,0\}$
$\{\mu_1-1,\mu_1-1\}$	$\{\mu_1-1,\mu_1-2\}$	$\cdots$	$\{\mu_1-1,0\}$	
$\vdots$	$\vdots$	$\cdots$		
$\{11\}$	$\{10\}$			
$\{00\}$				

simple since it has a scalar intrinsic u(2) irrep. Its Hilbert subspace of su(3) highest states is spanned by

$$\langle y,z^1|[\![\mu_1 0]\!]\theta\{\nu_1\nu_2\}(\nu_1\nu_2)\eta\rangle = (y^\theta/\sqrt{\theta!})P^{(0,-\lambda)}_{\eta+\theta-\mu_1}(z^1)|(\mu_1\mu_1)\mu_1\rangle, \tag{4.3}$$

where there is an unambiguous one-to-one correspondence between the parameters θ, λ in (4.2) and (4.3) and the rank parameters $b=\theta$, $w_1=\lambda$ in (3.16). Since the intrinsic u(2) irrep is one dimensional, the u(2) coupling in (4.3) is trivial. The space of su(3) highest-weight states is therefore, by construction, multiplicity-free. Furthermore, since the polynomials $P^{(0,-w_1)}(z^1)$ are well known and their construction unambiguous [as opposed to their counterparts in the $g_2\supset$so(4) case[4]], we are able to derive fully analytical results for the g_2 irrep $[\![\mu_1 0]\!]$ as we now show.

We first seek to compute $\mathscr{K}(\mu_1-\theta,\mu_1-\theta)^2$. From (3.21a), we have

$$\mathscr{K}(\mu_1-\theta-1,\mu_1-\theta-1)^2\times\sqrt{\theta+1} = \left\{\sqrt{\theta+1}(2\mu_1-\theta) - \frac{(-)\sqrt{2\theta(\theta+1)}}{(u_1-\theta+2)}\times(-)\sqrt{2\theta}\right\}\mathscr{K}(\mu_1-\theta,\mu_1-\theta)^2, \tag{4.4a}$$

which yields

$$\frac{\mathscr{K}(\mu_1-\theta-1,\mu_1-\theta-1)^2}{\mathscr{K}(\mu_1-\theta,\mu_1-\theta)^2} = \frac{(2\mu_1-\theta+4)(\mu_1-\theta)}{(\mu_1-\theta+2)} \tag{4.4b}$$

or

$$\mathscr{K}(\mu_1-\theta,\mu_1-\theta)^2 = \frac{(2\mu_1+4)!}{(2\mu_1+4-\theta)!}\cdot\frac{\mu_1!}{(\mu_1-\theta)!}\cdot\frac{(\mu_1+2-\theta)!}{(\mu_1+2)!}. \tag{4.4c}$$

Similarly, Eq. (3.21b) yields

$$\frac{\mathscr{K}(\mu_1-\theta-1,\mu_1-\theta-\lambda-1)^2}{\mathscr{K}(\mu_1-\theta,\mu_1-\theta-\lambda)^2} = (\mu_1-\theta-\lambda) \tag{4.5a}$$

or

$$\frac{\mathscr{K}(\mu_1-\theta,\mu_1-\theta-\lambda)^2}{\mathscr{K}(\mu_1-\theta,\mu_1-\theta)^2} = \frac{(\mu_1-\theta)!}{(\mu_1-\theta-\lambda)!}. \tag{4.5b}$$

Thus

$$\mathscr{K}(\mu_1-\theta,\mu_1-\theta-\lambda)^2 = \frac{(\mu_1-\theta)!}{(\mu_1-\theta-\lambda)!}\times\mathscr{K}(\mu_1-\theta,\mu_1-\theta)^2 \tag{4.6a}$$

$$= \frac{(2\mu_1+4)!}{(2\mu_1+4-\theta)!}\cdot\frac{\mu_1!}{(\mu_1-\theta-\lambda)!}\cdot\frac{(\mu_1+2-\theta)!}{(\mu_1+2)!} \tag{4.6b}$$

or, using (4.2),

$$\mathscr{K}(\nu_1\nu_2)^2 = \frac{(2\mu_1+4)!}{(\mu_1+\nu_1+4)!}\cdot\frac{\mu_1!}{\nu_2!}\cdot\frac{(\nu_1+2)!}{(\mu_1+2)!}. \tag{4.6c}$$

Note that Eqs. (4.4b) and (4.5a) clearly indicate that $\mathscr{K}(\nu_1\nu_2)^2$ will vanish beyond the edges of the triangle in Table I.

With the help of (3.19) and (3.20), we find

$$\langle[\![\mu_1 0]\!]\{\nu_1-1,\nu_2-1\}(\nu_1-1,\nu_2-1)\times\|\gamma(e^{(-1,-1)})\|[\![\mu_1 0]\!]\{\nu_1\nu_2\}(\nu_1\nu_2)\rangle = \left[\frac{\nu_2(\mu_1-\nu_1+1)(\mu_1+\nu_1+4)}{(\nu_1+2)}\right]^{1/2}, \tag{4.7a}$$

$$\langle[\![\mu_1 0]\!]\{\nu_1-1,\nu_2\}(\nu_1-1,\nu_2)\times\|\gamma(f^{(0,-1)})\|[\![\mu_1 0]\!]\{\nu_1\nu_2\}(\nu_1\nu_2)\rangle = \left[\frac{(\nu_1-\nu_2+1)(\mu_1-\nu_1+1)(\mu_1+\nu_1+4)}{(\nu_1+2)}\right]^{1/2}, \tag{4.7b}$$

$$\langle[\![\mu_1 0]\!]\{\nu_1,\nu_2-1\}(\nu_1,\nu_2-1)\|\gamma(f^{(0,-1)})\times\|[\![\mu_1 0]\!]\{\nu_1\nu_2\}(\nu_1\nu_2)\rangle = [\nu_2(\nu_1-\nu_2+1)]^{1/2}, \tag{4.7c}$$

which, when divided by the proper su(3)$\supset$u(2) reduced Wigner coefficients,[9] yield the following $g_2\supset$su(3) (triple-bar) reduced matrix elements: for the su(3) tensor $e^{\{1\}}$,

$$\langle[\![\mu_1 0]\!]\{\nu_1-1,\nu_2-1\}|||\gamma(e^{\{1\}})|||[\![\mu_1 0]\!]\{\nu_1\nu_2\}\rangle = \left[\frac{(\nu_2+1)(\mu_1-\nu_1+1)(\mu_1+\nu_1+4)}{(\nu_1+1)}\right]^{1/2}, \tag{4.8a}$$

$$\langle[\![\mu_1 0]\!]\{\nu_1+1,\nu_2\}|||\gamma(e^{\{1\}})|||[\![\mu_1 0]\!]\{\nu_1\nu_2\}\rangle = \left[\frac{(\nu_1-\nu_2+1)(\mu_1-\nu_1)(\mu_1+\nu_1+5)}{(\nu_1+3)}\right], \tag{4.8b}$$

$$\langle[\![\mu_1 0]\!]\{\nu_1,\nu_2+1\}|||\gamma(e^{\{1\}})|||[\![\mu_1 0]\!]\{\nu_1\nu_2\}\rangle = (-)[(\nu_2+1)(\nu_1-\nu_2+1)]^{1/2}, \tag{4.8c}$$

and, using the conjugation property[9]

$$\langle[\![\mu_1\mu_2]\!]b'\{\nu_1'\,\nu_2'\}|||\gamma(t^{\{\alpha_1\alpha_2\}})|||[\![\mu_1\mu_2]\!]b\{\nu_1\nu_2\}\rangle = (-1)^{\phi(\{\nu_1\nu_2\})+\phi(\{\alpha_1\alpha_2\})-\phi(\{\nu_1'\,\nu_2'\})}\times\left[\frac{\dim\{\nu_1\nu_2\}}{\dim\{\nu_1'\,\nu_2'\}}\right]^{1/2}\times\langle[\![\mu_1\mu_2]\!]b\{\nu_1\nu_2\}|||\gamma(t^{\{\alpha_1,\alpha_1-\alpha_2\}})|||[\![\mu_1\mu_2]\!]b'\{\nu_1'\,\nu_2'\}\rangle, \tag{4.9}$$

for the su(3) tensor $f^{\{11\}}$,

$$\langle[\![\mu_1 0]\!]\{\nu_1+1,\nu_2+1\}|||\gamma(f^{\{11\}})|||[\![\mu_1 0]\!]\{\nu_1\nu_2\}\rangle = \left[\frac{(\nu_2+1)(\mu_1-\nu_1)(\mu_1+\nu_1+5)}{(\nu_1+3)}\right]^{1/2}, \tag{4.10a}$$

$$\langle[\![\mu_1 0]\!]\{\nu_1-1,\nu_2\}|||\gamma(f^{\{11\}})|||[\![\mu_1 0]\!]\{\nu_1\nu_2\}\rangle = \left[\frac{(\nu_1-\nu_2+1)(\mu_1-\nu_1+1)(\mu_1+\nu_1+4)}{(\nu_1+1)}\right], \tag{4.10b}$$

$$\langle[\![\mu_1 0]\!]\{\nu_1,\nu_2-1\}|||\gamma(f^{\{11\}})|||[\![\mu_1 0]\!]\{\nu_1\nu_2\}\rangle = [(\nu_2+1)(\nu_1-\nu_2+1)]^{1/2}. \tag{4.10c}$$

With the normalization (2.2) for the $g_2 \supset \mathrm{su}(3)$ Lie algebra, the g_2 quadratic Casimir invariant is given by

$$I^{(2)}_{g_2} = I^{(2)}_{\mathrm{su}(3)} + \tfrac{2}{3} e\cdot f, \quad e\cdot f = e_i f_i, \tag{4.11a}$$

with eigenvalue[10]

$$\begin{aligned} I^{(2)}_{g_2}(\mu_1\mu_2) &\\ &= \tfrac{1}{9}[(2\mu_2+\mu_1)(2\mu_2+\mu_1+8) + (\mu_1-\mu_2) \\ &\quad \times(\mu_1-\mu_2+2) + (2\mu_1+\mu_2)(2\mu_1+\mu_2+10)] \end{aligned} \tag{4.11b}$$

on the g_2 irrep $[\mu_1\mu_2]$. It is easy to verify, with

$$I^{(2)}_{\mathrm{su}(3)}(\nu_1\nu_2) = \tfrac{2}{3}[\nu_1^2 - \nu_1\nu_2 + \nu_2^2 + 3\nu_1] \tag{4.12}$$

and

$$\begin{aligned} &\langle [\mu_1\mu_2] b\{\nu_1\nu_2\} \| | \gamma(e\cdot f) | \| [\mu_1\mu_2] b\{\nu_1\nu_2\}\rangle \\ &= \sum_{\bar b \bar\nu_1 \bar\nu_2} \left[\frac{\dim\{\bar\nu_1\bar\nu_2\}}{\dim\{\nu_1\nu_2\}}\right] \\ &\quad \times \langle [\mu_1\mu_2]\bar b\{\bar\nu_1\bar\nu_2\} | \, \| \gamma(f^{\{11\}}) \| \, | [\mu_1\mu_2] b\{\nu_1\nu_2\}\rangle^2, \end{aligned} \tag{4.13}$$

that the su(3)-reduced matrix elements (4.8) and (4.9) obey the sum rule

$$I^{(2)}_{g_2}(\mu_1 0) = \tfrac{2}{3}\mu_1(\mu_1+5) \tag{4.14}$$

in agreement with Eq. (4.11b).

We remark that the g_2 irreps belonging to the $[\mu_1 0]$ class of irreps are, from a VCS point of view, the only genuine multiplicity-free irreps and therefore the only class of irreps for which one can expect to derive analytical results. Irreps of the $[\mu_1\mu_1]$ kind belong to the only other class of multiplicity-free irreps but do so in a manner which, from a VCS point of view, is more characteristic of the generic situation as we now discuss.

B. The g_2 irrep $[\mu_1,\mu_2=\mu_1]$

The $g_2 \downarrow \mathrm{su}(3)$ reduction for the irrep $[\mu_1,\mu_2=\mu_1]$ is multiplicity-free. But, whereas it was *a priori* obvious from the trivial nature of the u(2) coupling in (4.3) that the $[\mu_1 0]$ irrep should be multiplicity-free, one might have expected exactly the opposite for a $[\mu_1\mu_1]$ irrep.

One recalls that the intrinsic (highest-weight) irrep for a generic g_2 irrep $[\mu_1\mu_2]$ carries a nonscalar u(2) irrep $(\mu_1+\mu_2,\mu_1)$ and, consequently, an intrinsic "angular momentum" $j=\frac{1}{2}\mu_2$. Thus one might expect that, among the class of g_2 irreps $[\mu_1\mu_2]$ with $\mu_1+\mu_2=\mathrm{const}=2\mu$, the multiplicity would increase with increasing μ_2. Indeed, this happens, as Table I of King and Qubanchi[5] illustrates very graphically, until $\mu_1=\mu_2$ when the irrep suddenly becomes multiplicity-free. It is therefore of considerable interest and a critical test of K-matrix theory to see how this comes about.

The $g_2 \downarrow \mathrm{su}(3)$ branching rule for the irrep $[\mu_1\mu_1]$ is given by

$$[\mu_1\mu_1]\downarrow \sum_{\substack{\mu_1\leqslant\nu_1\leqslant 2\mu_1 \\ \nu_1-\mu_1\leqslant\nu_2\leqslant\mu_1}} \{\nu_1\nu_2\}, \tag{4.15}$$

where each su(3) irrep appears only once. We parametrize the various su(3) partitions $\{\nu_1\nu_2\}$ appearing in (4.15) by (see Table II)

$$\begin{aligned} \{\nu_1\nu_2\} &= \{2\mu_1-\lambda-\theta,\mu_1-\theta\}, \\ &\quad 0\leqslant\theta\leqslant\mu_1, \quad 0\leqslant\lambda\leqslant\mu_1-\theta. \end{aligned} \tag{4.16}$$

The Bargmann subspace of su(3) highest-weight states is then spanned by the states

$$\begin{aligned} &\langle y,z^1 | [\mu_1\mu_1] i\{\nu_1\nu_2\}(\nu_1\nu_2)\eta\rangle \\ &= [(y^i/\sqrt{i!})[P^{(0,-(\lambda+2\theta-2i))}(z^1) \\ &\quad \times |(2\mu_1,\mu_1)\rangle]^{(\nu_1+i,\nu_2+i)}]^{(\nu_1\nu_2)}_\eta, \\ &\quad 0\leqslant i\leqslant \min(\theta,\mu_1-\lambda), \end{aligned} \tag{4.17}$$

i.e., we have set $b=i$ and $w_1=\lambda+2\theta-2i$ in (3.16).

According to the branching rule (4.15), we expect each $\mathscr{K}^2$ matrix to have only one nonvanishing eigenvalue. Thus only one specific linear combination of Bargmann basis states belonging to the multiplicity space (4.17) will survive under the map K to the corresponding one-dimensional VCS space.

Starting with the one-dimensional ($\theta=0$) matrices

$$\mathscr{K}(2\mu_1-\lambda,\mu_1)^2 = (3\mu_1+3)!/(3\mu_1+3-\lambda)! \tag{4.18}$$

easily obtained from (3.21b), one proceeds to the right of Table II in a stepwise fashion by increasing the value of the parameter θ in (4.16). Using simultaneously (3.21a) and (3.21b), one obtains in the i-ordered basis (4.17) (since $\mathscr{K}$ is Hermitian we give only its right upper elements): for $\theta=1$,

TABLE II. Parametrization $\{\nu_1\nu_2\}=\{2\mu_1-\lambda-\theta,\mu_1-\theta\}$ for the su(3) irreps belonging to the g_2 irrep $[\mu_1\mu_1]$.

λ strings↓	θ strings→				
	$\{2\mu_1,\mu_1\}$	$\{2\mu_1-1,\mu_1-1\}$	$\{2\mu_1-2,\mu_1-2\}$	$\cdots$	$\{\mu_1,0\}$
	$\{2\mu_1-1,\mu_1\}$	$\{2\mu_1-2,\mu_1-1\}$	$\cdots$	$\{\mu_1,1\}$	
	$\vdots$	$\vdots$	$\iddots$		
	$\{\mu_1+2,\mu_1\}$	$\{\mu_1+1,\mu_1-1\}$	$\{\mu_1,\mu_1-2\}$		
	$\{\mu_1+1,\mu_1\}$	$\{\mu_1,\mu_1-1\}$			
	$\{\mu_1,\mu_1\}$				

$$\mathscr{K}(2\mu_1-\lambda-1,\mu_1-1)^2=\frac{(2\mu_1+2)}{(\mu_1+1)}\cdot\frac{1}{(2\mu_1+2-\lambda)}\cdot\mathscr{K}(2\mu_1-\lambda-1,\mu_1)^2$$
$$\times\begin{pmatrix}(\lambda+1)(\mu_1+2) & \sqrt{(\lambda+1)(\mu_1+2)(\mu_1-\lambda)}\\ \cdots & (\mu_1-\lambda)\end{pmatrix}, \tag{4.19}$$

for $\theta=2$,

$$\mathscr{K}(2\mu_1-\lambda-2,\mu_1-2)^2$$
$$=\frac{(2\mu_1+2)(2\mu_1+1)}{(\mu_1+1)(\mu_1)}\cdot\frac{1}{(2\mu_1+2-\lambda)(2\mu_1+2-\lambda-1)}\cdot\mathscr{K}(2\mu_1-\lambda-2,\mu_1)^2$$
$$\times\begin{pmatrix}(\lambda+2)(\lambda+1)(\mu_1+3)(\mu_1+2) & (\lambda+1)(\mu_2+2)\sqrt{2(\mu_1-\lambda-1)(\lambda+2)(\mu_1+3)} & \sqrt{(\lambda+2)(\lambda+1)(\mu_1+3)(\mu_1+2)(\mu_1-\lambda)(\mu_1-\lambda-1)}\\ \cdots & 2(\mu_1-\lambda-1)(\lambda+1)(\mu_1+2) & (\mu_1-\lambda-1)\sqrt{2(\lambda+1)(\mu_1+2)(\mu_1-\lambda)}\\ \cdots & \cdots & (\mu_1-\lambda)(\mu_1-\lambda-1)\end{pmatrix} \tag{4.20}$$

from which one can infer by induction that, for given λ and θ, $\mathscr{K}^2$ is given by the $(\theta+1)\times(\theta+1)$ square matrix

$$\mathscr{K}(2\mu_1-\lambda-\theta,\mu_1-\theta)^2_{ij}$$
$$=\frac{(2\mu_1+2)!}{(2\mu_1+2-\theta)!}\cdot\frac{(\mu_1+1-\theta)!}{(\mu_1+1)!}\cdot\frac{(2\mu_1+2-\lambda-\theta)!}{(2\mu_1+2-\lambda)!}\cdot\frac{(3\mu_1+3)!}{(3\mu_1+3-\lambda-\theta)}!\times\sqrt{\beta_i^{\mu_1}(\lambda\theta)\beta_j^{\mu_1}(\lambda\theta)} \tag{4.21}$$

with

$$\beta_i^{\mu_1}(\lambda\theta)=\binom{\theta}{i}\frac{(\mu_1-\lambda-\theta+i)!}{(\mu_1-\lambda-\theta)!}.\frac{(\mu_1+\theta+1-i)!}{(\mu_1+1)}\cdot\frac{(\lambda+\theta-i)!}{\lambda!}. \tag{4.21$'$}$$

Note that only the right lower $\theta\times\theta$ submatrix survives when $\lambda=\mu_1-\theta$ and that this submatrix is proportional to $(\mu_1-\lambda-\theta+1)$, i.e., it vanishes identically, as it should, beyond the right edge of Table II.

Since all rows of the Hermitian matrix (4.21a) are proportional to each other, this matrix is singular and, as expected, has only one nonvanishing eigenvalue given by the trace tr $\mathscr{K}^2$. Its corresponding normalized eigenvector $\mathbf{x}=(x_i)$ has components

$$x_i^{\mu_1}(\lambda\theta)=\left(\frac{\beta_i^{\mu_1}(\lambda\theta)}{[\Sigma_k\beta_k^{\mu_1}(\lambda\theta)]}\right)^{1/2} \tag{4.22}$$

on the i-ordered basis (4.17).

As expected for a generic multiplicity case, matrix elements in this example involve sums that are irreducible to simple monomials and thus can only give the following semianalytical results: using interchangeably ν_1,ν_2, or λ,θ through the parametrization (4.16), we have

$$\langle[\mu_1\mu_1]\{\nu_1-1,\nu_2-1\}(\nu_1-1,\nu_2-1)\|\gamma(e^{(-1,-1)})\|[\mu_1\mu_1]\{\nu_1\nu_2\}(\nu_1\nu_2)\rangle$$
$$=\left[\frac{\text{tr}\,\mathscr{K}(\nu_1-1,\nu_2-1)^2}{\text{tr}\,\mathscr{K}(\nu_1,\nu_2)^2}\right]^{1/2}\sum_{ij}\left\{\left(\frac{\beta_i^{\mu_1}(\lambda,\theta-1)}{[\Sigma_k\beta_k^{\mu_1}(\lambda,\theta-1)]}\right)^{1/2}\langle[\mu_1\mu_1]i\{\nu_1-1,\nu_2-1\}(\nu_1-1,\nu_2-1)\|y\|\right.$$
$$\times[\mu_1\mu_1]j\{\nu_1\nu_2\}(\nu_1\nu_2)\rangle\left(\frac{\beta_j^{\mu_1}(\lambda\theta)}{[\Sigma_l\beta_l^{\mu_1}(\lambda\theta)]}\right)^{1/2}, \tag{4.23a}$$

$$\langle[\mu_1\mu_1]\{\nu_1-1,\nu_2\}(\nu_1-1,\nu_2)\|\gamma(f^{(0,-1)})\|[\mu_1\mu_1]\{\nu_1\nu_2\}(\nu_1\nu_2)\rangle$$
$$=\left[\frac{\text{tr}\,\mathscr{K}(\nu_1-1,\nu_2)^2}{\text{tr}\,\mathscr{K}(\nu_1,\nu_2)^2}\right]^{1/2}\sum_{ij}\left\{\left(\frac{\beta_i^{\mu_1}(\lambda+1,\theta)}{[\Sigma_k\beta_k^{\mu_1}(\lambda+1,\theta)]}\right)^{1/2}\langle[\mu_1\mu_1]i\{\nu_1-1,\nu_2\}(\nu_1-1,\nu_2)\|(z^1+y\partial^1)\|\right.$$
$$\left.\times[\mu_1\mu_1]j\{\nu_1\nu_2\}(\nu_1\nu_2)\rangle\left(\frac{\beta_j^{\mu_1}(\lambda\theta)}{[\Sigma_l\beta_l^{\mu_1}(\lambda\theta)]}\right)^{1/2}\right\}, \tag{4.23b}$$

$$\langle[\mu_1\mu_1]\{\nu_1,\nu_2-1\}(\nu_1,\nu_2-1)\|\gamma(f^{(0,-1)})\|[\mu_1\mu_1]\{\nu_1\nu_2\}(\nu_1\nu_2)\rangle$$
$$=\left[\frac{\text{tr}\,\mathscr{K}(\nu_1,\nu_2-1)^2}{\text{tr}\,\mathscr{K}(\nu_1,\nu_2)^2}\right]^{1/2}\sum_{ij}\left\{\left(\frac{\beta_i^{\mu_1}(\lambda-1,\theta+1)}{[\Sigma_k\beta_k^{\mu_1}(\lambda-1,\theta+1)]}\right)^{1/2}\langle[\mu_1\mu_1]i\{\nu_1,\nu_2-1\}(\nu_1,\nu_2-1)\|(z^1+y\partial^1)\|\right.$$
$$\left.\times[\mu_1\mu_1]j\{\nu_1\nu_2\}(\nu_1\nu_2)\rangle\right\}\left(\frac{\beta_j^{\mu_1}(\lambda\theta)}{[\Sigma_l\beta_l^{\mu_1}(\lambda\theta)]}\right)^{1/2}, \tag{4.23c}$$

which, with the help of the Wigner–Eckart theorem and of (4.9), would yield all desired $g_2 \supset \mathrm{su}(3)$-reduced matrix elements.

APPENDIX: STRUCTURAL PROPERTIES OF THE K MATRICES

The purpose of this Appendix is to prove the equalities defining the matrix elements (3.19) and (3.20), and to derive the recursion formulas (3.21) for the restriction $\mathscr{K}(\nu)$ defined by Eq. (3.17b). Most of the derivations to be given below depend on specific structural properties of the full transformation matrix K on which we first elaborate.

First, we show that

$$\langle [\mu] i\{\nu'\}(h) | K | [\mu] j\{\nu\}(h)\rangle = 0 \quad \text{for } \nu' > \nu, \quad (\mathrm{A1})$$

i.e., the matrix K is lower triangular in the (h) blocks.

To prove (A1), start with the equality

$$P_{\eta'}^{(2m,m)}(\gamma(g^{(21)}))|[\mu] j\{\nu\}(h)\eta\rangle = 0 \quad \text{for } 3m > \nu - h, \quad (\mathrm{A2})$$

where

$$\nu - h = (\nu_1 + \nu_2) - (h_1 + h_2)$$

and where $P_{\eta'}^{(2m,m)}(\gamma(g^{(21)}))$ is a (symmetric) polynomial in the commuting raising operators $\gamma(g^{(21)})$ of su(3) mod u(2). Equation (A2) states that there is a highest-weight state for the su(3) irrep $\{\nu\}$ and that the various u(2) unirreps (h) lying in this unirrep can be ordered by the decreasing eigenvalues of the $\mathrm{u}(1) \subset \mathrm{u}(2)$ number operator $g_{11} + g_{22} - 2g_{33}$, where the former are integer multiples of 3. Since $\gamma(g^{(21)}) = K^{-1}\,\Gamma(g^{(21)})K$, we can rewrite (A2),

$$\begin{aligned} &P_{\eta'}^{(2m,m)}(K^{-1}\partial^2 K)|[\mu] j\{\nu\}(h)\eta\rangle \\ &\quad = K^{-1}\sum_{k\bar{\nu}} P_{\eta'}^{(2m,m)}(\partial^2)|[\mu]k\{\bar{\nu}\}(h)\eta\rangle \\ &\qquad \times \langle [\mu]k\{\bar{\nu}\}(h)|K|[\mu] j\{\nu\}(h)\rangle \\ &\quad = 0, \quad \text{for } 3m > \nu - h. \end{aligned} \quad (\mathrm{A3})$$

Now, for the last equation to be satisfied, we must require (A1) since, in general,

$$P_{\eta'}^{(2m,m)}(\partial^2)|[\mu]k\{\bar{\nu}\}(h)\eta\rangle \neq 0 \text{ for } \bar{\nu} > \nu \text{ and } 3m < \bar{\nu} - h.$$

Note that Eq. (3.17a) is a special case of (A1).

Equation (A1) implies that $K^\dagger$ is upper triangular in the (h) blocks; i.e.,

$$\langle [\mu] i\{\nu'\}(h)|K^\dagger|[\mu] j\{\nu\}(h)\rangle = 0 \quad \text{for } \nu' < \nu. \quad (\mathrm{A4})$$

The triangular structure of K and $K^\dagger$ furthermore implies that

$$\langle [\mu] i\{\nu'\}(h)|K^{-1}|[\mu] j\{\nu\}(h)\rangle = 0, \quad \text{for } \nu' > \nu \quad (\mathrm{A5})$$

and

$$\langle [\mu] i\{\nu'\}(h)|(K^{-1})^\dagger|[\mu] j\{\nu\}(h)\rangle = 0 \quad \text{for } \nu' < \nu. \quad (\mathrm{A6})$$

It is convenient to require that the restriction (3.17b) be Hermitian. This requirement and the triangular properties of K imply

$$\begin{aligned} &\langle [\mu] j\{\nu\}(\nu)|K^\dagger|[\mu]\{\nu'\}(\nu)\rangle^* \\ &\quad = \langle [\mu] i\{\nu'\}(\nu)|K|[\mu] j\{\nu\}(\nu)\rangle \quad (\mathrm{A7a}) \\ &\quad = \delta_{\nu\nu'}\mathscr{K}(\nu)_{ij}, \quad (\mathrm{A7b}) \end{aligned}$$

$$\begin{aligned} &\langle [\mu] j\{\nu\}(\nu)|(K^{-1})^\dagger|[\mu] i\{\nu'\}(\nu)\rangle^* \\ &\quad = \langle [\mu] i\{\nu'\}(\nu)|K^{-1}|[\mu] j\{\nu\}(\nu)\rangle \quad (\mathrm{A7c}) \\ &\quad = \delta_{\nu\nu'}\mathscr{K}(\nu)_{ij}^{-1}. \quad (\mathrm{A7d}) \end{aligned}$$

Now, using definition (3.13a),

$$\gamma(t^{(\alpha_1,\alpha_2)}) = K^{-1}\Gamma(t^{(\alpha_1,\alpha_2)})K \quad (\mathrm{A8a})$$

for the K transformation, the Hermiticity requirement (3.13b)

$$\gamma(t^{(\alpha_1,\alpha_2)}) = K^\dagger[\Gamma(t^{(-\alpha_2,-\alpha_1)})]^\dagger(K^{-1})^\dagger \quad (\mathrm{A8b})$$

the triangular properties of the K matrices, the Hermiticity properties of the restriction $\mathscr{K}(\nu)$ and its inverse, and the z^2 independence of the su(3) highest-weight states (3.16a), it is easy to prove Eqs. (3.19) and (3.20).

Note that Eq. (A8a) turns out to be the most useful one for tensors $t^{(\alpha_1,\alpha_2)} \subset g_2$ with positive u(2) rank(α_1,α_2) while (A8b) is the equation one should use when this rank is negative. Note also that the various expansions $[\Gamma(t^{(\alpha_1\alpha_2)})]^\dagger$ do not belong to the VCS expansions (3.10) and, therefore, we are not guaranteed that application of a given Bargmann operator $[\Gamma(t^{(\alpha_1\alpha_2)})]^\dagger$ to the VCS (physical) space will leave it invariant. This precludes an arbitrary truncation scheme for a singular restriction $\mathscr{K}(\nu)$ when one uses Eq. (A8b); the correct procedure then is to restrict consideration to the space of eigenvectors of $\mathscr{K}(\nu)$ with corresponding nonvanishing eigenvalues.

To find the recursion formulas for $\mathscr{K}^2$, we start with the equality

$$\Gamma(t^{(-\alpha_2,-\alpha_1)})K = K\gamma(t^{(-\alpha_2,-\alpha_1)}), \quad (\mathrm{A9})$$

easily obtained from (A8a) and where, from now on, the partition $(\alpha_1\alpha_2)$ is strictly positive. Taking matrix elements of Eq. (A9) between the basis of su(3) highest-weight states (3.16), we have that

$$\begin{aligned} &\langle [\mu] i\{\nu'\}(\nu')\|\Gamma(t^{(-\alpha_2,-\alpha_1)})\mathscr{K}(\nu)\|[\mu] j\{\nu\}(\nu)\rangle \\ &\quad = \langle [\mu] i\{\nu'\}(\nu')\|\mathscr{K}(\nu')\gamma(t^{(-\alpha_2,-\alpha_1)})\|[\mu] j\{\nu\}(\nu)\rangle \\ &\qquad + \sum_{l,\bar{\nu}>\nu'} \langle [\mu] i\{\nu'\}(\nu')|K|[\mu] l\{\bar{\nu}\}(\nu')\rangle \\ &\qquad \times \langle [\mu] l\{\bar{\nu}\}(\nu')\|\gamma(t^{(-\alpha_2,-\alpha_1)})\|[\mu] j\{\nu\}(\nu)\rangle. \end{aligned} \quad (\mathrm{A10})$$

From

$$\Gamma(g^{(-1,-2)})K = K\gamma(g^{(-1,-2)}), \quad (\mathrm{A11})$$

we have

$$\begin{aligned} &\langle [\mu] i\{\nu'\}(\nu')\|\Gamma(g^{(-1,-2)})\mathscr{K}(\bar{\nu})\|[\mu] l\{\bar{\nu}\}(\bar{\nu})\rangle \\ &\quad = \langle [\mu] i\{\nu'\}(\nu')|K|[\mu] l\{\bar{\nu}\}(\nu')\rangle \\ &\qquad \times \langle [\mu] l\{\bar{\nu}\}(\nu')\|\gamma(g^{(-1,-2)})\|[\mu] l\{\bar{\nu}\}(\bar{\nu})\rangle, \end{aligned} \quad (\mathrm{A12})$$

where we have used the fact that $\gamma(t)$, $t \in \mathrm{su}(3)$, is diagonal in the su(3) label $\{\nu\}$ and in the multiplicity label l. Thus

$$\begin{aligned} &\langle [\mu] i\{\nu'\}(\nu')|K|[\mu] l\{\bar{\nu}\}(\nu')\rangle \\ &\quad = \frac{\langle [\mu] i\{\nu'\}(\nu')\|\Gamma(g^{(-1,-2)})\mathscr{K}(\bar{\nu})\|[\mu] l\{\bar{\nu}\}(\bar{\nu})\rangle}{\langle [\mu] l\{\bar{\nu}\}(\nu')\|\gamma(g^{(-1,-2)})\|[\mu] l\{\bar{\nu}\}(\bar{\nu})\rangle}. \end{aligned} \quad (\mathrm{A13})$$

Inserting (A13) into (A10), we obtain

$$\langle [\mu] i\{v'\}(v') \| \mathscr{K}(v') \gamma(t^{(-\alpha_2, -\alpha_1)}) \| [\mu] j\{v\}(v) \rangle$$
$$= \langle [\mu] i\{v'\}(v') \| \Gamma(t^{(-\alpha_2, -\alpha_1)}) \mathscr{K}(v) \| [\mu] j\{v\}(v) \rangle - \sum_{l, \bar{v} > v'} \frac{\langle [\mu] i\{v'\}(v') \| \Gamma(g^{(-1, -2)}) \mathscr{K}(\bar{v}) \| [\mu] l\{\bar{v}\}(\bar{v}) \rangle}{\langle [\mu] l\{\bar{v}\}(v') \| \gamma(g^{(-1, -2)}) \| [\mu] l\{\bar{v}\}(\bar{v}) \rangle}$$
$$\times \langle [\mu] l\{\bar{v}\}(v') \| \gamma(t^{(-\alpha_2, -\alpha_1)}) \| [\mu] j\{v\}(v) \rangle . \tag{A14}$$

Multiplying Eq. (A14) on the right by $\mathscr{K}(v)$ and using Eqs. (A7) and (A8), we obtain

$$\langle [\mu] i\{v'\}(v') \| \mathscr{K}(v')^2 \Gamma^\dagger(t^{(\alpha_1 \alpha_2)}) \| [\mu] j\{v\}(v) \rangle$$
$$= \langle [\mu] i\{v'\}(v') \| \Gamma(t^{(-\alpha_2, -\alpha_1)}) \mathscr{K}(v)^2 \| [\mu] j\{v\}(v) \rangle - \sum_{l, \bar{v} > v'} \frac{\langle [\mu] i\{v'\}(v') \| \Gamma(g^{(-1, -2)}) \mathscr{K}(\bar{v}) \| [\mu] l\{\bar{v}\}(\bar{v}) \rangle}{\langle [\mu] l\{\bar{v}\}(v') \| \gamma(g^{(-1, -2)}) \| [\mu] l\{\bar{v}\}(\bar{v}) \rangle}$$
$$\times \langle [\mu] l\{\bar{v}\}(v') \| K^{-1} \Gamma(t^{(-\alpha_2, -\alpha_1)}) \mathscr{K}(v)^2 \| [\mu] j\{v\}(v) \rangle. \tag{A15}$$

From

$$\gamma(g^{(-1, -2)}) K^\dagger = K^\dagger \Gamma^\dagger(g^{(21)}) = K^\dagger z^2, \tag{A16}$$

we derive

$$\langle [\mu] i\{\bar{v}\}(h) | K^\dagger | [\mu] l\{\bar{v}\}(h) \rangle$$
$$= \delta_{\bar{v}\bar{v}} \langle [\mu] i\{\bar{v}\}(h) \| \gamma(g^{(-1, -2)}) \| [\mu] j\{\bar{v}\}(\bar{v}) \rangle$$
$$\times \mathscr{K}(\bar{v})_{jl}, \quad (\bar{v}), \ (\bar{v}) \in (h) \times (21), \tag{A17}$$

which implies

$$\langle [\mu] i\{\bar{v}\}(h) | K | [\mu] l\{\bar{v}\}(h) \rangle$$
$$= \delta_{\bar{v}\bar{v}} \langle [\mu] i\{\bar{v}\}(h) \| \gamma(g^{(-1, -2)}) \|$$
$$\times [\mu] j\{\bar{v}\}(\bar{v}) \rangle^* \mathscr{K}(\bar{v})_{jl}. \tag{A18}$$

From (A5), (A17), and (A18), we derive that

$$\sum_m \mathscr{K}(\bar{v})_{lm} \langle [\mu] m\{\bar{v}\}(h) | K^{-1} | [\mu] n\{\bar{v}\}(h) \rangle$$
$$= \frac{\delta_{ln} \delta_{\bar{v}\bar{v}}}{\langle [\mu] l\{\bar{v}\}(h) \| \gamma(g^{(-1, -2)}) \| [\mu] l\{\bar{v}\}(\bar{v}) \rangle^*}. \tag{A19}$$

Finally, upon substitution of (A19) in (A15), we find

$$\langle [\mu] i\{v'\}(v') \| \mathscr{K}(v')^2 \Gamma^\dagger(t^{(\alpha_1 \alpha_2)}) \| [\mu] j\{v\}(v) \rangle$$
$$= \langle [\mu] i\{v'\}(v') \| \Gamma(t^{(-\alpha_2, -\alpha_1)}) \mathscr{K}(v)^2 \| [\mu] j\{v\}(v) \rangle$$
$$- \sum_{l, \bar{v} > v'} \frac{\langle [\mu] i\{v'\}(v') \| \Gamma(g^{(-1, -2)}) \| [\mu] l\{\bar{v}\}(\bar{v}) \rangle}{\langle [\mu] l\{\bar{v}\}(v') \| \Gamma(g^{(-1, -2)}) \| [\mu] l\{\bar{v}\}(\bar{v}) \rangle}$$
$$\times \langle [\mu] l\{\bar{v}\}(v') \| \Gamma(t^{(-\alpha_2, -\alpha_1)}) \mathscr{K}(v)^2 \|$$
$$\times [\mu] j\{v\}(v) \rangle, \tag{A20}$$

where we have used the equality

$$\langle [\mu] l\{\bar{v}\}(v') \| \Gamma(g^{(-1, -2)}) \| [\mu] l\{\bar{v}\}(\bar{v}) \rangle$$
$$= | \langle [\mu] l\{\bar{v}\}(v') \| \gamma(g^{(-1, -2)}) \| [\mu] l\{\bar{v}\}(\bar{v}) \rangle |^2, \tag{A21}$$

which is easily verified by direct computation. Equation (A20) yields Eqs. (3.21a) and (3.21b) upon substituting (11) and (10), respectively, for $(\alpha_1 \alpha_2)$. It should be compared to Eqs. (3.52) and (3.60) of Rowe *et al.*[3] that were derived for the special case of nilpotent raising subalgebra of order 2. The final equations are in fact identical and this is indicative that Eq. (A20) has a wide range of applicability as we now discuss.

We conjecture that the derivation of the recursion formula (A20) for the restrictions $\mathscr{K}(v)$ is general and is not restricted to $g_2 \supset su(3)$. The order of nilpotency n of a given Lie algebra chain is arbitrary as can easily be seen from the fact that the above derivation is independent of the number of ranks (α) of the various $\mathbf{h}$ tensors in $\mathbf{g}$ mod $\mathbf{l}$. In fact, the formula should apply to every Lie algebra chain $\mathbf{g} \supset \mathbf{l} \supset \mathbf{h}$, where $\mathbf{g}$ is the Lie algebra and $\mathbf{h}$ is the stability subalgebra of the highest-weight representation, and $\mathbf{l}$ is the intermediate subalgebra [here su(3)] such that $\mathbf{l} = \mathbf{h} + \mathbf{n}_+^m + \mathbf{n}_-^m$. When $m = 1$, the raising subalgebra is Abelian, the second term in the right-hand side of Eq. (A20) vanishes and the evaluation of the various matrix elements in the recursion formula is then facilitated[2,3] by the use of commutator methods and the introduction of $\mathbf{h}$ invariants.[2]

[1] J. Deenen and C. Quesne, J. Math. Phys. **25**, 2354 (1984); **26**, 2705 (1985); C. Quesne, *ibid.* **27**, 428, 869 (1986).

[2] D. J. Rowe, J. Math. Phys. **25**, 2662 (1984); D. J. Rowe, G. Rosensteel, and R. Carr, J. Phys. A: Math. Gen. **17**, L399 (1984); D. J. Rowe, "Coherent states, contractions and classical limits of the non-compact symplectic groups," in *Proceedings of the XIII International Colloquium on Group Theoretical Methods in Physics*, edited by W. W. Zachary, College Park, Maryland (World Scientific, Singapore, 1984); D. J. Rowe, in *Phase Space Approach to Nuclear Dynamics*, edited by M di Toro, W. Nörenberg, M. Rosina, and S. Stringari (World Scientific, Singapore, 1986), p. 546.

[3] D. J. Rowe, R. Le Blanc, and K. T. Hecht, preprint, 1987.

[4] R. Le Blanc and D. J. Rowe, J. Math. Phys. **29**, 758 (1988).

[5] R. C. King and A. H. A. Qubanchi, J. Phys. A: Math. Gen. **11**, 1491 (1978).

[6] M. Perroud, J. Math. Phys. **17**, 1998 (1976).

[7] G. Racah, Lectures notes on *Group Theory and Spectroscopy* (Institute of Advanced Study, Princeton, NJ, 1951).

[8] D. J. Rowe, G. Rosensteel, and R. Gilmore, J. Math. Phys. **26**, 2787 (1985).

[9] K. T. Hecht, R. Le Blanc, and D. J. Rowe, J. Phys. A: Math. Gen. **20**, 2241 (1988); R. Le Blanc and K. T. Hecht, *ibid.* **20**, 4613 (1988); L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. **8**, 89 (1968); K. T. Hecht, Nucl. Phys. **62**, 1 (1965).

[10] B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).

The Demazure–Tits subgroup of a simple Lie group

L. Michel, J. Patera,[a] and R. T. Sharp[b]
Institut des Hautes Études Scientifiques, Bures-sur-Yvette, France

(Received 5 February 1987; accepted for publication 19 August 1987)

The Demazure–Tits subgroup of a simple Lie group **G** is the group of invariance of Clebsch–Gordan coefficients tables (assuming an appropriate choice of basis). The structure of the Demazure–Tits subgroups of A_n, B_n, C_n, D_n, and G_2 is described. Orbits of the permutation action of the DT group in any irreducible finite-dimensional representation space of A_2, C_2, and G_2 are decomposed into the sum of irreducible representations of the DT group.

I. INTRODUCTION

The purpose of this paper is to study a certain finite subgroup of any simple compact Lie group **G**. We call the subgroup the Demazure[1]–Tits[2] group and denoted it by **DT** or **DT**(**G**).

The maximal tori (called the Cartan subgroups) of a compact semisimple Lie group **G** are all conjugate. They are isomorphic to $\mathrm{U}(1)^l$, where l is the rank of **G**. The centralizer $\mathbf{C}_G(g)$ of g in **G** contains a Cartan subgroup; the elements $g \in \mathbf{G}$, whose centralizer is exactly a Cartan subgroup, are called regular. They form an open dense set in **G**.

Given a Cartan subgroup $\mathbf{H} \subset \mathbf{G}$, one considers its normalizer $\mathbf{N}_G(\mathbf{H})$ (the largest **G** subgroup containing **H** as an invariant subgroup). The quotient $\mathbf{N}_G(\mathbf{H})/\mathbf{H} = \mathbf{W}(\mathbf{G})$ is the Weyl group of **G**. This is a finite group with a natural action on the Cartan subalgebra **h** (the Lie algebra of **H**) of **G** generated by reflections along the simple roots. The importance of the Weyl group in the theory of Lie algebras, Lie groups, and their representations is well recognized. However, the exact sequence

$$1 \to \mathrm{U}(1)^l \to \mathbf{N}_G \mathrm{U}(1)^l \xrightarrow{\vartheta'} \mathbf{W}(\mathbf{G}) \to 1\,, \tag{1.1}$$

in general does not split, so **W** is not a subgroup of **G**, where **G** is simply connected compact. Among the finite subgroups of the normalizer $\mathbf{N}_G(\mathrm{U}(1)^l)$ that are mapped by ϑ onto **W** there is a natural one $\mathbf{DT}(G)$, defined by (2.15) below, that has been first pointed out by Demazure[1] and Tits.[2] Its intersection with $\mathrm{U}(1)^l$ is the group of square roots of 1, hence it is the extension

$$1 \to Z_2^l \to \mathbf{DT}(\mathbf{G}) \xrightarrow{\vartheta} \mathbf{W}(\mathbf{G}) \to 1\,, \tag{1.2}$$

which is naturally deduced from (1.1).

Physicists' interest in the Demazure–Tits group **DT**(**G**) is most likely to originate either from the similarity of its action in representation space to the action of the Weyl group in weight space, or from the fact that it permutes (with some changes of sign) the physical states of a **G**-irreducible space, thus making it possible to keep the same states even without the full Lie group symmetry. It is a finite subgroup of **G** that preserves the root space decomposition (Cartan decomposition) of the Lie algebra of **G**. The group **DT**(**G**) has occasionally appeared in mathematics literature; however, recognition of its usefulness in applied problems relevant to physics is quite recent (cf. Ref. 3, where the group **DT** is denoted by **N**). A systematic use of **DT**(**G**) has been made as the group of invariance of table of the Clebsch–Gordan coefficients (relative to an appropriate basis choice). In computing Clebsch–Gordan coefficients for $\mathbf{G} = \mathrm{SU}(5), \mathrm{O}(10)$, and E_6 (cf. Refs. 4–6) **DT** was used as a group of transformations among CGC of the same values. Practically it allows a small fraction of nonzero CGC to represent all.

In this article we give in Sec. II the structure of **DT**(**G**) for the classical groups A_l, B_l, C_l, D_l, and for G_2. Section III contains some examples of the **DT** group in lowest representations. In general, it is very interesting to decompose an irreducible **G**-representation space V_Λ (Λ is the highest weight) into a direct sum of subspaces irreducible with respect to **DT**(**G**). For groups **G** of rank $l = 2$ we describe **DT**(**G**) in detail in Secs. IV–VI. Namely, we find its character table, decompose any V_Λ into **DT**-invariant subspaces, and identify each **DT**-conjugacy class as a **G** class of elements of finite order (Sec. VII). The last step opens the possibility of using the powerful computing methods[7–10] with elements of finite order in **G** for the study of conjugacy classes of **DT** in all representations of **G**. The simple Lie group **G** in this article is always the simply connected one. Section VIII contains a summary of our results and some open problems. The Appendix contains a summation formula, which, as far as we know, does not appear in literature.

We denote a group (finite or continuous) by bold capital letters; for a Lie algebra we use lowercase bold symbols except for groups or algebras of specific types like A_2 or SU(3), etc. The symbols **W**(**g**) and **W**(**G**), **DT**(**G**) and **DT**(**g**), etc., where **g** is the Lie algebra of **G**, are used as synonyms.

II. THE STRUCTURE OF THE DEMAZURE–TITS SUBGROUPS OF THE SIMPLE SIMPLY CONNECTED LIE GROUPS

We denote by (λ,μ) the Cartan–Killing positive definite scalar product on the compact semisimple Lie algebra **g**, and let the roots be $\alpha_i \in \Delta$, its root system in a chosen Cartan subalgebra **h**; Δ is the root system of **g**. If l is the rank of **g**

[a] Permanent address: Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec H3C 3J7, Canada.
[b] Permanent address: Department of Physics, McGill University, Montreal, Quebec H3A 2T8, Canada.

 0022-2488/88/040777-20$02.50

then the Weyl group $\mathbf{W}(\mathbf{g})$ is generated by the reflections r_i, $i = 1,...,l$, along the simple roots α_i,

$$r_i\lambda = \lambda - 2(\alpha_i,\lambda)(\alpha_i,\alpha_i)^{-1}\alpha_i \,. \tag{2.1}$$

When λ itself is a simple root, say α_i,

$$r_j\alpha_i = \alpha_i - \alpha_j A_{ij} \,, \tag{2.2}$$

where

$$A_{ij} = 2(\alpha_i,\alpha_j)(\alpha_j,\alpha_j)^{-1} \tag{2.3}$$

are the matrix elements of the Cartan matrix of $\mathbf{g}$.

We denote by $l = \dim \mathbf{h}$ the rank of $\mathbf{g}$. Let $\{r_i, 1 \leqslant i \leqslant l\}$ be a minimal set of generators of $\mathbf{W}(\mathbf{g})$ (the corresponding simple roots α_i form a base of $\mathbf{h}$); this group is completely characterized by the relations

$$1 \leqslant i,j \leqslant l, \quad (r_ir_j)^{m_{ij}} = I, \quad m_{ii} = 1, \quad 2 \leqslant m_{ij} = m_{ji} \leqslant 6 \,. \tag{2.4}$$

Note that $r_ir_j = r_jr_i$ when $m_{ij} = 2$. The list of possible values of m_{ij} was given by Coxeter and is summarized in the Coxeter–Dynkin diagram of $\mathbf{g}$. Namely, $m_{ij} = (1 - \theta_{ij}/\pi)^{-1}$, where θ_{ij} is the angle between α_i and α_j; it is 2, 3, 4, or 6 according to whether there are zero, one, two, or three lines joining vertices i and j. To specify the structure of $\mathbf{W}(\mathbf{g})$, we define first a family of matrix groups (see, e.g., Ref. 11).

A. The groups G(m, p, n)

Let m, p, n be integers with p dividing m; we denote by $\mathbf{A}(m,p,n)$ the group of diagonal $n \times n$ unitary matrices a that satisfy the relations

$$(a_{ii})^m = 1, \quad 1 \leqslant i \leqslant n, \quad \det(a)^{m/p} = 1 \,. \tag{2.5}$$

Let $\mathbf{\Pi}_n$ be the group of $n \times n$ permutation matrices; they have one 1 in each row and each column and zeros elsewhere. It is a faithful representation of $\mathbf{S}_n$, the group of permutations of n objects. The determinant of a permutation matrix is ± 1 according to the parity of the permutation. We denote by $\mathbf{G}(m,p,n)$ the matrix group generated by the groups $\mathbf{A}(m,p,n)$ and $\mathbf{\Pi}_n$. Obviously, $\mathbf{G}(m,p,n)$ is the semidirect product,

$$\mathbf{G}(m,p,n) = \mathbf{A}(m,p,n) \rtimes \mathbf{\Pi}_n \,. \tag{2.6}$$

All the matrix groups $\mathbf{G}(m,p,n)$, except $\mathbf{G}(1,1,n) = \mathbf{\Pi}_n$ and $\mathbf{G}(2,2,2)$ are irreducible over $\mathbb{C}$. The only pair of conjugate groups is $\mathbf{G}(4,4,2)$ and $\mathbf{G}(2,1,2)$. For a finite group $\mathbf{G}$, we denote by $|\mathbf{G}|$ the number of its elements. Then

$$|\mathbf{G}(m,p,n)| = m^n p^{-1} n! \,. \tag{2.7}$$

The linear action of the Weyl group $\mathbf{W}(\mathbf{g})$ on the Cartan subalgebra $\mathbf{h}$ is represented by

$$\mathbf{W}(A_l) = \mathbf{G}(1,1,l+1), \quad \mathbf{W}(B_l) = \mathbf{W}(C_l) = \mathbf{G}(2,1,l) \,,$$
$$\mathbf{W}(D_l) = \mathbf{G}(2,2,l), \quad \mathbf{W}(G_2) = \mathbf{G}(6,6,2) \,. \tag{2.8}$$

Exceptionally, for $A_l \sim \mathrm{SU}_{l+1}$, we have used the Cartan algebra of U_{l+1}; in it the Cartan algebra of A_l is the hyperplane orthogonal to a vector with all coordinates equal.

For a matrix group $\mathbf{G}$ we denote by $\mathbf{SG}$, or sometimes by $\mathbf{G}^+$, its unimodular subgroup (i.e., the group of matrices with determinant 1). Note the isomorphism,

$$\mathbf{SG}(2,1,3) = \mathbf{W}(B_3)^+ \sim \mathbf{S}_4 \,. \tag{2.9}$$

We recall now, at least in a particular case, the definition of the *wreath product*: given a group $\mathbf{K}$, the wreath product by $\mathbf{S}_n$, which we denote by $K \uparrow n$, is the semidirect product

$$\mathbf{K}\uparrow n = \mathbf{K}^n \rtimes \mathbf{S}_n \tag{2.10}$$

of $\mathbf{S}_n$ by n copies of $\mathbf{K}$, $\mathbf{S}_n$ acting by permutations on the n factors of $\mathbf{K}^n$. For a finite group $\mathbf{K}$,

$$|\mathbf{K}\uparrow n| = |\mathbf{K}|^n n! \,. \tag{2.11}$$

Let us point out that

$$\mathbf{G}(m,1,n) \sim \mathbf{Z}_m \uparrow n; \quad \text{e.g.,} \ \mathbf{W}(B_l) \sim \mathbf{Z}_2 \uparrow l \,. \tag{2.12}$$

We will need the following properties of Weyl groups. The Lie algebras of types B_l and C_l have roots of two different lengths; the corresponding reflections form two conjugacy classes in $\mathbf{W}(B_l) = \mathbf{W}(C_l)$ with, respectively, l and $l(l-1)$ elements. The elements of the conjugacy class with l elements are the reflections of $\mathbf{A}(m,1,l)$. They commute and generate the Abelian group $\mathbf{A}(m,1,l)$. Here $\mathbf{W}(D_l)$ is an index 2 subgroup of $\mathbf{W}(B_l)$; when l is odd, $-I \notin \mathbf{W}(D_l)$. That is,

$$\mathbf{W}(B_l) = \mathbf{W}(D_l) \times \mathbf{Z}_2(-I), \quad \text{for } l \text{ odd} \,. \tag{2.13}$$

While the Weyl group $\mathbf{W}(\mathbf{g})$ is the same for all groups $\mathbf{G}$ that have the same Lie algebra $\mathbf{g}$, the Demazure–Tits group $\mathbf{DT}(G)$ does depend on the choice of $\mathbf{G}$; here we consider only simple simply connected compact Lie groups $\mathbf{G}$. We use the notation

$$\mathrm{prod}(n,x,y) = xyxy\cdots \,, \tag{2.14}$$

for a product of n factors, alternately x and y. Tits[2] defines $\mathbf{DT}(\mathbf{G})$ by its generators q_i and their relations

$$1 \leqslant i \leqslant l, \quad q_i^4 = 1, \quad q_i^2 q_j^2 = q_j^2 q_i^2 \,, \tag{2.15a}$$

$$\mathrm{prod}(m_{ij},q_i,q_j) = \mathrm{prod}(m_{ij},q_j,q_i), \qquad q_i q_j^2 q_i^{-1} = q_j^2 q_i^{2A_{ij}} \,. \tag{2.15b}$$

The q_i^2 are the square roots of 1 in the Cartan subgroup, they generate the kernel of ϑ in Eq. (1.2). The presence of the exponent $2A_{ij}$ in (2.15b) implies that $\mathbf{DT}(B_l)$ and $\mathbf{DT}(C_l)$ are different although $\mathbf{W}(B_l) = \mathbf{W}(C_l)$. Since we will use these relations often we give them more explicitly:

$$q_i^4 = 1, \quad q_i^2 q_j^2 = q_j^2 q_i^2 \,, \tag{E1}$$

$$m_{ij} = 2\colon\ q_i q_j = q_j q_i \,, \tag{E2}$$

$$m_{ij} = 3\colon\ q_i q_j q_i = q_j q_i q_j, \quad q_i q_j^2 = q_j^2 q_i^{-1} \,, \tag{E3}$$

$$m_{ij} = 2k\colon\ (q_i q_j)^k = (q_j q_i)^k, \qquad q_i q_j^2 q_i^{-1} = q_j^2 q_i^{2A_{ij}} \,. \tag{E4}$$

Consider two semisimple Lie groups $\mathbf{G}$ and $\mathbf{G}'$ both of rank l. If the Coxeter–Dynkin diagram of $\mathbf{G}$ is a subdiagram of the extended Coxeter–Dynkin diagram of $\mathbf{G}'$, then one has for the corresponding $\mathbf{DT}$ groups,

$$\mathbf{DT}(\mathbf{G}) \subset \mathbf{DT}(\mathbf{G}') \,. \tag{2.16}$$

Clearly $\mathbf{G}$ and $\mathbf{G}'$ have the same Cartan subgroup $\sim U_1^l$ and $\mathbf{N}_G(U_1^l) \subset \mathbf{N}_{G'}(U_1^l)$. Since the corresponding $\mathbf{DT}$ groups have the same kernel $\mathbf{Z}_2^l$, (2.16) holds. If the rank of $\mathbf{G}'$ is lower than l, (2.16) still holds provided the Coxeter–Dynkin diagram of $\mathbf{G}'$ is a subdiagram of the (nonextended) diagram of $\mathbf{G}$.

Let $\mathbf{C(G)}$ be the center of $\mathbf{G}$. The intersection $\mathbf{C(G)} \cap \mathbf{DT(G)}$ is the group of square roots of $\mathbf{C(G)}$. We recall the nature of $\mathbf{C(G)}$ in Table I.

B. The DT subgroup of A_l

In the natural $(l+1)$-dimensional representation of SU_{l+1}, a Cartan subgroup is represented by diagonal matrices; its subgroup of square roots of the unit is $\mathbf{A}(2,2,l+1) \sim \mathbf{Z}_2^l$. The Weyl group $\sim \mathbf{S}_{l+1}$ permutes the elements of these diagonal matrices; it can be represented by the group of permutation matrices $\mathbf{\Pi}_{l+1}$. The reflections correspond to permutations of two elements, the r_i corresponding to the permutations of neighboring elements. In $\mathbf{\Pi}_{l+1}$ their determinant is -1. The unimodular matrices that represent them in $\mathbf{DT}(\mathrm{SU}_{l+1})$ have been given in Ref. 3 (where they are denoted R_i). They are

$$a_i = I_{i-1} \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus I_{l-i}, \tag{2.17}$$

where I_k is the $k \times k$ unit matrix.

Let us introduce the $(l+1) \times (l+1)$ diagonal matrices:

$$v_1 = -1 \oplus I_l, \tag{2.18}$$

$$v_i = I_{i-1} \oplus -1 \oplus I_{l-i+1} = v_1 \prod_{k=1}^{i-1} a_k^2, \quad 2 \leqslant i \leqslant l+1. \tag{2.19}$$

They are the reflections of the group $\mathbf{A}(2,1,l+1)$ that they generate. For $1 \leqslant i \leqslant l$, the matrices $v_i a_i$ belong to Π_{l+1} and generate it since they represent the permutations $(i,i+1)$. Hence we have shown that v_1 and the a_i's generate $\mathbf{G}(2,1,l+1)$. Since $\det(a_i) = 1 = -\det(v_1)$, the a_i's generate the unimodular subgroup $\mathbf{SG}(2,1,l+1)$. This proves that

$$\begin{aligned} \mathbf{DT}(A_l) &= \mathbf{DT}(\mathrm{SU}_{l+1}) \\ &= \mathbf{SG}(2,1,l+1) \sim \mathbf{W}(B_{l+1})^+ . \end{aligned} \tag{2.20}$$

When l is even, $\det(-I_{l+1}) = -1$, so we obtain a unimodular representation $\mathbf{\Pi}_{l+1}$ of $\mathbf{S}_{l+1}$ by multiplying by -1 the matrices representing odd permutations; since $\mathbf{\Pi}_{l+1} \subset \mathbf{SG}(2,1,l+1)$, this shows that the exact sequence (1.2) splits for l even,

$$\mathbf{DT}(A_l) = \mathbf{Z}_2^l \rtimes \mathbf{W}(A_l) \quad (l \text{ even}). \tag{2.21}$$

This is not the case for odd l; e.g., for $l=1$, $\mathbf{DT}(A_1) = \mathbf{Z}_4$ (see also at the end of this section). When l is even, we can write explicitly a choice of representatives $\tilde{a}_i$ of the a_i's that realizes the splitting (2.21). We define the a_i's using the sets of indices

TABLE I. Structure of the center of a classical simple Lie group $\mathbf{G}$.

Algebra	A_l	B_l	C_l	D_l
$\mathbf{G}$	SU_{l+1}	Spin_{2l+1}	Sp_{2l}	Spin_{2l}
$\mathbf{C(G)}$	Z_{l+1}	Z_2	Z_2	$Z_2 \times Z_2$ (l even) Z_4 (l odd)

$$\mathbf{F}(i,l) = \{k, (0 < k \text{ odd} < i) \cup (i \leqslant k \text{ even} \leqslant l)\}, \tag{2.22a}$$

$$\tilde{a}_i = a_i \prod_{k \in F(i,l)} a_k^2 . \tag{2.22b}$$

These $\tilde{a}_i$ generate a subgroup of $\mathbf{DT}(A_l)$ isomorphic to $\mathbf{W}(A_l) \sim \mathbf{S}_{l+1}$.

The center of A_l is the cyclic group $\mathbf{Z}_{l+1}$. When l is odd, the center has a nontrivial square root of unity that is in every Cartan subalgebra and therefore in $\mathbf{DT}(A_l)$. Indeed, the irreducible matrix group $\mathbf{SG}(2,1,l+1)$ has a nontrivial center $\mathbf{C}(\mathbf{SG}(2,1,l+1))$ only when it contains the $-I$ matrix, i.e., for odd l. Thus

$$\mathbf{C}(\mathbf{DT}(A_l)) = 1 \text{ or } Z_2(\alpha), \quad \text{for } l \text{ even or odd};$$

$$\alpha = \prod_{k \text{ odd}} a_k^2 . \tag{2.23}$$

C. The DT subgroup of C_l

Next we consider the $\mathbf{DT}$ of the symplectic group Sp_{2l}. We denote by c_i the generators of this group. The equations (E) applied to them become

$$\begin{aligned} &c_i^4 = 1, \quad c_i^2 c_j^2 = c_j^2 c_i^2, \quad c_i c_{i+1}{}^2 = c_{i+1}{}^2 c_i{}^{-1}, \\ &c_i c_{i+1} c_i = c_{i+1} c_i c_{i+1} \quad (1 \leqslant i \leqslant l-1), \\ &c_{l-1} c_l c_{l-1} c_l = c_l c_{l-1} c_l c_{l-1}, \\ &c_{l-1} c_l{}^2 = c_l{}^2 c_{l-1}{}^{-1}, \quad c_l c_{l-1}{}^2 = c_{l-1}{}^2 c_l . \end{aligned} \tag{2.24}$$

According to (2.16), for $1 \leqslant i \leqslant l-1$, the c_i's generate $\mathbf{DT}(A_{l-1}) \subset \mathbf{DT}(C_l)$. In order to complete our study of C_l, our strategy is to consider its l elements s_i, $1 \leqslant i \leqslant l$, "above" the l *commuting* reflections r_i generating $\mathbf{A}(2,1,l) \subset \mathbf{W}(C_l)$, i.e.,

$$\begin{aligned} &\theta(s_i) = r_i, \quad s_l = c_l, \quad 1 \leqslant i \leqslant l-1, \\ &s_i = u_i s_l u_i^{-1} \quad \text{with } u_i = \prod_{k=i}^{l-1} c_k . \end{aligned} \tag{2.25}$$

(In the Π symbol, when the factors do not commute, they always are assumed to be placed in order of increasing index value: $u_i = c_i c_{i+1} \cdots c_{l-2} c_{l-1}$.) We know that these reflections commute among themselves. We now prove the following lemma.

Lemma 1: The elements s_i commute among themselves.

We first verify it for s_{l-1} and s_l. Indeed from (2.24) and (2.25), we compute

$$\begin{aligned} s_{l-1} s_l &= c_{l-1} c_l c_{l-1}{}^{-1} c_l \\ &= c_{l-1} c_l c_{l-1} c_l c_{l-1}{}^2 \\ &= c_l c_{l-1} c_l c_{l-1}{}^{-1} = s_l s_{l-1} . \end{aligned} \tag{2.26}$$

Because c_i and c_j commute when $|i-j| > 1$, with $\tilde{u}_i = \Pi_{k=i}^{l-1} c_k$, we have

$$\begin{aligned} s_i s_l &= \tilde{u}_i s_{l-1} \tilde{u}_i^{-1} s_l \\ &= \tilde{u}_i s_{l-1} s_l \tilde{u}_i^{-1} \\ &= \tilde{u}_i s_l s_{l-1} \tilde{u}_i^{-1} \\ &= s_l \tilde{u}_i s_{l-1} \tilde{u}_i{}^{-1} = s_l s_i \quad (i \leqslant l-2) . \end{aligned} \tag{2.27}$$

We need the relation [use (2.24) twice]

$$s_i = c_{i+1} s_i c_{i+1}^{-1} \quad (1 \leqslant i \leqslant l-2) \tag{2.28}$$

to prove by recursion that s_i and s_{i+1} commute. It is true for $i = l-2$:

$$\begin{aligned} s_{l-2}s_{l-1} &= c_{l-1}s_{l-2}s_l c_{l-1}^{-1} \\ &= c_{l-1}s_l s_{l-2}c_{l-1}^{-1} \\ &= c_{l-1}s_l c_{l-1}^{-1}s_{l-2} = s_{l-1}s_{l-2}\,. \end{aligned} \tag{2.29}$$

Assuming that it is true for $i = k$, we prove it for $i = k-1$,

$$\begin{aligned} s_{k-1}s_k &= c_k s_{k-1} c_k^{-1} s_k \\ &= c_k c_{k-1} s_k c_{k-1}^{-1} s_{k+1} c_k^{-1} \\ &= c_k c_{k-1} s_k s_{k+1} c_{k-1}^{-1} c_k^{-1} \\ &= c_k c_{k-1} s_{k+1} s_k c_{k-1}^{-1} c_k^{-1} \\ &= c_k s_{k+1} c_{k-1} s_k c_{k-1}^{-1} c_k^{-1} \\ &= c_k s_{k+1} s_{k-1} c_k^{-1} \\ &= c_k s_{k+1} c_k^{-1} s_{k-1} = s_k s_{k-1}\,. \end{aligned} \tag{2.30}$$

Finally when $i \leqslant j-2$, we define as before $u = u_i u_{j-1}^{-1}$. Then

$$\begin{aligned} s_i s_j &= u s_{j-1} u^{-1} s_j \\ &= u s_{j-1} s_j u^{-1} = u s_j s_{j-1} u^{-1} \\ &= s_j u s_{j-1} u^{-1} = s_j s_i\,. \end{aligned} \tag{2.31}$$

Using (2.24), we find

$$s_i^{\,2} = \prod_{k=1}^{l} c_k^{\,2}\,, \tag{2.32}$$

and remark that all the squares are different. Similarly,

$$c_i^{\,2} = s_i^{\,2} s_{i+1}^{\,2}, \quad c_l^{\,2} = s_l^{\,2} \quad (1 \leqslant i \leqslant l-1)\,. \tag{2.33}$$

Hence the s_i commute also with the $c_i^{\,2}$. They generate an Abelian group containing the kernel in (1.2) of $\mathbf{DT}(C_l)$. Moreover, the commutation of the s_i's shows that the covering of $\mathbf{A}(2,1,l) \subset \mathbf{W}(C_l)$ in $\mathbf{DT}(C_l)$ is

$$\vartheta^{-1}(\mathbf{A}(2,1,l)) = \mathbf{Z}_4^{\,l}\,. \tag{2.34}$$

When $1 \leqslant i \leqslant l-1$, we choose other representatives $\tilde{c}_i$ of the r_i's,

$$\begin{aligned} &\vartheta(\tilde{c}_i) = \vartheta(c_i) = r_i, \\ &\tilde{c}_i = s_i^{\,2} c_i = c_i s_{i+1}^{\,2} \quad (1 < i \leqslant l-1)\,, \end{aligned} \tag{2.35}$$

where the last equality is obtained by a repeated use of Eqs. (2.24). We verify that

$$\tilde{c}_i^{\,2} = 1, \quad 1 \leqslant i \leqslant l-2, \quad (\tilde{c}_i \tilde{c}_{i+1})^3 = 1 \quad (1 \leqslant i \leqslant l-1)\,. \tag{2.36}$$

This shows that $\mathbf{DT}(C_l)$ contains a subgroup isomorphic to $\mathbf{W}(A_{l-1}) \sim \mathbf{S}_l$. We verify that it acts on the s_i by permutations

$$\begin{aligned} &\tilde{c}_i s_{i+1} \tilde{c}_i^{-1} = s_i, \quad \tilde{c}_i s_i \tilde{c}_i^{-1} = s_{i+1}\,, \\ &\tilde{c}_i s_j \tilde{c}_i^{-1} = s_j \quad (i<j \text{ or } i>j+1)\,. \end{aligned} \tag{2.37}$$

This completes the proof of the isomorphism

$$\mathbf{DT}(C_l) \sim Z_4 \uparrow l \sim \mathbf{G}(4,1,l). \tag{2.38}$$

The center, $\mathbf{C}(\mathbf{DT}(C_l)) = \mathbf{Z}_4(s)$, of this group is the diagonal subgroup of $\mathbf{Z}_4^{\,l}$. It is generated by

$$s = \prod_{k=1}^{l} s_i\,. \tag{2.39}$$

Observe that

$$\mathbf{C}(\mathrm{Sp}_{2l}) \cap \mathbf{C}(\mathbf{DT}(C_l)) = Z_2(s^2)\,, \tag{2.40}$$

where α has been defined in (2.23),

$$s^2 = \prod_{k \text{ odd}} c_k^{\,2} = \alpha\,. \tag{2.41}$$

The matrices representing c_i's in the $2l$-dimensional faithful representation of the symplectic group C_l are shown in Sec. III. All equations of this section can be thus verified.

D. The DT subgroup of B_l

Let us now consider the **DT** of Spin_{2l+1}. We denote by b_i its generators. For $1 \leqslant i \leqslant l-1$, like the c_i, these satisfy (2.24) and (E1). But the last line of Eq. (2.24) is replaced by

$$\begin{aligned} &b_{l-1} b_l b_{l-1} b_l = b_l b_{l-1} b_l b_{l-1}\,, \\ &b_l b_{l-1}^{\,2} b_l = b_{l-1}^{\,2} b_l^{-1}, \quad b_{l-1} b_l^{\,2} = b_l^{\,2} b_{l-1}\,, \end{aligned} \tag{2.42}$$

and $m_{ij} = 2$ when $|i-j| > 1$, so (E2) applies

$$b_i b_j = b_j b_i \quad (|i-j| > 1). \tag{2.43}$$

From these equations we obtain

$$\mathbf{Z}_2(\eta) \subseteq \mathbf{C}(\mathbf{DT}(B_l)), \quad \eta = b_l^{\,2}\,. \tag{2.44}$$

Here $\mathbf{Z}_2(\eta)$ denotes the $\mathbf{Z}_2$ group generated by η. The group $\mathbf{Z}_2(\eta)$ is exactly $\mathbf{C}(\mathrm{Spin}_{2l+1})$. As we will see later, $\mathbf{C}(\mathbf{DT}(B_l))$ might be larger.

Since $\mathbf{W}(B_l) = \mathbf{W}(C_l)$, we follow the same strategy as for the study of $\mathbf{DT}(C_l)$: we introduce the representatives t_i of the $l-1$ reflections conjugate to b_l,

$$t_l = b_l, \quad t_i = b_i t_{i+1} b_i^{-1} = u_i b_l u_i^{-1} \quad (1 \leqslant i \leqslant l), \tag{2.45}$$

where the u_i are defined as in (2.25). This time we find that the t_i's all have the same square,

$$t_i^{\,2} = \eta\,, \quad \eta^2 = 1\,, \tag{2.46}$$

and, instead of commuting among themselves, we demonstrate that they "anticommute." More precisely their commutator is η,

$$t_i t_j t_i^{-1} t_j^{-1} = \eta \quad (1 \leqslant i, j \leqslant l)\,. \tag{2.47}$$

For this we follow the same path of computations as in Eq. (2.26)–(2.31):

$$\begin{aligned} t_{l-1} t_l &= b_{l-1} b_l b_{l-1}^{-1} b_l \\ &= b_{l-1} b_l b_{l-1} b_l b_{l-1}^{\,2} \eta \\ &= b_l b_{l-1} b_l b_{l-1}^{-1} \eta = \eta t_l t_{l-1}\,. \end{aligned} \tag{2.48}$$

Replacing the s_i's by t_i's and (2.26) by (2.48), Eq. (2.27) carries through:

$$t_i t_l = \eta t_l t_i \quad (1 \leqslant i \leqslant l-2)\,. \tag{2.49}$$

Equation (2.28) depends only on (2.24) which is common for both $\mathbf{DT}(C_l)$ and $\mathbf{DT}(B_l)$. It reads for the latter group,

$$t_i = b_{i+1} t_i b_{i+1}^{-1} \quad (1 \leqslant i \leqslant l-2)\,. \tag{2.50}$$

To prove by recursion that t_i and t_{i+1} anticommute, we prove it first for $i=l-2$. For this we use (2.50), then (2.49),

$$\begin{aligned} t_{l-2}t_{l-1} &= b_{l-1}t_{l-2}t_l b_{l-1}{}^{-1} \\ &= \eta b_{l-1}t_l t_{l-2}b_{l-1}{}^{-1} \\ &= \eta b_{l-1}t_l b_{l-1}{}^{-1}t_{l-2} = \eta t_{l-1}t_{l-2}\,. \end{aligned} \tag{2.51}$$

We assume it true for $i+1$ and prove it for i. For this replace the s and c's of (2.30) by t and b's; use (2.51) instead of (2.29). An η will appear and this will conclude the proof of (2.47).

The group defined by Eqs. (2.46) and (2.47) is called a *Clifford group*. It is also called the *extra special two-group* in mathematics literature. We denote it by $\mathbf{CL}_l$. Its elements are the monomials of the symbolic polynomial $(1+\eta)\Pi_{i=1}^{l}(1+t_i)$. Thus its order is

$$|\mathbf{CL}_l| = 2^{l+1} \quad (1\leqslant i,j\leqslant l)\,. \tag{2.52}$$

The group $\mathbf{CL}_2$ is the quaternionic group, generated by two $i\sigma_k$, where the σ_k, $k=1,2,3$, are the three Pauli matrices. We define

$$t=\prod_{k=1}^{l} t_k\,. \tag{2.53}$$

From Eqs. (2.46) and (2.47) we get

$$\begin{aligned} &t_i t = t t_i \eta^{l-1}, \quad t^2=\eta, \quad \text{for } l\equiv 1,2 \bmod 4\,; \\ &t^2 = 1, \quad \text{for } l\equiv 0,3 \bmod 4\,. \end{aligned} \tag{2.54}$$

We have seen that in $\mathbf{W}(B_l)$, the subgroup $\mathbf{W}(A_{l-1})$ generated by the r_k's, $1\leqslant k\leqslant l-1$, acts as the group of permutations $\mathbf{S}_l$ on the l reflections in $\mathbf{A}(2,1,l)\lhd\mathbf{W}(B_l)$ ($\lhd$ reads "invariant subgroup"). The corresponding action of b_k, $1\leqslant k\leqslant l-1$, on the t_i will be, by permutations modulo elements in Ker, $\mathbf{DT}(B_l)=\Pi_{i=1}^{l}\mathbf{Z}_2(b_i{}^2)$. By computation we find that this action is only modulo η; explicitly,

$$\begin{array}{lllll} b_i t_j b_i{}^{-1} = t_j, & \eta t_{j+1}, & t_{j-1}, & t_j, \\ \text{when } j<i, & j=i, & j=i+1, & j>i+1\,. \end{array} \tag{2.55}$$

This also shows that $\mathbf{CL}_l\lhd\mathbf{DT}(B_l)$. Moreover, since the two subgroups $\mathbf{CL}_l$ and $\mathbf{DT}(A_{l-1})$ generate $\mathbf{DT}(B_l)$ and their intersection is only 1, this proves that

$$\mathbf{DT}(B_l)\sim\mathbf{CL}_l ⓢ \mathbf{DT}(A_{l-1})\sim\mathbf{CL}_l ⓢ \mathbf{SG}(2,1,l)\,, \tag{2.56}$$

with the action defined in (2.55). From this equation we obtain the action of the b_i's on t defined in (2.53); it is trivial:

$$b_i t b_i{}^{-1} = t\,. \tag{2.57}$$

From (2.54), we see that when l is odd, $t\in\mathbf{C}(\mathbf{DT}(B_l))$. Finally, with (2.54) we obtain

$$\begin{array}{llll} \mathbf{C}(\mathbf{DT}(B_l)) = \mathbf{Z}_2(\eta), & \mathbf{Z}_4(t), & Z_2(\eta)\times\mathbf{Z}_2(t), \\ \quad l \bmod 4\equiv 0,2, & 1, & 3. \end{array} \tag{2.58}$$

We recall that for all values of l, $\mathbf{C}(B_l)=\mathbf{Z}_2(\eta)$.

In Sec. III we give an explicit representation of the b_i's in the 2^l-dimensional faithful representation of Spin_{2l+1}.

We denote by φ the homomorphism from Spin_{2l+1} onto $\mathrm{SO}_{2l+1}\sim\mathrm{Spin}_{2l+1}/\mathbf{Z}_2(\eta)$. These two groups are the images of the nontrivial irreducible representations of B_l. In the tensorial representations, $\mathbf{DT}(B_l)$ is represented by the splitting image

$$\varphi(\mathbf{DT}(B_l)) = \mathbf{Z}_2{}^{l-1} ⓢ \mathbf{W}(B_l)\sim(\mathbf{Z}_2{}^{l-1}\times\mathbf{Z}_2{}^{l}) ⓢ \mathbf{S}_l\,. \tag{2.59}$$

E. The DT subgroup of D_l

We denote by d_i the generators of $\mathbf{DT}(D_l)\subset\mathrm{Spin}_{2l}$. Since $D_l=\mathrm{Spin}_{2l}$ is a maximal subgroup of $B_l=\mathrm{Spin}_{2l+1}$ with the same rank l, we know from (2.16) that

$$\mathbf{DT}(D_l)\subset\mathbf{DT}(B_l)\,, \tag{2.60}$$

and that it is of index 2, i.e., the same as $\mathbf{W}(D_l)$ in $\mathbf{W}(B_l)$, since we pass from the latter group to the former one by replacing $\mathbf{A}(2,1,l)$ in it by its subgroup of unimodular matrices $\mathbf{A}(2,2,l)=\mathbf{SA}(2,1,l)$. It contains only the products of an even number of reflections r_i. We will write the generators w_i of $\vartheta^{-1}(\mathbf{SA}(2,1,l))$ as products of pairs of the t_i's. More generally, it follows from the structure of $\mathbf{W}$ that we can write the generators of $\mathbf{DT}(D_l)$ in terms of those of $\mathbf{DT}(B_l)$. Namely,

$$d_k=b_k, \quad d_l=b_l b_{l-1} b_l{}^{-1} \quad (1\leqslant k\leqslant l-1)\,. \tag{2.61}$$

We can verify that the d_i's satisfy the equations corresponding to (E2), and (E3). In particular,

$$d_{l-1}d_l = d_l d_{l-1}\,. \tag{2.62}$$

Since $\eta\in\mathbf{C}(\mathbf{DT}(B_l))$, it is also in $\mathbf{C}(\mathbf{DT}(D_l))$. It can now be defined by

$$\eta = d_{l-1}{}^2 d_l{}^2\,. \tag{2.63}$$

We can choose for the generators of $\mathbf{SA}(2,1,l)$,

$$\begin{aligned} &w_i = t_i t_l = v_i d_{l-1}{}^{-1} d_l v_i{}^{-1}\,, \\ &w_{l-1} = t_{l-1}t_l = d_{l-1}{}^{-1} d_l\,, \\ &v_i = \prod_{k=i}^{l-2} d_k \quad (1\leqslant i\leqslant l-2)\,. \end{aligned} \tag{2.64}$$

From Eqs. (2.46) and (2.47) we find immediately that the $l-1$ w's satisfy the same equations so they generate a subgroup $\sim\mathbf{CL}_{l-1}$. This is an invariant subgroup of $\mathbf{DT}(D_l)$ that has a trivial intersection with the subgroup $\mathbf{DT}(A_{l-1})$. These two subgroups generate $\mathbf{DT}(D_l)$. Hence

$$\mathbf{DT}(D_l) = \mathbf{CL}_{l-1} ⓢ \mathbf{DT}(A_{l-1})\,, \tag{2.65}$$

where the action of the d_i's on the w_j's is defined implicitly by (2.55) when the d_i's and the w_j's are expressed, respectively, as functions of b_i and t_j [see (2.61) and (2.64)].

Let us now consider the center of $\mathbf{DT}(D_l)$. As in (2.53) we define

$$\begin{aligned} w = \prod_{k=1}^{l-1} w_k &= t, \quad &&\text{for } l \text{ even}, \\ &= \eta t t_l\,, &&\text{for } l \text{ odd}\,. \end{aligned} \tag{2.66}$$

Similarly to (2.54) we obtain

$$\begin{aligned} w w_i = w_i w\,, \quad &w^2=1, \quad \text{for } l\equiv 0,1 \bmod 4, \\ &w^2=\eta, \quad \text{for } l\equiv 2,3 \bmod 4. \end{aligned} \tag{2.67}$$

When l is even,

$$\alpha = \prod_{k \text{ odd}} d_k{}^2\,, \tag{2.68}$$

already defined in (2.23), is in $\mathbf{C}(\mathbf{DT}(A_{l-1}))$. It anticommutes with b_l, so it commutes with d_l. Hence it is in the

TABLE II. Structure of the center of the Demazure–Tits subgroup of the simple Lie group D_l and its intersection with the center of the Lie group. $\mathbf{Z}_k(y)$ denotes a cyclic group generated by y.

$l \pmod 4$	0	1	2	3
$\mathbf{C}(\mathbf{DT}(D_l))$	$\mathbf{Z}_2(\alpha)\times\mathbf{Z}_2(\eta)\times\mathbf{Z}_2(w)$	$\mathbf{Z}_2(\eta)\times\mathbf{Z}_2(w)$	$\mathbf{Z}_2(\alpha)\times\mathbf{Z}_4(w)$	$\mathbf{Z}_4(w)$
$\mathbf{C}(D_l)$	$\mathbf{Z}_2^2$	$\mathbf{Z}_4$	$\mathbf{Z}_2^2$	$\mathbf{Z}_4$
$\mathbf{C}(\mathbf{DT}(D_l))\cap(D_l)$	$\mathbf{Z}_2(\alpha)\times\mathbf{Z}_2(\eta)$	$\mathbf{Z}_2(\eta)$	$\mathbf{Z}_2(\alpha)\times\mathbf{Z}_2(\eta)$	$\mathbf{Z}_2(\eta)$

center of $\mathbf{DT}(D_l)$. We summarize the description of the center of $\mathbf{DT}(D_l)$ and its intersection with the center of G in Table II.

For l even, there are no faithful irreducible representations of D_l. We denote again by φ the homomorphism from Spin_{2l} onto $\mathrm{SO}_{2l}\sim\mathrm{Spin}_{2l+1}/\mathbf{Z}_2(\eta)$. In the tensorial representations, $\varphi(\mathbf{DT}(B_l))$ is represented by the splitting image,

$$\varphi(\mathbf{DT}(B_l)) = \mathbf{Z}_2^{l-1}\otimes\mathbf{W}(D_l)\sim(\mathbf{Z}_2^{l-1}\times\mathbf{Z}_2^{l-1})\otimes\mathbf{S}_l\,. \tag{2.69}$$

F. The DT subgroup of G_2

The Weyl group of G_2 is the dihedral group of 12 elements isomorphic to $\mathbf{S}_3\times\mathbf{Z}_2$. Therefore the order of $|\mathbf{DT}(G_2)|$ is 48. From (2.16) we know that $\mathbf{DT}(\mathrm{SU}_3)\subset\mathbf{DT}(G_2)$ and it has index 2. Note that $\mathbf{DT}(\mathrm{SU}_3)$ is isomorphic to $\mathbf{S}_4$ [see (2.20) and (2.9)]; so it is complete. That means it has no center and no outer automorphism. Hence from a known theorem[11] one has the isomorphism

$$\mathbf{DT}(G_2)\sim\mathbf{S}_4\times\mathbf{Z}_2\,. \tag{2.70}$$

We have seen that $\mathbf{DT}(A_2)\sim\mathbf{Z}_2^2\times\mathbf{S}_3\sim\mathbf{S}_4$ splits. Since $\mathbf{W}(G_2)=\mathbf{S}_3\times\mathbf{Z}_2$, (2.70) implies that $\mathbf{DT}(G_2)$ also splits,

$$\mathbf{DT}(G_2)=\mathbf{Z}_2^2\times\mathbf{W}(G_2)\sim\mathbf{Z}_2^2\otimes\mathbf{S}_3\times\mathbf{Z}_2\,. \tag{2.71}$$

We recall that $\mathbf{C}(G_2)=1$; however, $\mathbf{C}(\mathbf{DT}(G_2))\sim\mathbf{Z}_2$.

In his paper Tits[2] asks the question: What is the smallest subgroup $\mathbf{W}'$ of $\mathbf{DT}(\mathbf{G})$ that covers $\mathbf{W}(\mathbf{G})$, i.e., $\vartheta(\mathbf{W}')=\mathbf{W}(\mathbf{G})$? With the knowledge of the explicit structure of the $\mathbf{DT}(\mathbf{G})$ groups we can give the answer. It is found in Table III.

To end this section we summarize in Table IV the information obtained on the structure of the $\mathbf{DT}(\mathbf{G})$ and their centers.

TABLE III. The smallest subgroups of the Demazure–Tits group $\mathbf{DT}(\mathbf{G})$ covering the Weyl group $\mathbf{W}(\mathbf{G})$. $\mathbf{K}=\ker\vartheta$. The exception for $\mathbf{DT}(A_l)$ is due to the solvability of $\mathbf{S}_4\sim\mathbf{Z}_2^2\times\mathbf{S}_3$; the result can be understood from $A_3\sim D_3$.

G	rank l	$\mathbf{W}'$	$\mathbf{W}'\cap\mathbf{K}$
A_l	l even	$\sim\mathbf{W}$	1
	$3\neq l$ odd	$\mathbf{DT}(A_l)$	$\mathbf{Z}_2^l$
	$l=3$	$\mathbf{CL}_2\otimes\mathbf{S}_3$	$\mathbf{Z}_2(\alpha)$
C_l		$\mathbf{DT}(C_l)$	$\mathbf{Z}_2^l$
B_l		$\mathbf{CL}_l\otimes\mathbf{S}_l$	$\mathbf{Z}_2(\eta)$
D_l		$\mathbf{CL}_{l-1}\otimes\mathbf{S}_l$	$\mathbf{Z}_2(\eta)$
G_2		$\sim\mathbf{W}$	1

III. REPRESENTATIONS OF THE DEMAZURE–TITS GROUPS AND EXAMPLES

Let us underline some common features as well as differences between the well-known group $\mathbf{W}(\mathbf{G})$ and the group $\mathbf{DT}(\mathbf{G})$ that are used subsequently and provide some examples of elements R_i, $i=1,\dots,l$, generating $\mathbf{DT}(\mathbf{G})$ in some low-dimensional representations of $\mathbf{G}$ of several types and many ranks. The rank $l=2$ cases are studied in much greater detail in Secs. IV–VI. Other properties of $\mathbf{DT}(\mathbf{G})$ can be found in Sec. III of Ref. 3.

The fundamental weights $\omega_1,\dots,\omega_l$ are defined by

$$(\alpha_i,\omega_k)=\delta_{ik}(\alpha_i,\alpha_i)/2\,. \tag{3.1}$$

The weight lattice Q is the $\mathbf{Z}$ span of the fundamental weights of $\mathbf{G}$,

$$Q=\{\mu:=(a_1,\dots,a_l)\,|\,\mu=a_1\omega_1+\dots+a_l\omega_l,\ a_i\in\mathbf{Z}\}\,. \tag{3.2}$$

The sector of Q containing only dominant weights (all $a_i\geqslant 0$) is denoted Q^+. Each orbit of $\mathbf{W}$ in Q is a set of weights that contains precisely one dominant weight, say λ^+. By definition, the set of lattice points

$$\mathrm{O}(\lambda^+)=\{\mu\,|\,\mu=w\lambda^+,\ w\in\mathbf{W}\}\,, \tag{3.3}$$

is a $\mathbf{W}$ orbit, it is $\mathbf{W}$ invariant and is usually specified by its dominant weight λ^+. Subsequently, when no ambiguity could arise, we often use λ^+ for $\mathrm{O}(\lambda^+)$; similarly $\mathrm{O}(\lambda^+)$ is often denoted by $\mathbf{W}\lambda^+$. The number of elements of $\mathrm{O}(\lambda^+)$ is equal to the ratio

$$|\mathrm{O}(\lambda^+)|=|\mathbf{W}\lambda^+|=|\mathbf{W}|/|\mathrm{Stab}_w\lambda^+| \tag{3.4}$$

of the order of $\mathbf{W}$ to the order of the stabilizer of λ^+ in $\mathbf{W}$. It is tabulated in Ref. 13:

$$\mathrm{Stab}_W\lambda^+=\{w\,|\,w\lambda^+=\lambda^+\ \text{and}\ w\in\mathbf{W}\}\,. \tag{3.5}$$

$\mathrm{Stab}_W\lambda^+$ is the Weyl group of a (semisimple) Lie algebra obtained easily as follows. Take the Coxeter–Dynkin diagram of $\mathbf{G}$ ($\mathbf{W}$ is the Weyl group of $\mathbf{G}$) and attach the coordinates of the dominant weight λ^+ in the basis of the fundamental weights to the corresponding nodes of the Coxeter–Dynkin diagram. Remove nodes with nonzero coordinates. What remains is the diagram of a semisimple Lie subgroup of $\mathbf{G}$ whose Weyl group is $\mathrm{Stab}_W\lambda^+$.

An irreducible representation is specified up to $\mathbf{G}$ conjugacy by its highest weight $\Lambda\in Q^+$. Therefore a representation is usually denoted by Λ. An efficient algorithm for finding all λ^+ in $\Omega(\Lambda)$ is given in Refs. 12 and 13. For most cases of interest, λ^+ have been tabulated in Ref. 13 together with the multiplicity of their occurrences in $\Omega(\Lambda)$.

The weight system $\Omega(\Lambda)$ of a representation Λ is in-

TABLE IV. Structure of the Demazure–Tits subgroups of simple Lie groups. Symbols α, s, η, t, w, are, respectively, defined by the following equations: α: (2.23), (2.41), s: (2.39), η: (2.44), t:(2.53), w: (2.66). Here $\mathbf{Z}_n(y)$ denotes a cyclic group of order n generated by y. The Clifford group $\mathbf{CL}_l$ is defined by (2.53) and (2.54).

G	l mod 4	DT(G)	C(DT(G))	C(G)	C(DT(G))∩C(G)
A_l	0,2	$\mathbf{Z}_2{}^l ⓢ \mathbf{S}_{l+1}$	1	Z_{l+1}	1
	1,3	$\sim\mathbf{W}(B_{l+1})^+$	$\mathbf{Z}_2(\alpha)$	$\mathbf{Z}_{l+1}$	1
B_l	0,2	$\mathbf{CL}_l ⓢ \mathbf{DT}(A_{l-1})$	$\mathbf{Z}_2(\eta)$	$\mathbf{Z}_2(\eta)$	$\mathbf{Z}_2(\eta)$
B_l	1	$\mathbf{CL}_l ⓢ (\mathbf{Z}_2{}^l ⓢ \mathbf{S}_{l-1})$	$\mathbf{Z}_4(t)$	$\mathbf{Z}_2(\eta)$	$\mathbf{Z}_2(\eta)$
	3	$\mathbf{CL}_l ⓢ (\mathbf{Z}_2{}^l ⓢ \mathbf{S}_{l-1})$	$\mathbf{Z}_2(\eta)\times\mathbf{Z}_2(t)$	$\mathbf{Z}_2(\eta)$	$\mathbf{Z}_2(\eta)$
C_l		$\mathbf{Z}_4(s)\uparrow l$	$\mathbf{Z}_4(s)$	$\mathbf{Z}_2(\alpha)$	$\mathbf{Z}_2(\alpha)$
D_l	0	$\mathbf{CL}_{l-1} ⓢ \mathbf{DT}(A_{l-1})$	$\mathbf{Z}_2(\alpha)\times\mathbf{Z}_2(\eta)\times\mathbf{Z}_2(w)$	$\mathbf{Z}_2{}^2$	$\mathbf{Z}_2(\alpha)\times\mathbf{Z}_2(\eta)$
	1	$\mathbf{CL}_{l-1} ⓢ (\mathbf{Z}_2{}^l ⓢ \mathbf{S}_{l-1})$	$\mathbf{Z}_2(\eta)\times\mathbf{Z}_2(w)$	$\mathbf{Z}_4$	$\mathbf{Z}_2(\eta)$
	2	$\mathbf{CL}_{l-1} ⓢ \mathbf{DT}(A_{l-1})$	$\mathbf{Z}_2(\eta)\times\mathbf{Z}_4(w)$	$\mathbf{Z}_2{}^2$	$\mathbf{Z}_2(\alpha)\times\mathbf{Z}_2(\eta)$
	3	$\mathbf{CL}_{l-1} ⓢ (\mathbf{Z}_2{}^l ⓢ \mathbf{S}_{l-1})$	$\mathbf{Z}_4(w)$	$\mathbf{Z}_4$	$\mathbf{Z}_2(\eta)$
G_2		$\mathbf{S}_4\times\mathbf{Z}_2$	$\mathbf{Z}_2$	1	1

variant under $\mathbf{W}$ and decomposes into several $\mathbf{W}$ orbits $O(\lambda^+) = O(\mathbf{W}\lambda^+)$:

$$\Omega(\Lambda) = \bigcup_{\lambda^+} O(\lambda^+) . \tag{3.6}$$

The same orbit $O(\lambda^+)$ often occurs with multiplicity $\text{mult}_\Lambda(\lambda^+) > 1$ in $\Omega(\Lambda)$. We use n for the multiplicity $\text{mult}_\Lambda(\lambda^+)$ of λ^+ in $\Omega(\Lambda)$ whenever there is no ambiguity as to what Λ and λ^+ are. The orbit $O(\Lambda)$ of the highest weight Λ is always unique in $\Omega(\Lambda)$, i.e., $\text{mult}_\Lambda(\Lambda) = 1$.

Consider the representation space V_Λ and its decomposition

$$V_\Lambda = \bigoplus_{\lambda^+\in\Omega(\Lambda)} V_W(\lambda^+) = \bigoplus_{\lambda^+\in\Omega(\Lambda)} \bigoplus_{\mu\in O(\lambda^+)} V_\Lambda(\mu) \tag{3.7}$$

parallel to the decomposition (3.6) of $\Omega(\Lambda)$, where the subspace $V_W(\lambda^+)$ corresponds to $O(\lambda^+)$. Indeed $V_W(\lambda^+)$ is the direct sum of weight subspaces $V_\Lambda(\mu)$, $\mu\in O(\lambda^+)$. The dimensions are given by

$$\begin{aligned}\dim V_W(\lambda^+) &= |\mathbf{W}\lambda^+|\dim V_\Lambda(\mu)\\ &= |\mathbf{W}\lambda^+|\text{mult}_\Lambda \lambda^+ .\end{aligned} \tag{3.8}$$

The permutation of weights

$$\mu' = r_i\mu, \quad \mu,\mu'\in\Omega(\Lambda), \quad r_i\in W,$$

by r_i's of (2.1) exactly corresponds to the permutation of weight subspaces $V_\Lambda(\mu)$ by the elements $R_i\in\mathbf{DT}$. Namely,

$$R_i V_\Lambda(\mu) = V_\Lambda(r_i\mu) = V_\Lambda(\mu'), \quad R_i\in\mathbf{DT}, \quad 1\leqslant i\leqslant l. \tag{3.9}$$

In Ref. 3 the elements R_i are called *charge conjugation operators*. In practice one is more interested in the transformation properties of individual vectors $v_\mu\in V_\Lambda(\mu)$,

$$R_i v_\mu = v_{\mu'} = v_{r_i\mu}, \quad v_\mu\in V_\Lambda(\mu), \quad v_{r_i\mu}\in V_\Lambda(r_i\mu), \tag{3.10}$$

rather than in (3.9). Since there may be n, $n\geqslant 0$, linearly independent vectors v_μ, it turns out that the action of R_i on $V_\Lambda(\mu)$ is quite nontrivial even if r_i acts trivially on μ, i.e., if $r_i\mu = \mu$. Although one still has (3.9), it does not imply that $v_\mu = v_{\mu'}$. For examples see Ref. 3 and Appendix C of Ref. 14.

It follows from (3.9) and (3.10) that one can write symbolically

$$\mathbf{DT}V_W(\lambda^+) = V_W(\lambda^+) = \bigoplus_i m_i V(\Gamma_i), \quad m_i\in\mathbf{Z}_{>0} . \tag{3.11}$$

The action of $\mathbf{DT}$ is necessarily reducible in subspaces $V_W(\lambda^+)$ of V_Λ. Indeed, $\mathbf{DT}$, being a finite group, has finitely many irreducible representations Γ_i, $i = 1,2,\ldots,k < \infty$, while the dimension of $V_W(\lambda^+)$ has no upper limit; it grows with Λ. The summation in (3.11) extends over the irreducible representations of $\mathbf{DT}$.

Before turning to specific examples let us recall some notations and conventions. Consider l isomorphic copies of the complex Lie algebra $\mathrm{sl}(2,\mathbb{C})_i$, $1\leqslant i\leqslant l$, in 1–1 correspondence with the simple roots of $\mathbf{G}$. The basis elements e_i, f_i, h_i of each $\mathrm{sl}(2,\mathbb{C})_i$ are chosen to satisfy

$$\begin{aligned}&[e_i,f_i] = h_i, \quad [h_i,e_i] = 2e_i,\\ &[h_i,f_i] = -2f_i, \quad 1\leqslant i\leqslant l .\end{aligned} \tag{3.12}$$

The generator of $\mathbf{G}$ can be written as linear combinations of $e_i - f_i$ and $\sqrt{-1}(f_i + e_i)$ for $i\in\{1,\ldots,l\}$ and their commutators. Since we make no direct use of these other generators, there is no need to write them down here. However, we *always* assume that a Chevalley basis[11] of $\mathbf{G}$ has been chosen. It amounts to having the structure constants integer.

The charge conjugation operators[3] $R_i\in\mathbf{G}$ can be written as

$$\begin{aligned}R_i &= \exp(f_i)\exp(-e_i)\exp(f_i)\\ &= \exp \tfrac{1}{2}\pi(f_i - e_i), \quad 1\leqslant i\leqslant l .\end{aligned} \tag{3.13}$$

They generate the Demazure–Tits group $\mathbf{DT}$. It has been shown in Ref. 3 that

$$R_i{}^4 = 1, \quad R_i v_\lambda = (-1)^{(\Lambda-\lambda)/2} v_{-\lambda}, \quad v_\lambda\in V_\Lambda(\lambda), \tag{3.14}$$

where Λ (= twice the angular momentum) denotes the irreducible representation of A_1 of dimension $\Lambda + 1$ and λ is a weight of its weight system $\Omega(\Lambda) = \{\lambda,\lambda - 2,\ldots, -\lambda\}$.

Let us consider examples of R_i in the lowest representations of simple Lie groups of different types.

(A_l) The faithful representation $\Lambda = (100\cdots0)$ of dimension $l + 1$

$$R_i = I_{i-1} \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus I_{l-i}, \quad 1 \leqslant i \leqslant l. \tag{3.15}$$

Here I_k is the $k \times k$ identity matrix. In matrixlike symbols we write negative signs over the digits.

(B_l) The matrices R_i, $1 \leqslant i \leqslant l-1$ (denoted by b_i in Sec. II) corresponding to $r_i \in \mathbf{W}$ in the (faithful) 2^l-dimensional spinor representation of Spin_{2l+1} are

$$\begin{aligned} R_i &= (\oplus^{i-1} I_2) \otimes P \otimes (\otimes^{l-i-1} I_2), \quad 1 \leqslant i \leqslant l-1, \\ R_l &= (\otimes^{l-1} I_2) \otimes \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix}, \end{aligned} \tag{3.16}$$

where P is the matrix

$$P = \tfrac{1}{2}(I_2 \otimes I_2 + \sigma_3 \otimes \sigma_3 + i\sigma_1 \otimes \sigma_2 - i\sigma_2 \otimes \sigma_1) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \bar{1} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In particular, one has for $l = 3$ the B_3 representation of dimension 2^3 in a direct sum form, as

$$\begin{aligned} R_1 &= I_2 \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus I_2, \\ R_2 &= I_1 \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus I_2 \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus \mathrm{I}_1, \\ R_3 &= \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 & \bar{1} & 0 \\ 0 & 0 & 0 & \bar{1} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix}. \end{aligned} \tag{3.17}$$

Similarly one has the B_l representation of dimension $2l+1$ that is not faithful (trivial center),

$$\begin{aligned} R_i &= I_{i-1} \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus I_{2l-2i-1} \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus I_{i-1}, \\ &\quad 1 \leqslant i \leqslant l-1, \\ R_l &= I_{l-1} \oplus \begin{pmatrix} 0 & 0 & 1 \\ 0 & \bar{1} & 0 \\ 1 & 0 & 0 \end{pmatrix} \oplus I_{l-1}, \quad l \geqslant 2. \end{aligned} \tag{3.18}$$

(C_l) Representation of dimension $2l$,

$$\begin{aligned} R_i &= I_{i-1} \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus I_{2l-2i-2} \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus I_{i-1}, \\ R_l &= I_{l-1} \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus I_{l-1}. \end{aligned} \tag{3.19}$$

Note that, for $l = 2$, B_2 is identical to C_2 up to a renumbering $\alpha_1 \leftrightarrow \alpha_2$ of simple roots. In this case (3.18) and (3.19) refer to the same group in representations of dimension 5 and 4, respectively.

(D_l) When l is even no irreducible representation of $D_l = \mathrm{Spin}_{2l}$ is faithful because the center is not cyclic, $C(D_l) = Z_2^{\,2}$. In order to have a faithful representation one can consider the direct sum of the two 2^{l-1}-dimensional spinor representations. It can be obtained from the 2^l-dimensional representation of $B_l = \mathrm{Spin}_{2l+1}$. The matrices R_i corresponding to $r_i \in \mathbf{W}$ are

$$\begin{aligned} &R_i \ \text{as in (3.16)}, \quad \text{for } 1 \leqslant i \leqslant l-1, \\ &R_l = (\oplus^{l-2} I_2) \otimes Q, \end{aligned} \tag{3.20}$$

with

$$Q = \tfrac{1}{2}(I_2 \otimes I_2 - \sigma_3 \otimes \sigma_3 + i\sigma_1 \otimes \sigma_2 - i\sigma_2 \otimes \sigma_1) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \bar{1} & 0 & 0 & 0 \end{pmatrix}.$$

The D_l representation of dimension $2l$ has

$$\begin{aligned} R_i &= I_{i-1} \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus I_{2l-2i-2} \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus I_{i-1}, \\ R_l &= I_{l-2} \oplus \begin{pmatrix} 0 & 0 & \bar{1} & 0 \\ 0 & 0 & 0 & \bar{1} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \oplus I_{l-2}. \end{aligned} \tag{3.21}$$

Somewhat special is the case $l = 4$. There are three representations of dimension 8. They differ by the following permutations of R_i's,

$$\begin{aligned} &10^{0}_{0} \quad \text{as in Eq. (3.21)}, \\ &00^{0}_{1} \quad R_1 \leftrightarrow R_4, \\ &00^{1}_{0} \quad R_1 \leftrightarrow R_3. \end{aligned} \tag{3.22}$$

(G_2) Representation of dimension 7,

$$\begin{aligned} R_1 &= I_1 \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus I_1 \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus I_1, \\ R_2 &= \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 & 1 \\ 0 & \bar{1} & 0 \\ 1 & 0 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & \bar{1} \\ 1 & 0 \end{pmatrix}. \end{aligned} \tag{3.23}$$

IV. THE DEMAZURE–TITS SUBGROUP of A_2

In Secs. IV–VI we consider each of the simple Lie groups of rank 2. The description of the Demazure–Tits group **DT** in these cases is carried much further than for higher ranks because one may expect that the lowest ranks will be used most frequently; also, the derivations and results are simpler. Our analysis serves as a model of what can be learned, at least in principle, about each case, besides being a particularly useful illustration.

Each of the three groups is specified up to an isomorphism by its simple roots α_1 and α_2 or, equivalently, by the Cartan matrix

$$(A_{ij}) = \frac{2(\alpha_i, \alpha_j)}{(\alpha_j, \alpha_j)} = \begin{pmatrix} 2 & -A \\ -B & 2 \end{pmatrix}, \tag{4.1}$$

where

$$\begin{aligned} A = B = 1, &\quad \text{for } A_2, \\ A = 2B = 2, &\quad \text{for } B_2, \\ A = 3B = 3, &\quad \text{for } G_2. \end{aligned} \tag{4.2}$$

The Weyl group **W** acts on the weight lattice Q, which is the **Z** span of two fundamental weights ω_1 and ω_2. In particular,

$$\alpha_1 = 2\omega_1 - A\omega_2, \quad \alpha_2 = -B\omega_1 + 2\omega_2 , \tag{4.3}$$

and therefore

$$\begin{aligned} \omega_1 &= [1/(4-AB)](2\alpha_1 + A\alpha_2), \\ \omega_2 &= [1/(4-AB)](B\alpha_1 + 2\alpha_2) . \end{aligned} \tag{4.4}$$

The elements r_1 and r_2 generate **W** by their action (2.1) on the weights $\mu = a\omega_1 + b\omega_2 = (a,b) \in Q$, where $a,b \in \mathbb{Z}$. Namely,

$$r_1(a,b) = (-a, b + Aa), \quad r_2(a,b) = (a + Bb, -b) . \tag{4.5}$$

In particular, one has for the simple roots, $r_1\alpha_1 = r_1(2,-A) = (-2,A) = -\alpha_1$, $r_2\alpha_2 = r_2(-B,2) = (B,-2) = -\alpha_2$. A weight is called dominant if $a,b \geqslant 0$.

The "lifting" of the action of **W** on Q to the action of **DT** on V_Λ, i.e., the homomorphism $\mathbf{DT} \to \mathbf{W}$, can be set up in several equivalent but not identical ways. To avoid possible ambiguities, we adopt from now on the following prescription. The elementary reflections $r_1, r_2 \in \mathbf{W}$ of (3) are lifted into R_1, R_2 as given in (3.13) and (3.14). Any other $w \in \mathbf{W}$ is expressed as a word $r_{i_1} r_{i_2} \cdots$ of minimal length in elementary reflections. Then as it is lifted we take the result to be $R_{i_1} R_{i_2} \cdots$. The group **W** also contains one element (opposite involution) of maximal length $k_{\max}$ = number of positive roots of **G**.

The decomposition of $V_W(\lambda^+)$ into **DT**-irreducible subspaces in the three cases of rank 2 is the main problem solved in the rest of this article. Our task is to find the multiplicities m_i of occurrence of the subspaces $V(\Gamma_i)$, irreducible with respect to the representations Γ_i of **DT** in the direct sum [cf. (3.11)],

$$V_W(\lambda^+) = \oplus_i m_i V(\Gamma_i), \quad m_i \in \mathbb{Z}_{>0} . \tag{4.6}$$

Unlike the **W** orbit $\mathrm{O}(\lambda^+)$, which is independent of the rest of a weight system $\Omega(\Lambda)$ to which it may belong, the decomposition (4.6) depends on Λ and the multiplicity $n = \mathrm{mult}_\Lambda \lambda^+$. For simplicity of notation we write (4.6) as

$$\lambda^+ = \oplus_i m_i \Gamma_i . \tag{4.6$'$}$$

Let us now turn to the particular case of the Lie algebra A_2 [or Lie group SU(3)]. The multiplicity n of a dominant weight $\lambda^+ = (a,b)$ in an SU(3) representation $\Lambda = (p,q)$ is the coefficient of the term $P^p Q_q A^a B^b$ in the power expansion of the generating function[15]

$$\begin{aligned} \frac{1}{(1-PQ)^2} \Bigg\{ & \frac{1}{(1-PA)(1-QB)(1-P^2B)} \\ & + \frac{Q^2A}{(1-PA)(1-QB)(1-Q^2A)} \\ & + \frac{P^3}{(1-PA)(1-P^2B)(1-P^3)} \\ & + \frac{Q^3}{(1-QB)(1-Q^2A)(1-Q^3)} \Bigg\} . \end{aligned} \tag{4.7}$$

From (4.7) we deduce that $n = 0$ unless $p - q + b - a = 0$ mod 3, $2p + q \geqslant 2a + b$, and $p + 2q \geqslant a + 2b$. Then the orbit multiplicity n is given by

$$\begin{aligned} n = \min [& p, q, \tfrac{1}{3}(2p + q - 2a - b), \\ & \tfrac{1}{3}(p + 2q - a - 2b)] + 1 . \end{aligned} \tag{4.8}$$

The four expressions in the minimum symbol arise, respectively, from terms 4, 3,2, 1 in (4.7); there is no overlap (i.e., for given p,q,a,b at most one term contributes, namely the one giving the smallest value).

The Weyl group of A_2 is isomorphic to $\mathbf{S}_3$, the group of permutations of three objects. It is also the dihedral group $\mathbf{D}_3$. Its character table is given in Table V. That table contains as well the characters of the $\mathbf{DT}(A_2)$ group, the homomorphism between the classes of elements of W and DT groups, and the SU(3)-conjugacy classes of elements of **DT**.

The character values afforded by the three conjugacy classes of **W** are easily deduced using the action of representative elements on the points of a generic orbit (a,b), illustrated on Fig. 1.

The decomposition of Weyl group orbits on the A_2 weight lattice into direct sums of irreducible representations of W is presented in Table VI.

The structure of the Demazure–Tits subgroup $\mathbf{DT} \subset \mathrm{SU}(3)$ is found either from the SU(n) case of Sec. II or by a direct computation.[3] It turns out to be the octahedral

TABLE V. The character table of the $\mathbf{DT}(A_2)$ and $\mathbf{W}(A_2)$ groups. Subscript of the class symbol indicates the order of its elements. EFO denotes the conjugacy class in SU(3) and IR means irreducible representation.

	Weyl group					
Number of elements	1		3		2	
Representative element	I		r_1		r_1r_2	
Class / IR	C_1		C_2		C_3	
Γ_1	1	1	1	1	1	Γ_1
Γ_2	1	1	-1	-1	1	Γ_2
Γ_3	2	2	0	0	-1	Γ_3
	3	-1	-1	1	0	Γ_4
	3	-1	1	-1	0	Γ_5
Class	C_1	C_2	C_2'	C_4	C_3	IR
Representative element	I	R_1^2	$R_1R_2^2$	R_1	R_1R_2	
EFO	[100]	[011]	[011]	[211]	[111]	
Number of elements	1	3	6	6	8	
Demazure - Tits group						

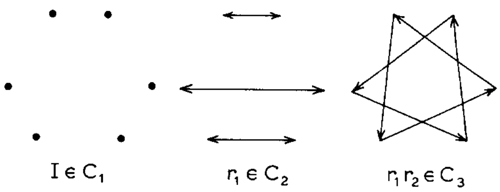


FIG. 1. Action of representative elements of conjugacy classes of the Weyl group of A_2 on weights of a generic orbit.

TABLE VI. Decomposition of the orbits of the Weyl group acting as a permutation group on the A_2 lattice. Character of each class on the orbits is shown.

W orbit	Shape	Characters E	C_2	C_3	W orbit decomposition
(a,b) $a,b>0$	hexagonal	6	0	0	$\Gamma_1 \oplus \Gamma_2 \oplus 2\Gamma_3$
$(a,0)$ or $(0,b)$ $a,b>0$	triangular	3	1	0	$\Gamma_1 \oplus \Gamma_3$
$(0,0)$	point	1	1	1	Γ_1

group. Its character table is in Table V. Each element of **W** corresponds to four elements of **DT**. The correspondences are shown in Table V. The irreducible representations Γ_1, Γ_2, and Γ_3 of **DT** coincide with Γ_1, Γ_2, Γ_3 of **W**. Our notations Γ_i, $i = 1,...,5$, for the representations of the octahedral group are taken from Ref. 16. Table V contains as well a sample element of each conjugacy class of **DT** and **W**, and its SU(3) conjugacy class is identified[7] in the case of **DT**.

Table VI contains the decomposition of **W** orbits in the weight lattice Q into direct sums of irreducible components. Let us point out that the action of **W** is reducible under a general linear transformation but cannot be further reduced when it is confined to permutations of the lattice points.

We now consider the decomposition of the **DT** orbits into direct sums of irreducible representations of **DT**. The results are summarized in Table VII.

The analysis is simplest for the generic (hexagonal) orbit; we need to consider only the classes C_1 and C_2 that correspond to Weyl class C_1. We use $R_1{}^2$ as the representative element for C_2. Its eigenvalue is $(-1)^{m_1}$, where m_1 is the SU(2) weight in the α_1 (horizontal) direction; thus the eigenvalue is $(-1)^a, (-1)^b, (-1)^{a+b}$ each for $2n$ states of the orbit and the trace (character) for C_2 is $6n$ for a,b both even, $-2n$ otherwise, as given in Table VII.

We can treat the two types of triangular orbit simultaneously by letting (b) stand for $(0,b)$ or $(b,0)$ according as b is positive or negative. Then b is the second weight component of the states of the orbit for which $m_1 = 0$. The classes C_1 and C_2 are treated as for the hexagonal orbit and have the characters given in Table VII. We must consider in addition the classes C_4 and C_2' whose representatives we take as R_1 and $R_1R_2{}^2$, respectively. Only the $m_1 = 0$ states contribute to their trace; for them the eigenvalue of $R_2{}^2$ is $(-1)^b$ and that of R_1 is $(-1)^{s_1/2}$, where s_1 is the representation label of the SU(2) group in the α_1 direction (s_1 is even for such states).

We will now derive a generating function for the characters of the classes C_4 and C_2'. The generating function for $\mathrm{SU}(3) \supset \mathrm{SU}(2) \times \mathrm{U}(1)$ is

$$F(P,Q,S,Z) = [(1-PSZ)(1-PZ^{-2}) \times (1-QSZ^{-1})(1-QZ^2)]^{-1}. \quad (4.9)$$

In the expansion of (4.9) the coefficient of $P^pQ^qS^sZ^z$ is the multiplicity of the irreducible representation (s,z) of $\mathrm{SU}(2)\times\mathrm{U}(1)$ in (p,q) of SU(3). To convert (4.9) to a generating function for the C_4 characters we retain only the part even in S [only even s representations of SU(2) contain an $m=0$ state], set $S^2 = -1$ [the eigenvalue of R_1 is $(-1)^{s/2}$], set $Z = \sqrt{B}$ and separate the result into non-negative and negative powers of B. The non-negative power part turns out to be

TABLE VII. Decomposition of orbits of the Demazure–Tits group in an SU(3) representation (p,q) into the direct sum of irreducible representations $\Gamma_1,...,\Gamma_5$ of **DT**. A **DT** orbit is specified by an SU(3) dominant weight (a,b); n is the multiplicity of (a,b) in (p,q). It is known that for (0,0) weight $n = 1 + \min\{p,q\}$; $k = p - q \bmod 2$.

DT orbit in (p,q)						Decomposition					
Dominant weight	Characters C_1	C_2	C_2'	C_4	C_3	Multiplicities of irreps of DT group Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	Restrictions
(a,b)	$6n$	$6n$	0	0	0	n	n	$2n$	⋯	⋯	a,b even
$a,b>0$	$6n$	$-2n$	0	0	0	⋯	⋯	⋯	n	n	a,b not both even
	$3n$	$3n$	0	0	0	$n/2$	$n/2$	n	⋯	⋯	b,n even
$(0,b)$	$3n$	$-n$	0	0	0	⋯	⋯	⋯	$n/2$	$n/2$	b odd, n even
for $b>0$	$3n$	$3n$	1	1	0	$(n+1)/2$	$(n-1)/2$	$(n+1)/2$	$(n+1)/2$	$(n+1)/2$	b even, n odd, $p-q$ even
	$3n$	$3n$	-1	-1	0	$(n-1)/2$	$(n+1)/2$	n	⋯	⋯	b even, n odd, $p-q$ odd
$(-b,0)$	$3n$	$-n$	-1	1	0	⋯	⋯	⋯	$(n+1)/2$	$(n-1)/2$	b,n odd, $p-q$ odd
for $b<0$	$3n$	$-n$	1	-1	0	⋯	⋯	⋯	$(n-1)/2$	$(n+1)/2$	b,n odd, $p-q$ even
	n	n	0	0	0	$n/6$	$n/6$	$n/3$	⋯	⋯	$n=0 \bmod 6$
	n	n	0	0	-1	$(n-2)/6$	$(n-2)/6$	$(n+1)/3$	⋯	⋯	$n=2 \bmod 6$
	n	n	0	0	1	$(n+2)/6$	$(n+2)/6$	$(n-1)/3$	⋯	⋯	$n=4 \bmod 6$
	n	n	1	1	1	$(n+5)/6$	$(n-1)/6$	$(n-1)/3$	⋯	⋯	$n=1 \bmod 6, k=0$
(0,0)	n	n	-1	-1	1	$(n-1)/6$	$(n+5)/6$	$(n-1)/3$	⋯	⋯	$n=1 \bmod 6, k=1$
	n	n	1	1	0	$(n+3)/6$	$(n-3)/6$	$n/3$	⋯	⋯	$n=3 \bmod 6, k=0$
	n	n	-1	-1	0	$(n-3)/6$	$(n+3)/6$	$n/3$	⋯	⋯	$n=3 \bmod 6, k=1$
	n	n	1	1	-1	$(n+1)/6$	$(n-5)/6$	$(n+1)/3$	⋯	⋯	$n=5 \bmod 6, k=0$
	n	n	-1	-1	-1	$(n-5)/6$	$(n+1)/6$	$(n+1)/3$	⋯	⋯	$n=5 \bmod 6, k=1$

$$
\frac{1}{(1-P^2Q^2)}\left(\frac{1}{(1+P^3)(1+P^2B)} + \frac{QB}{(1+P^2B)(1-QB)} - \frac{Q^3}{(1-QB)(1+Q^3)}\right). \tag{4.10}
$$

The coefficient of $P^pQ^qB^b$ in the expansion of (4.10) is the character of the class C_4 in the orbit $(0,b)$ in (p,q) of SU(3). The three terms in (4.10) never overlap (at most one contributes to the character in each case) and the character is $(-1)^{p-q+b}$ for n odd, 0 for n even, as shown in Table VII. To get the C_2' character, replace B by $-B$ in the generating function, or equivalently, multiply the C_4 character by $(-1)^b$. The characters for $(-b,0)$ orbits are obtained from the negative power (in B) part of the generating function with similar results, found in Table VII.

Finally we come to the (0,0) point orbit. The characters of C_1, C_2, C_2', C_4 are found as before. In addition we now get nonzero contributions from C_3. Since C_3 contributes nothing to the characters of other orbits, its character for the point orbit is equal to that for the whole irreducible representation of SU(3). It is given by the generating function[17]

$$
(1-PQ)/(1-P^3)(1-Q^3), \tag{4.11}
$$

i.e., 1 for $p=q=0 \bmod 3$, -1 for $p=q=1 \bmod 3$, 0 for $p=q=2 \bmod 3$, as shown in Table VII. There is no point orbit for $p-q\neq 0 \bmod 3$.

V. THE DEMAZURE-TITS SUBGROUP OF B_2

The irreducible representation (p,q) of the Lie algebra B_2 [or Lie group Sp(4) and also O(5)] has the highest weight $p\omega_1 + q\omega_2$; in particular, (1,0) and (0,1) are the representations of dimensions 5 and 4, respectively. Similarly (a,b), $a,b\geqslant 0$, denotes a dominant weight or the Weyl group orbit of the B_2 lattice containing (a,b); the multiplicity of (a,b) in the weight system of (p,q) is denoted by n.

The multiplicity n of a dominant weight $\Lambda^+ = (a,b)$ is the coefficient of the term $P^pQ^qA^aB^b$ in the power expansion of the generating function[15]

$$
\frac{1}{(1-P)(1-PA)(1-Q^2)(1-QB)} \times\left[\frac{1+PQB}{(1-P^2B^2)(1-P^2)} + \frac{Q^2}{(1-P)(1-Q^2)} + \frac{Q^2A}{(1-Q^2)(1-Q^2A)}\right]. \tag{5.1}
$$

TABLE VIII. The character table of the groups $\mathbf{DT}(B_2)$ and $\mathbf{W}(B_2)$. Subscripts of the class symbol indicate the order of its elements. Here *EFO* denotes a B_2-conjugacy class; IR is an irreducible representation.

Weyl group

Class / IR															
	1			2		2				1			2		Number of elements
	I			r_2		r_1				$r_1r_2r_1r_2$			r_1r_2		Repres. element
	C_1			C_2		C_2'				C_2''			C_4		
Γ_1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Γ_1
Γ_2	1	1	1	-1	-1	1	1	1	1	1	1	1	-1	-1	Γ_2
Γ_3	1	1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	Γ_3
Γ_4	1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	Γ_4
Γ_5	2	2	2	0	0	0	0	0	0	-2	-2	-2	0	0	Γ_5
	1	1	-1	-1	1	i	-i	i	-i	-1	-1	1	i	-i	Γ_6
	1	1	-1	-1	1	-i	i	-i	i	-1	-1	1	-i	i	Γ_7
	1	1	-1	1	-1	i	-i	i	-i	-1	-1	1	-i	-i	Γ_8
	1	1	-1	1	-1	-i	i	-i	i	-1	-1	1	i	-i	Γ_9
	2	-2	0	0	0	1+i	1-i	-1-i	-1+i	2i	-2i	0	0	0	Γ_{10}
	2	-2	0	0	0	1-i	1+i	-1+i	-1-i	-2i	2i	0	0	0	Γ_{11}
	2	-2	0	0	0	-1-i	-1+i	1+i	1-i	2i	-2i	0	0	0	Γ_{12}
	2	-2	0	0	0	-1+i	-1-i	1-i	1+i	-2i	2i	0	0	0	Γ_{13}
	2	2	-2	0	0	0	0	0	0	2	2	-2	0	0	Γ_{14}
	C_1	C_2	C_2'	C_2''	C_4	C_4'	C_4''	C_4'''	C_4^{iv}	C_4^{v}	C_4^{vi}	C_4^{vii}	C_8	C_8'	IR
Number of elements	1	1	2	4	4	2	2	2	2	1	1	2	4	4	
EFO	[100]	[010]	[001]	[001]	[110]	[201]	[201]	[021]	[021]	[110]	[110]	[110]	[111]	[111]	
Representative element	1	R_2^2	R_1^2	$R_1^2R_2$	R_2	R_1	R_1^3	$R_1R_2^2$	$R_2^2R_1^3$	$R_2R_1^3R_2R_1^3$	$R_1R_2R_1R_2$	$R_1^3R_2R_1R_2$	R_1R_2	$R_1^3R_2$	Class

Demazure-Tits group

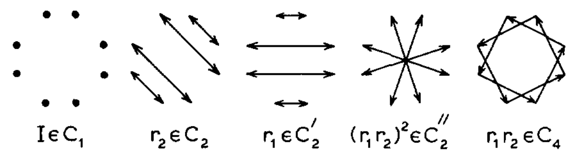


FIG. 2. Action of representative elements of conjugacy classes of the Weyl group of C_2 on weights of a generic orbit.

The character tables of the **W** and **DT** groups are given in Table VIII. The character values of the five conjugacy classes of **W** are found from the action of representative elements on the points of a generic orbit (a,b), $a>0$, $b>0$, illustrated in Fig. 2. Thus one finds the decomposition of the Weyl orbits into the direct sums shown in Table IX.

We turn to the decomposition of **DT** orbits of an arbitrary irreducible representation (p,q) of B_2. As usual the analysis is simplest for the generic (octagonal) orbit (a,b) with $a>0$ and $b>0$; only the classes C_1, C_2, C_2', which correspond to **W** class C_1 have nonzero characters. The weight vectors are eigenvectors of these classes' representative elements with the following eigenvalues:

$$I \to 1, \quad R_2^2 \to (-1)^{m_2}, \quad R_1^2 \to (-1)^{m_1}.$$

Here m_1 and m_2 are the SU(2) weights in the α_1 and α_2 directions. Thus for R_1^2 one has the eigenvalue $(-1)^a$ for the two top and two bottom states of each orbit, and $(-1)^{a+b}$ for the remaining four in the middle of the orbit. For R_2^2 one has the eigenvalue $(-1)^b$ for all eight states. In Table X one finds the decompositions.

For square representations [i.e., highest weights $(a,0)$ and $(0,b)$, $a>0$, $b>0$] the eigenvalues of representatives of the additional classes needed depend not only on the weights of the states, but also on labels s_1 and s_2 of the representation the SU(2) along the α_1,α_2 directions. We use generating functions to keep track of these additional labels.

First we consider the orbits $(a,0)$, squares with horizontal and vertical sides. The new classes are C_4 and C_2'' with representatives R_2 and $R_1^2R_2$, respectively. The characters of the classes C_1, C_2, C_2' are found as for the octagonal orbits. Only the upper right and lower left ($m_2=0$) states contribute to the characters of C_4 and C_2'', for them the eigenvalues of R_2 and $R_1^2R_2$ are, respectively, $(-1)^{s_2}$ and $(-1)^{a+s_2}$. We now derive a generating function for the characters of the classes C_4 and C_2''.

The generating function for $\mathrm{Sp}(4) \supset \mathrm{SU}(2) \times \mathrm{U}(1)$ branching rules is

$$F(P,Q;S_2,Z) = \frac{1}{(1-PZ^2)(1-PZ^{-2})(1-QS_2Z)(1-QS_2Z^{-1})} \left(\frac{1}{1-PS_2^{\,2}} + \frac{Q^2}{1-Q^2} \right). \tag{5.2}$$

In the expansion of (5.2) the coefficient of $P^pQ^qS_2^{s_2}Z^z$ is the multiplicity of the representation (s_2,z) of $\mathrm{SU}(2)\times\mathrm{U}(1)$ in (p,q) of Sp(4). To convert (5.2) to a generating function for half the C_4 character (because two states contribute), we retain the part even in S_2 [only odd-dimensional SU(2) representations have even valued weights, in particular, the weight $m_2=0$]. Then we set $S_2^2=-1$ [the eigenvalue of R_2 is $(-1)^{s_2/2}$], and set $Z^2=A$ and keep only the positive power part in A. The result is

$$\frac{1}{(1-P^2)(1+P)(1+Q^2A)} \left(\frac{1}{1-PA} + \frac{Q^4-PQ^2}{1-Q^4} \right). \tag{5.3}$$

Twice the coefficient of $P^pQ^qA^a$ is the character of C_4 for the orbit $(a,0)$. To get a generating function for half the C_2'' character substitute $A \to -A$ in (5.3) or, equivalently, multiply the C_4 character by $(-1)^a$. The coefficients of the expansions have been evaluated and the results are summarized in Table IX. We give below the multiplicity n of $(a,0)$ orbits, obtained from the generating function (5.1) with $B=0$, for all six cases q is even and $p+\frac{1}{2}q \geqslant a$:

$$\begin{array}{llll}
(1) & p,\ \tfrac12 q \geqslant a, & p-a \text{ even}, & n = 1 + \tfrac12(pq+p+q-a^2)\,, \\
(2) & p,\ \tfrac12 q \geqslant a, & p-a \text{ odd}; & n = \tfrac12(pq+p+q-a^2+1)\,, \\
(3) & p \geqslant a \geqslant \tfrac12 q, & p-a \text{ even}; & n = \tfrac14 q(\tfrac12 q+3) + \tfrac12(p-a)(q+1)+1\,, \\
(4) & p \geqslant a \geqslant \tfrac12 q, & p-a \text{ odd}; & n = \tfrac14 q(\tfrac12 q+3) + \tfrac12((p-a)(q+1)+1), \\
(5) & \tfrac12 q \geqslant a \geqslant p; & & n = \tfrac12(p+1)(p+q-2a+2), \\
(6) & a \geqslant \tfrac12 q,\ p; & & n = \tfrac12(p+\tfrac12 q-a+1)(p+\tfrac12 q-a+2).
\end{array} \tag{5.4}$$

TABLE IX. Decomposition of the orbits of the Weyl group $W(B_2)$ acting as a permutation group on the B_2 lattice. Characters of each class on the orbit are shown.

		Characters					
W orbit	Shape	C_1	C_2	C_2'	C_2''	C_4	W orbit decomposition
(a,b), $a,b>0$	octagon	8	0	0	0	0	$\Gamma_1 \oplus \Gamma_2 \oplus \Gamma_3 \oplus \Gamma_4 \oplus 2\Gamma_5$
$(a,0)$, $a>0$	square	4	2	0	0	0	$\Gamma_1 \oplus \Gamma_3 \oplus \Gamma_5$
$(0,b)$, $b>0$	square	4	0	2	0	0	$\Gamma_1 \oplus \Gamma_2 \oplus \Gamma_5$
$(0,0)$	point	1	1	1	1	1	Γ_1

TABLE X. Decomposition of the generic (octagonal) orbit of $\mathbf{DT}(B_2)$ into a direct sum of irreducible representations. n is the multiplicity of the orbit (a,b), $a,b>0$, in the representation (p,q) of B_2.

Nonzero characters			Orbit decomposition	Restrictions
C_1	C_2	C_2'		
$8n$	$8n$	$8n$	$n(\Gamma_1\oplus\Gamma_2\oplus\Gamma_3\oplus\Gamma_4\oplus 2\Gamma_5)$	a,b even
$8n$	$-8n$	0	$n(\Gamma_{10}\oplus\Gamma_{11}\oplus\Gamma_{12}\oplus\Gamma_{13})$	b odd
$8n$	$8n$	$-8n$	$n(\Gamma_6\oplus\Gamma_7\oplus\Gamma_8\oplus\Gamma_9\oplus 2\Gamma_{14})$	a odd, b even

For the square orbit $(0,b)$, with diagonal sides, the classes with nonzero trace are C_1, C_2, C_2', C_4', C_4'', C_4''', and C_4^{iv}. The characters of C_1 C_2, C_2' are found as for the octagonal orbit. We take the representative elements of C_4', C_4'', C_4''', and C_4^{iv} to be, respectively, R_1, R_1^3, $R_1R_2^2$, $R_2^2R_1^3$. Only the top and bottom ($m_1=0$) states of the orbit contribute to their characters; the eigenvalue of R_1 is $(-1)^{s_1}$ and that of R_2^2 is $(-1)^b$ for these states. We now derive a generating function for the characters of the classes in question.

The generating function for $\mathrm{Sp}(4)\supset\mathrm{SU}(2)\times\mathrm{SU}(2)$ branching rules is

$$F(P,Q;S_1,U) = [(1-P)(1-PS_1U)(1-QS_1)(1-QU)]^{-1}. \tag{5.5}$$

In the expansion of (5.5) the coefficient of $P^pQ^qS_1^{s_1}U^u$ is the multiplicity of the representation (s_1u) of $\mathrm{SU}(2)\times\mathrm{SU}(2)$ in (p,q) of $\mathrm{Sp}(4)$; here s_1 is the SU(2) representation label (highest weight) in the direction of α_1 and u is the representation label in the $\alpha_1+2\alpha_2$ (vertical) direction. To convert (5.5) into a generating function for half (because two states contribute) the C_4' character, we retain the part of (5.5) that is even in S_1 [only even s_1 representations of SU(2) have states with $m_1=0$]. Set $S_1^2=-1$ [the eigenvalue of R_1^2 is $(-1)^{s_1}$], multiply by $(1-U^{-2})(1-U^{-1}B)$ and keep the U^0 part (thereby retaining only positive u weights, which are just the orbit labels). The result is

$$\frac{1}{(1+P^2)(1+Q^2)}\left[\frac{1}{(1-P)(1+P^2Q^2)} + \frac{QB}{(1+P^2Q^2)(1-QB)} + \frac{Q^2}{(1-QB)(1-Q^2)}\right]. \tag{5.6}$$

Twice the coefficient of $P^pQ^qB^b$ is the character of C_4' (and C_4'') for the orbit $(0,b)$. To get a generating function for half the characters of C_4''' (and C_4^{iv}) for the orbit, substitute $B\to -B$ in (5.6) or, equivalently, multiply the C_4' characters by $(-1)^b$. The coefficients have been evaluated (they take only the values ± 1 and 0) and the result is found in Table XII, along with the reduction of $(0,b)$ to the direct sum of irreducible representations of **DT**. We give below the multiplicity n for $(0,b)$ orbits, obtained from the generating function (5.1) with $A=0$. For each case $q-b$ is even and $p+q\geqslant b$.

(1) p even, $q\geqslant b$;

$$n=\tfrac{1}{4}[(p-\beta+2)(\beta+1)+(p-\gamma)(\gamma+1) + (p+1)(q-b)],$$

(2) p odd, $q\geqslant b$;

$$n=\tfrac{1}{4}[(p-\delta+1)(\delta+1)+(p-\epsilon+1)(\epsilon+1) + (p+1)(q-b)],$$

(3) p even, $q<b$;

$$n=\tfrac{1}{4}[(p-\beta-\xi+2)(\beta-\xi+1) + (p-\gamma-\xi)(\gamma-\xi+1)],$$

(4) p odd, $q<b$;

TABLE XI. Decomposition of square orbit $(a,0)$ of $\mathbf{DT}(B_2)$ into the direct sum of its irreducible representations. Only nonzero characters are shown. The values of the multiplicity n are given in (5.4); $\alpha=(-1)^{q/2}(p+\frac{1}{2}q-a+2)$, $\beta=(-1)^{q/2}(p+\frac{1}{2}q-a+1)$, $\gamma=p+2$, $\delta=p+1$.

Characters					Decomposition	Restrictions
C_1	C_2	C_2'	C_2''	C_4		
$4n$	$4n$	$4n$	α	α	$(\frac{1}{2}n+\frac{1}{4}\alpha)(\Gamma_1\oplus\Gamma_3)\oplus(\frac{1}{2}n-\frac{1}{4}\alpha)(\Gamma_2\oplus\Gamma_4)\oplus 2\Gamma_5$	$a\geqslant\frac{1}{2}q$, a even, $p+\frac{1}{2}q$ even
$4n$	$4n$	$4n$	$-\beta$	$-\beta$	$(\frac{1}{2}n-\frac{1}{4}\beta)(\Gamma_1\oplus\Gamma_3)\oplus(\frac{1}{2}n+\frac{1}{4}\beta)(\Gamma_2\oplus\Gamma_4)\oplus 2\Gamma_5$	$a\geqslant\frac{1}{2}q$, a even, $p+\frac{1}{2}q$ odd
$4n$	$4n$	$-4n$	β	$-\beta$	$(\frac{1}{2}n-\frac{1}{4}\beta)(\Gamma_6\oplus\Gamma_7)\oplus(\frac{1}{2}n+\frac{1}{4}\alpha)(\Gamma_8\oplus\Gamma_9)\oplus 2\Gamma_{14}$	$a\geqslant\frac{1}{2}q$, a odd, $p+\frac{1}{2}q$ even
$4n$	$4n$	$-4n$	$-\alpha$	α	$(\frac{1}{2}n+\frac{1}{4}\alpha)(\Gamma_6\oplus\Gamma_7)\oplus(\frac{1}{2}n-\frac{1}{4}\alpha)(\Gamma_8\oplus\Gamma_9)\oplus 2\Gamma_{14}$	$a\geqslant\frac{1}{2}q$, a odd, $p+\frac{1}{2}q$ odd
$4n$	$4n$	$4n$	γ	γ	$(\frac{1}{2}n+\frac{1}{4}\gamma)(\Gamma_1\oplus\Gamma_3)\oplus(\frac{1}{2}n-\frac{1}{4}\gamma)(\Gamma_2\oplus\Gamma_4)\oplus 2\Gamma_5$	$a<\frac{1}{2}q$, a even, $p+\frac{1}{2}q$ even, p even
$4n$	$4n$	$-4n$	p	$-p$	$(\frac{1}{2}n-\frac{1}{4})(\Gamma_6\oplus\Gamma_7)\oplus(\frac{1}{2}n+\frac{1}{4}p)(\Gamma_8\oplus\Gamma_9)\oplus 2\Gamma_{14}$	$a<\frac{1}{2}q$, a odd, $p+\frac{1}{2}q$ even, p even
$4n$	$4n$	$4n$	p	p	$(\frac{1}{2}n+\frac{1}{4}p)(\Gamma_1\oplus\Gamma_3)\oplus(\frac{1}{2}n-\frac{1}{4})(\Gamma_2\oplus\Gamma_4)\oplus 2\Gamma_5$	$a<\frac{1}{2}q$, a even, $p+\frac{1}{2}q$ odd, p even
$4n$	$4n$	$-4n$	γ	$-\gamma$	$(\frac{1}{2}n-\frac{1}{4}\gamma)(\Gamma_6\oplus\Gamma_7)\oplus(\frac{1}{2}n+\frac{1}{4}\gamma)(\Gamma_8\oplus\Gamma_9)\oplus 2\Gamma_{14}$	$a<\frac{1}{2}q$, a odd, $p+\frac{1}{2}q$ odd, p even
$4n$	$4n$	$4n$	$-\delta$	$-\delta$	$(\frac{1}{2}n-\frac{1}{4}\delta)(\Gamma_1\oplus\Gamma_3)\oplus(\frac{1}{2}n+\frac{1}{4}\delta)(\Gamma_2\oplus\Gamma_4)\oplus 2\Gamma_5$	$a<\frac{1}{2}q$, a even, p odd
$4n$	$4n$	$-4n$	$-\delta$	δ	$(\frac{1}{2}n+\frac{1}{4}\delta)(\Gamma_6\oplus\Gamma_7)\oplus(\frac{1}{2}n-\frac{1}{4}\delta)(\Gamma_8\oplus\Gamma_9)\oplus 2\Gamma_{14}$	$a<\frac{1}{2}q$, a odd, p odd

TABLE XII. Decomposition of square orbit $(0,b)$ of $\mathbf{DT}(B_2)$ into irreducible representations of $\mathbf{DT}(B_2)$. Characters not shown are 0. Values of the multiplicity n are given in (5.7). For $p>b$, we have

$\alpha = +1$, for (p mod 4, q mod 4, b mod 4) = (0,0,0),(0,1,1),(0,1,3),(0,2,2),(1,0,0),(2,2,2);
$\alpha = -1$, for (p mod 4, q mod 4, b mod 4) = (1,2,0),(2,0,2),(2,1,1),(2,1,3),(2,2,0),(3,0,2);
$\alpha = 0$, otherwise.

For $p<b$, we have

$\alpha = +1$, for (p mod 4, q mod 4, b mod 4) = (0,0,0),(0,1,1),(0,2,2),(0,3,3);
$\alpha = -1$, for (p mod 4, q mod 4, b mod 4) = (2,2,0),(2,3,1),(2,0,2),(2,1,3);
$\alpha = 0$, otherwise.

Characters						
C_1	C_2	C_2'	C_4',C_4''	C_4''',C_4^{iv}	Decomposition	Restriction
$4n$	$4n$	$4n$	2α	2α	$\frac{1}{2}(n+\alpha)(\Gamma_1\oplus\Gamma_2)+\frac{1}{2}(n-\alpha)(\Gamma_3\oplus\Gamma_4)+n\Gamma_5$	b even
$4n$	$-4n$	0	2α	-2α	$\frac{1}{2}(n+\alpha)(\Gamma_{10}\oplus\Gamma_{11})+\frac{1}{2}(n-\alpha)(\Gamma_{12}\oplus\Gamma_{13})$	b odd

$$n=\tfrac{1}{2}[(p-\delta-\xi+1)(\delta-\xi+1)+(p-\epsilon-\xi+1)(\epsilon-\xi+1)]\,. \tag{5.7}$$

In the above

$$\beta=\mathrm{Min}\left\{\left[\frac{b}{2}\right],\frac{p}{2}\right\},\quad \gamma=\mathrm{Min}\left\{\left[\frac{b-1}{2}\right],\frac{p}{2}-1\right\},$$

$$\delta=\mathrm{Min}\left\{\left[\frac{b}{2}\right],\frac{p-1}{2}\right\},\quad \epsilon=\mathrm{Min}\left\{\left[\frac{b-1}{2}\right],\frac{p-1}{2}\right\},$$

$$\xi=\tfrac{1}{2}(b-q)\,.$$

Finally we turn to the (0,0) point orbit. The characters C_1, C_2, C_2', C_4', C_4'', C_4''', and C_4^{iv} are found as before. In addition we now get nonzero characters for C_4^{v}, C_4^{vi}, C_4^{vii}, C_8, and C_8'. Since their characters are zero for the other orbits, their characters on the point orbit are equal to those on the whole representation of the B_2 algebra. Thus they are given by the generating functions of Ref. 17 (replacing the variables A and B by Q and P, respectively):

$$\frac{(1+P)(1+PQ^2)}{(1-P^2)^2(1+Q^2)^2},\quad \text{for } C_4^{v},\ C_4^{vi},\ C_4^{vii};$$

$$\frac{(1-P)(1+PQ^2)}{(1-P^4)(1+Q^4)},\quad \text{for } C_8,\ C_8'\,.$$

For C_4^{v}, C_4^{vi}, C_4^{vii} we find the characters,

$$(-1)^{q/2}(\tfrac{1}{2}p+\tfrac{1}{2}q+1),\quad \text{for } p \text{ even},$$

$$(-1)^{q/2}\tfrac{1}{2}(p+1),\quad \text{for } p \text{ odd}.$$

For C_8 and C_8' we find the characters

$$\begin{array}{lll} (-1)^{q/4}, & \text{for } p=0 \bmod 4, & q=0 \bmod 4;\\ -(-1)^{q/4}, & \text{for } p=1 \bmod 4, & q=0 \bmod 4;\\ (-1)^{(q-2)/4}, & \text{for } p=1 \bmod 4, & q=2 \bmod 4;\\ -(-1)^{(q-2)/4}, & \text{for } p=2 \bmod 4, & q=2 \bmod 4;\\ 0, & \text{otherwise}. & \end{array}$$

There is no point orbit for q odd. The generating function for the multiplicity of the point is

$$(1+PQ^2)/(1-P)(1-P^2)(1-Q^2)^2, \tag{5.8}$$

which implies

$$\begin{aligned} n&=\tfrac{1}{2}(pq+p+q)+1, \quad \text{for } p \text{ even},\\ n&=\tfrac{1}{2}(p+1)(q+1), \quad \text{for } p \text{ odd}. \end{aligned} \tag{5.9}$$

The decomposition of the point orbit into irreducible representations of **DT** is given in Table XIII.

VI. THE DEMAZURE–TITS SUBGROUP OF G_2

As in the previous two cases, $(p,q)=p\omega_1+q\omega_2$ is the highest dominant weight denoting an irreducible representation of G_2. In particular, (1,0) and (0,1) are the representations of dimensions 14 and 7, respectively. A dominant weight $(a,b)=a\omega_1+b\omega_2$ denotes the **W** orbit in the G_2-weight (and also root) lattice containing it, as well as the **DT** orbit of subspaces in the representation space labeled by the highest weight (p,q). Naturally one assumes that $(a,b)\in\Omega(p,q)$, otherwise our problem is trivial.

TABLE XIII. Decomposition of the point orbit of $\mathbf{DT}(B_2)$ into its irreducible representations. The values of n are given in (5.9). $\alpha=(-1)^{q/2}\times(\frac{1}{2}p+\frac{1}{2}q+1)$, $\beta=(-1)^{q/2}\frac{1}{2}(p+1)$, $\gamma=2(-1)^{q/4}$, $\delta=2(-1)^{(q-2)/4}$.

Nonzero multiplicities of irreducible $\mathbf{DT}(B_2)$ representations					
Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	(p,q) mod 4
$\frac{1}{8}(n+p+\alpha+\gamma+4)$	$\frac{1}{8}(n-p+\alpha-\gamma)$	$\frac{1}{8}(n+p+\alpha-\gamma)$	$\frac{1}{8}(n-p+\alpha+\gamma-4)$	$\frac{1}{4}(n-\alpha)$	(0,0)
$\frac{1}{8}(n+p+\alpha)$	$\frac{1}{8}(n-p+\alpha)$	$\frac{1}{8}(n+p+\alpha)$	$\frac{1}{8}(n-p+\alpha)$	$\frac{1}{4}(n-\alpha)$	(0,2)
$\frac{1}{8}(n-p+\beta-\gamma+1)$	$\frac{1}{8}(n+p+\beta+\gamma+3)$	$\frac{1}{8}(n-p+\beta+\gamma-3)$	$\frac{1}{8}(n+p+\beta-\gamma-1)$	$\frac{1}{4}(n-\beta)$	(1,0)
$\frac{1}{8}(n-p+\beta+\delta-3)$	$\frac{1}{8}(n+p+\beta-\delta-1)$	$\frac{1}{8}(n-p+\beta-\delta+1)$	$\frac{1}{8}(n+p+\beta+\delta+3)$	$\frac{1}{4}(n-\beta)$	(1,2)
$\frac{1}{8}(n+p+\alpha+2)$	$\frac{1}{8}(n-p+\alpha-2)$	$\frac{1}{8}(n+p+\alpha+2)$	$\frac{1}{8}(n-p+\alpha-2)$	$\frac{1}{4}(n-\alpha)$	(2,0)
$\frac{1}{8}(n+p+\alpha-\delta-2)$	$\frac{1}{8}(n-p+\alpha+\delta-2)$	$\frac{1}{8}(n+p+\alpha+\delta+2)$	$\frac{1}{8}(n-p+\alpha-\delta+2)$	$\frac{1}{4}(n-\alpha)$	(2,2)
$\frac{1}{8}(n-p+\beta-1)$	$\frac{1}{8}(n+p+\beta+1)$	$\frac{1}{8}(n-p+\beta-1)$	$\frac{1}{8}(n+p+\beta+1)$	$\frac{1}{4}(n-\beta)$	(3,0),(3,2)

TABLE XIV. Character table of the $\mathbf{DT}(G_2)$ and $\mathbf{W}(G_2)$ groups. Representative element of each conjugacy class is shown. Subscript on class symbol is the order of its elements. Conjugacy classes of G_2 are given as EFO. IR is an irreducible representation.

Weyl group											
Class	I		$(r_1r_2)^3$		r_2		r_1		$(r_1r_2)^2$	r_1r_2	Representative element
	1		1		3		3		2	2	Number of elements
IR	C_1		C_2		C_2'		C_2''		C_3	C_6	
Γ_1	1	1	1	1	1	1	1	1	1	1	Γ_1
Γ_2	1	1	-1	-1	-1	-1	1	1	1	-1	Γ_2
Γ_3	1	1	1	1	-1	-1	-1	-1	1	1	Γ_3
Γ_4	1	1	-1	-1	1	1	-1	-1	1	-1	Γ_4
Γ_5	2	2	-2	-2	0	0	0	0	-1	1	Γ_5
Γ_6	2	2	2	2	0	0	0	0	-1	-1	Γ_6
	3	-1	-3	1	1	-1	-1	1	0	0	Γ_7
	3	-1	3	-1	-1	1	-1	1	0	0	Γ_8
	3	-1	-3	1	-1	1	1	-1	0	0	Γ_9
	3	-1	3	-1	1	-1	1	-1	0	0	Γ_{10}
	C_1	C_2	C_2'	C_2''	C_2'''	C_4	C_2''''	C_4'	C_3	C_6	IR
Number of elements	1	3	1	3	6	6	6	6	8	8	
EFO	[100]	[001]	[001]	[001]	[001]	[110]	[001]	[201]	[101]	[111]	
Representative element	I	R_1^2	$(R_1R_2)^3$	$R_1R_2R_1^3R_2R_1R_2$	$R_1^2R_2$	R_2	$R_1R_2^2$	R_1	$(R_1R_2)^2$	R_1R_2	Class
Demazure–Tits group											

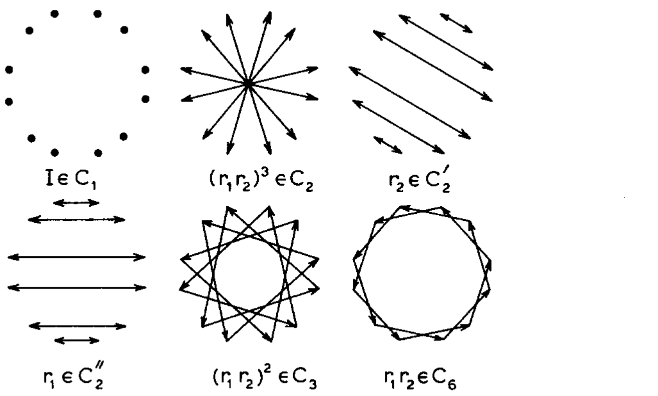


FIG. 3. Action of representative elements of conjugacy classes of the Weyl group of G_2 on weights of a generic orbit.

The multiplicity $n = \mathrm{mult}_{(p,q)}(a,b)$ of a weight (a,b) in the weight system $\Omega(p,q)$ is also the multiplicity of the **DT** orbit. It can be found either in the tables of Ref. 13 (for the lowest 100 representations) or it can be calculated using the G_2 character generator, Eq. (2.7) of Ref. 18. There in order to conform to present notation the following substitutions should be made: $A \rightarrow Q, B \rightarrow P, \eta \rightarrow AB^{-3/2}, \xi \rightarrow B^{1/2}$; then the coefficient of the term $P^pQ^qA^aB^b$ (a,b non-negative) in the power expansion of the generating function is the multiplicity n.

The character table of the Weyl group $\mathbf{W}(G_2)$ and the Demazure–Tits group $\mathbf{DT}(G_2)$ are found in Table XIV.

First consider $\mathbf{W}$ acting on the G_2 weight lattice. Representative elements of the $\mathbf{W}$-conjugacy classes are

$$\begin{aligned} &C_1:\ I, \quad C_2:\ (r_1r_2)^3, \quad C_2':\ r_2, \\ &C_2'':\ r_1, \quad C_3:\ (r_1r_2)^2, \quad C_6:\ r_1r_2. \end{aligned} \tag{6.1}$$

The subscript on a class symbol is the order of its elements; r_1 and r_2 are the elementary reflections (2.1). The traces of classes of each type are easy to determine as before: each point of the orbit that is not moved by the representative element contributes 1 to the trace. Hence it suffices to see the action of the representative of each class on $Q(G_2)$. It is shown in Fig. 3.

Consider the generic, or dodecagonal, orbit (a,b), $a>0$, $b>0$, of the Weyl group in the G_2 weight lattice Q. The class C_1 has trace 12, while all other classes have trace 0. Hence one has the decomposition $(a,b) = \Gamma_1 \oplus \Gamma_2 \oplus \Gamma_3 \oplus \Gamma_4 \oplus 2\Gamma_5 \oplus 2\Gamma_6$ as shown in Table XV. Similarly for the hexagonal orbit $(a,0)$, $a>0$, the class C_1 has trace 6, the class C_2' has trace 2, and all other classes have trace 0. We find the decomposition $(a,0) = \Gamma_1 \oplus \Gamma_4 \oplus \Gamma_5 \oplus \Gamma_6$ (cf. Table XV). For the other hexagonal orbit, $(0,b)$, $b>0$, the class C_1 has trace 6, the class C_2'' has trace 2, and the others are 0. The decomposition is $(0,b) = \Gamma_1 \oplus \Gamma_2 \oplus \Gamma_5 \oplus \Gamma_6$. Finally for the point orbit $(0,0)$ each class has trace 1 so that its decomposition is $(0,0) = \Gamma_1$. The decomposition of Weyl group orbits of $Q(G_2)$ is summarized in Table XV.

Next let us consider the **DT** group acting on the weight vector basis of V_Λ, $\Lambda = (p,q)$ and let us find the decomposition (3.11).

We consider first the generic orbit (a,b), $a>0$, $b>0$, which appears with multiplicity n in $V_{(p,q)}$. The classes with nonzero traces are C_1 and C_2. The trace of C_1 is $12n$. For C_2 we have the representative element R_1^2; its eigenvalue is $(-1)^{m_1}$, where m_1 is the SU(2) weight in the α_1 direction. The values of $|m_1|$ at the 12 points of the orbit are a, $a+b$,

TABLE XV. Decomposition of the Weyl group orbits of the G_2 lattice.

W orbit on G_2 lattice	Shape	Characters of classes C_1	C_2	C_2'	C_2''	C_3	C_6	Orbit decomposition
(a,b) $a,b>0$	dodecagonal	12	0	0	0	0	0	$\Gamma_1 \oplus \Gamma_2 \oplus \Gamma_3 \oplus \Gamma_4 \oplus 2\Gamma_5 \oplus 2\Gamma_6$
$(a,0)$ $a>0$	hexagonal	6	0	0	2	0	0	$\Gamma_1 \oplus \Gamma_4 \oplus \Gamma_5 \oplus \Gamma_6$
$(0,b)$ $b>0$	hexagonal	6	0	2	0	0	0	$\Gamma_1 \oplus \Gamma_2 \oplus \Gamma_5 \oplus \Gamma_6$
$(0,0)$	point	1	1	1	1	1	1	Γ_1

$2a+b$, each $4n$ times. Hence the trace for C_2 is $12n$ for a,b both even, and $-4n$ otherwise. Hence one has the decomposition as given in Table XVI.

The hexagonal orbit $(a,0)$, $a>0$, has two horizontal sides; the classes with nonzero character are C_1, C_2, C_2''', C_4, as follows from Fig. 3. The trace of C_1 is $6n$. For C_2 the trace is $6n$ if a is even, and $-2n$ if a is odd. We will derive generating functions for traces of C_2''' and C_4. Orient the $SU(2)\times SU(2)$ subgroup of G_2 so that α_2 points in the direction of the second SU(2) root. The states not moved by R_2 and $R_1{}^2R_2$, the representative elements of C_4 and C_2''', respectively, are those with dominant weight $(a,0)$ and opposite weight $(-a,0)$. On these states the eigenvalue of R_2 is $(-1)^{t/2}$, and that of $R_1{}^2$ is $(-1)^a$; $|m_s|$ takes the value $2a$, where (s,m_s) are the representation label and weight of the first SU(2) subgroup and (t,m_t) those of the second.

The even-even part of the $G_2\supset SU(2)\times SU(2)$ branching rules generating function is found from Ref. 18, Eq. (3.1) (to conform to our present notations, the substitutions $A\to Q$ and $B\to P$ should be made):

$$F(P,Q;S^2,T^2)=\frac{1}{(1-P^2)(1-PS^2)(1-QT^2)(1-Q^2S^2T^2)}\left[\frac{1+PQ^3S^2+Q^3S^2T^2+PQ^3S^2T^2}{(1-Q^3S^2)(1-Q^2)}+\frac{PT^2+PQT^2+P^2QS^2T^4+PQ^2S^2T^2}{(1-Q^2)(1-PT^2)}+\frac{P^2S^2T^6+P^3S^2T^6+PQS^2T^4+P^4QS^4T^{10}}{(1-PT^2)(1-P^2S^2T^6)}\right]. \tag{6.2}$$

Because $R_2=(-1)^{t/2}$, we set $T^2=-1$. The result is

$$F'(P,Q;S^2)=\frac{1}{(1-P^2)(1+Q)}\left[\frac{1}{(1-PS^2)(1-Q^2)(1+Q^2S^2)}-\frac{P}{(1+P)(1-PS^2)(1-Q^2)}-\frac{PQ}{(1+P)(1-Q^2)(1+Q^2S^2)}+\frac{PQS^2}{(1+P)(1-PS^2)(1+Q^2S^2)}-\frac{P^2S^2+P^3QS^4}{(1+P^2S^2)(1-PS^2)(1+Q^2S^2)}\right]. \tag{6.3}$$

Finally we convert this generating function for SU(2) representations to the corresponding one for non-negative SU(2) weights (or G_2 orbit labels, since $a=\frac{1}{2}m_s$) by computing

$$\begin{aligned}G(P,Q;A)&=\left.\frac{F'(P,Q;S^2)}{(1-S^{-2})(1-S^{-2}A)}\right|_{S^0}\\&=\frac{1}{1+Q}\left[\frac{1}{(1-P)(1-P^2)(1-Q^4)(1-PA)}-\frac{Q^2A}{(1-P^2)(1-Q^4)(1-PA)(1+Q^2A)}-\frac{P}{(1-P^2)^2(1-Q^2)(1-PA)}\right.\\&\quad-\frac{PQ}{(1-P^2)(1+P)(1-Q^4)(1+Q^2A)}+\frac{PQ}{(1-P^2)^2(1+Q^2)(1+Q^2A)}\\&\quad+\frac{PQA}{(1-P^2)^2(1-PA)(1+Q^2A)}-\frac{P^2+P^3Q}{(1-P^4)(1-P)(1+Q^2)(1-PA)}\\&\quad\left.-\frac{P^2A+P^3QA^2}{(1-P^4)(1+P^2A)(1-PA)(1+Q^2A)}+\frac{P^2Q^2A-P^3QA^2}{(1-P^4)(1+Q^2)(1-PA)(1+Q^2A)}\right].\end{aligned} \tag{6.4}$$

TABLE XVI. Decomposition of G_2 orbits of the Demazure–Tits group **DT** in a representation (p,q) into a direct sum of irreducible representations $\Gamma_1,\dots,\Gamma_{10}$ of **DT**. An orbit is given by a G_2 dominant weight (a,b); n is the multiplicity of (a,b) in (p,q). Notation: c, d, e, f, g are the coefficients of the term $P^pQ^qA^aB^b$ in the power series of Eqs. (6.5), (6.8), (6.9), (6.10), (6.11), respectively; $X_\pm=(n\pm e)/12$, $Y_\pm=(d\pm c)/4$, $Z_\pm=(f\pm g)/6$.

DT orbit in (p,q)												Decomposition									
Dominant weight		Characters										Multiplicities of irreps of DT(G_2)									
		C_1	C_2	C_2'	C_2''	C_2'''	C_2^{IV}	C_4	C_4'	C_3	C_6	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	Γ_6	Γ_7	Γ_8	Γ_9	Γ_{10}
(a,b)	a,b even	12n	12n	0	0	0	0	0	0	0	0	n	n	n	n	2n	2n	0	0	0	0
a,b>0	otherwise	12n	-4n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	n	n	n	n
(a,0)	a even	6n	6n	0	0	2c	0	2c	0	0	0	$\frac{n+c}{2}$	$\frac{n-c}{2}$	$\frac{n-c}{2}$	$\frac{n+c}{2}$	n	n	0	0	0	0
	a odd	6n	-2n	0	0	-2c	0	2c	0	0	0	0	0	0	0	0	0	$\frac{n-c}{2}$	$\frac{n+c}{2}$	$\frac{n+c}{2}$	$\frac{n-c}{2}$
(0,b)	b even	6n	6n	0	0	0	2d	0	2d	0	0	$\frac{n+d}{2}$	$\frac{n+d}{2}$	$\frac{n-d}{2}$	$\frac{n-d}{2}$	n	n	0	0	0	0
	b odd	6n	-2n	0	0	0	-2d	0	2d	0	0	0	0	0	0	0	0	$\frac{n+d}{2}$	$\frac{n+d}{2}$	$\frac{n-d}{2}$	$\frac{n-d}{2}$
(0,0)		n	n	e	e	c	d	c	d	f	g	$X_++Y_++Z_+$	$X_-+Y_-+Z_-$	$X_+-Y_++Z_+$	$X_--Y_-+Z_-$	$2X_--Z_-$	$2X_+-Z_+$	0	0	0	0

The symbol $|_{S^0}$ indicates that only the 0th power of S term of (6.4) should be retained. The power series expansion of $G(P,Q;A)$,

$$G(P,Q;A) = \sum_{pqa} P^p Q^q A^a c_{pqa}\,, \tag{6.5}$$

states that the trace of class C_4 is $2c_{pqa}$ for the orbit $(a,0)$ in (p,q); for C_2'' the trace is $2(-1)^a c_{pqa}$. The factor 2 appears because two states contribute to the trace. The decomposition of the $(a,0)$ orbit is shown in Table XVI.

For the hexagonal orbit $(0,b)$, $b>0$ (two vertical sides), the classes with nonzero trace are C_1, C_2, C_2^{iv}, C_4'. For C_1 the trace is $6n$. For C_2 it is $6n$ for b even, $-2n$ for b odd. We derive generating functions for the trace of C_2^{iv} and C_4', using the representative elements $R_2^2R_1$ and R_1, respectively. Orient the $\mathrm{SU}(2)\times\mathrm{SU}(2)$ subgroup with the first SU(2) root along α_1 of G_2. The states not moved by R_1 and $R_2^2R_1$ are those with weights $(0,\pm b)$. On these states the eigenvalue of R_1 is $(-1)^{s/2}$ [s is the first SU(2) representation label] and that of R_2^2 is $(-1)^b$; $|m_t|$ takes the value $2b$ [m_t is the second SU(2) weight]. Because $R_1=(-1)^{s/2}$, we set $S^2=-1$ in the generating function (6.2) with the result

$$F''(P,Q;T^2) = \frac{1}{(1-P^2)(1+Q^2T^2)}\left[\frac{1}{(1-Q^2)(1+Q^3)} + \frac{QT^2}{(1+Q^3)(1-QT^2)} - \frac{P^2T^6}{(1-QT^2)(1+P^2T^6)} - \frac{P-PT^2-P^2T^4}{(1+P^2T^6)(1+P)} - \frac{PQ^2}{(1+P)(1-Q^2)}\right]. \tag{6.6}$$

Finally we convert this generating function for SU(2) representations into the corresponding one for non-negative weights (or G_2 orbits labels, since $b=\frac{1}{2}m_t$) by computing

$$\begin{aligned} H(P,Q;B) &= \frac{F''(P,Q;T^2)}{(1-T^{-2})(1-T^{-2}B)}\Bigg|_{T^0} \\ &= \frac{1}{1+Q^2B}\left[\frac{1+Q+Q^2}{(1-P^2)(1+Q^3)(1-Q^4)} + \frac{P^2}{(1+P)(1-P^4)(1+Q^2)} - \frac{P^2}{(1-P^4)(1-Q)(1+Q^2)} + \frac{PQ^2}{(1+P)(1-P^2)(1-Q^4)}\right. \\ &\quad - \frac{P^2B+P^2B^2+P^4B^3}{(1-P^4)(1-Q)(1+P^2Q^3)} + \frac{PB+P^2B+P^2B^2+P^3B^3+P^4B^3-P^3-P^3B}{(1+P)(1-P^4)(1+P^2B^3)} \\ &\quad \left. + \frac{QB}{(1-P^2)(1-Q)(1+Q^3)(1-QB)} - \frac{P^2B^3}{(1-P^2)(1-Q)(1-QB)(1+P^2B^3)}\right]. \end{aligned} \tag{6.7}$$

The power series expansion of $H(P,Q;B)$,

$$H(P,Q,B) = \sum_{pqb} P^pQ^qB^b d_{pqb}\,, \tag{6.8}$$

gives the trace of the class C_4 for the orbit $(0,b)$ in the G_2 representation (p,q) as $2d_{pqb}$; for C_2''' the trace is $2(-1)^b d_{pqb}$. The decomposition of the orbit $(0,b)$ is given in Table XVI.

Finally we deal with the point orbit $(0,0)$. All classes can now have nonzero trace. The traces of classes C_1, C_2, C_2''', C_2^{iv}, C_4, C_4' are computed as above for the hexagonal orbits. Thus the trace of C_1 and C_2 is n, the multiplicity of the orbit. A generating function for n is obtained from (6.2) by setting $S=T=1$, since each even $\mathrm{SU}(2)\times\mathrm{SU}(2)$ representation has just one state at the origin; n for (p,q) is the coefficient of P^pQ^q in the power series expansion. For C_2^{iv} and C_4' the trace is $c=c_{pq0}$, the coefficient of $P^pQ^qA^0$ in the expansion of (6.4). For C_2''' and C_4 the trace is $d=d_{pq0}$, the coefficient of $P^pQ^qB^0$ in the expansion of (6.7). Since the remaining classes have zero trace for all but the point orbit, their trace for the point orbit is their character in the whole irreducible representation (p,q). Accordingly we can get it from the known generating functions for the characters of the corresponding G_2-conjugacy class of elements of finite order in G_2, Ref. 17. For C_2' and C_2'' we have

$$\begin{aligned}\sum_{pq} P^pQ^q e_{pq} &= \frac{1}{(1+P)^2(1-P^2)^2(1+Q)^2(1-Q^2)^2} \\ &\quad\times[1+P-2PQ-P^2Q-PQ^2 \\ &\quad +Q^3+2P^3Q-2P^2Q^2+2PQ^3 \\ &\quad +P^4Q-P^3Q^2-P^2Q^3-2P^3Q^3 \\ &\quad +P^3Q^4+P^4Q^4]\,. \end{aligned} \tag{6.9}$$

For C_3 we have

$$\begin{aligned}\sum_{pq} P^pQ^q f_{pq} &= \frac{1}{(1-P^3)^2(1+Q+Q^2)^3} \\ &\quad\times[1+P+2Q+2Q^2+PQ^2+PQ \\ &\quad +Q^3+P^4Q+P^3Q^2 \\ &\quad +2P^4Q^2+P^3Q^3+2P^4Q^3 \\ &\quad +P^3Q^4+P^4Q^4]\,. \end{aligned} \tag{6.10}$$

For C_6 we have

$$\sum_{pq} P^pQ^q g_{pq} = \frac{(1-Q^2)(1-P+Q-P^4Q+P^3Q^2-P^4Q^2)}{(1-P^6)(1-Q^6)}\,. \tag{6.11}$$

Our result, the decomposition of the point orbit, is given in Table XVI.

VII. CONJUGACY CLASSES OF ELEMENTS GENERATING THE DEMAZURE–TITS GROUPS

In this section we consider the elements R_k, $k\in\{1,2,...,l\}$, which generate the Demazure–Tits group **DT(G)** up to equivalence transformation by the simple connected Lie group **G**, and identify the **G**-conjugacy classes to which they belong. Since part of that has been done already in Ref. 7, here we just complete Table III of that article.

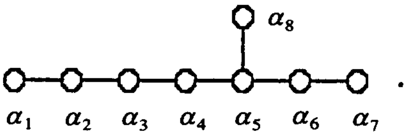
First let us show that R_k, $k\in\{1,2,...,l\}$, are rational elements in any **G**. (An element is rational if its character values for any representation of **G** are integers.) Consider $R_k \in \mathrm{SU}_k(2) \in \mathbf{G}$, and the subgroup $\mathrm{SU}_k(2)$ whose simple root is α_k. The character value of R_k for any representation $\Lambda(\mathbf{G})$ of **G** is by definition its character for the subgroup representation $\Lambda(\mathrm{SU}_k(2)) \subset \Lambda(\mathbf{G})$. Then recalling[3,17] that R_k is a rational element of $\mathrm{SU}_k(2)$, it has to be rational also in **G**.

We know[3] that all R_k are of order 4 and that those R_k corresponding to simple roots α_k of the same length are **G** conjugate, while any two R_k corresponding to roots of different lengths are not **G** conjugate. Therefore here we have to identify one conjugacy class of elements of order 4 in D_l, E_6, E_7, and E_8 and two such conjugacy classes in F_4. For all other cases the conjugacy classes were found.[7] All the conjugacy classes of R_k are shown in Table XVII.

From now on we assume the conventions and results of Ref. 7. In particular, elements of finite order in **G** are denoted by relatively prime non-negative integers attached to the nodes of extended Coxeter–Dynkin diagram; we use the Dynkin numbering of the nodes (cf., for instance, Ref. 7 or Ref. 13). It is not difficult to list all conjugacy classes of elements of order 4 in any **G**. Thus, for example, there are only seven such classes of elements in E_8. Since this is clearly the most complicated case we have to face, we illustrate in this example how one can proceed.

Let $g\in E_8$ belong to one of the seven E_8-conjugacy classes of elements of order 4, $g^4=1$. Note that all E_8 representations are self-contragredient. Therefore g and $g^{-1}=g^3$ are conjugate, $g\sim g^3$. That is, all powers of g relatively prime to 4 are conjugate to g. Consequently,[7] the character $\chi_\Lambda(g)$ of any element of our seven conjugacy classes is an integer in any representation Λ of E_8.

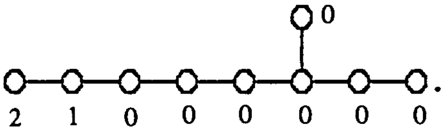
Since all eight R_k, $k\in\{1,2,...,8\}$, are E_8 conjugate, it suffices to consider only, say, R_1. In adopted conventions the E_8 simple roots are numbered as



Let us find the character value $\chi_{\mathrm{Ad}(E_8)}(R_1)$ of R_1 on the 248-dimensional (adjoint) representation $\mathrm{Ad}(E_8)$. But h_1 is orthogonal to the diagram with α_1 removed. That is, the diagram of E_7 (its dimension is 133) on which the SU(2) with simple root α_1 acts trivially. Otherwise R_1 sends h_1 to $-h_1$ and merely transposes the remaining root vectors in pairs which contributes nothing to the character. Therefore one has

$$\chi_{\mathrm{Ad}(E_8)}(R_k)=132, \quad k\in\{1,2,...,8\}.$$

Next we find which of the seven elements of order 4 in E_8 has that character value on $\mathrm{Ad}(E_8)$. It turns out that there is just one such element [21000000]. Using the extended diagram, it is given as



In order to verify that its character is indeed 132, one can consult the table of positive roots of E_8 (pp. 62 and 63 of Ref. 13), this time reading the roots in the simple root basis (α basis). We need to know only the α_1 coordinate of each root. That coordinate takes only five values ± 2, ± 1, 0, negative values occurring for negative roots only. An E_8 root with the α_1 coordinate m contributes[7] to the character value $\exp(2\pi i m/4)$. Moreover since the character must be integer, the values $m=1$ and 3 can be disregarded; they must cancel out. Among the positive roots one finds 63 times $m=0$ and once $m=2$; the negative roots contribute similarly. Adding the eight zero weights of the adjoint representation as another $m=0$ eight times, one gets the character as 132. In the same way, but much more quickly, one can determine the rest of the conjugacy classes of R_k in any other simple **G**.

TABLE XVII. G-conjugacy classes of elements generating the Demazure–Tits group and their second powers. Subscript short (long) corresponds to short (long) simple roots of a simple Lie algebra.

G		R_{long}	R^2_{long}	R_{short}	R^2_{short}
A_1		[11]	[01]	···	···
A_l	$l>2$	[210···01]	[010···01]	···	···
B_l	$l>2$	[2010···0]	[0010···0]	[110···0]	[010···0]
C_l	$l>3$	[210···0]	[010···0]	[2010···0]	[0010···0]
D_l	$l>4$	[2010···0]	[0010···0]	···	···
E_6		[2000001]	[0000001]	···	···
E_7		[21000000]	[01000000]	···	···
E_8		[210000000]	[010000000]	···	···
F_4		[2100]	[0100]	[2001]	[0001]
G_2		[210]	[010]	[101]	[010]

VIII. CONCLUDING REMARKS

The Weyl group has been the most important device in virtually any extensive work with representations of high rank ($\gg 1$) simple Lie algebras/groups. The higher the rank the more difficult it is to proceed without it.

Physical states "live" in representation spaces rather than in spaces populated by roots of an algebra or weights of its representations. Consequently, the symmetries of the Weyl group are no more than an (homomorphic) image of the general symmetries of physical states. Moreover, interesting problems at any period of time are usually at (or beyond) the limits of what one can calculate with present day methods. Therefore using only the Weyl group is helpful but one can often proceed much more effectively.

A motivation to carry out large scale computations is often present in physics but only rarely in mathematics. That is perhaps the reason that a tool of prime importance like the Demazure–Tits group has been relatively little studied by mathematicians.

This independent sequel to Ref. 3 is an attempt to partially rectify the situation. The principal results are the following: Description of the **DT** in the classical series of simple Lie groups and G_2; identification of the conjugacy classes (under the Lie group action) of the elements generating **DT**; finding the character table of **DT** in simple Lie groups of rank 2; and decomposition of all finite-dimensional representations of rank 2 Lie groups into direct sums of irreducible components of **DT**.

There remain unsolved other equally interesting problems involving **DT**. We name a few.

The character tables of **DT** group in simple Lie groups of rank >2. An extension of known character tables of W to those of **DT**, as exemplified here for rank 2, is possible and it may not even be difficult.

The structure of **DT** in E_6, E_7, E_8, and F_4. The following appears to be true: $\mathbf{DT}(E_k)\subset\mathbf{DT}(E_{k+1})$ for $k=6$ and 7. The homomorphism $\mathbf{DT}(E_k)\to\mathbf{W}(E_k)$ is nonsplit.

Branching rules for Lie groups of rank >2 to **DT**. The multiplicities of Weyl group orbits in corresponding weight systems are either known[13] or can easily be found right now for every case which may conceivably ever be needed.

Integrity bases of invariants and covariants of **DT**. Their description along the lines, for instance, Ref. 16 is possible at least for lower ranks.

Let us finish the article with a remark concerning the action of $\mathbf{DT}(\mathbf{G})$ on a generic orbit $V_W(\lambda^+)$. Its dominant weight $\lambda^+=(\lambda_1,\ldots,\lambda_l)$ has only trivial stabilizer in $\mathbf{W}$; equivalently, λ^+ has no zero coordinates in the basis of fundamental weights, $\lambda_j>0$ for any $1\leqslant j\leqslant l$. The decomposition (3.11) in this case depends only on the values λ_j mod 2, $1\leqslant j\leqslant l$ and not on the highest weight Λ of any representation of $\mathbf{G}$.

The only elements of $\mathbf{DT}(\mathbf{G})$ which have nonzero trace on $V_W(\lambda^+)$ are the 2^l elements which are mapped under ϑ' of (1.2) to the identity element of $\mathbf{W}$. All other elements of **DT** move every vector of $V_W(\lambda^+)$. The 2^l elements are of the form

$$\prod_{i=1}^{l}(R_i^2)^{\delta_i},\quad \delta_i=0\quad\text{or}\ 1\,.$$

The eigenvalue of $R_i{}^2$ acting on any vector of weight $\Sigma_k m_k\omega_k$ is just $(-1)^{m_i}$. The weight component m_i is also the SU(2) weight in the α_i direction.

The eigenvalues of all elements of **DT** with nonzero trace thus depend only on the weights of the orbit. Their characters and hence their orbit decomposition, therefore depend only on the parity of λ_j's.

ACKNOWLEDGMENTS

The authors are grateful for the hospitality of the Institute des Hautes Etudes Scientifiques (J. P. and R. S.), to the Centre de recherches mathématiques, Université de Montréal (L. M.), and to the Aspen Center for Physics (J. P.) where this work was pursued. We would also like to thank A. J. Coleman, R. Griess, and J. McKay for helpful remarks and for reading the manuscript.

APPENDIX: A SUMMATION FORMULA

Here we derive the following identity:

$$\sum_{x=0}^{[q/2]}\frac{(-1)^{q-x}(q-x)!}{x!(q-2x)!}=(q+2)\ \mathrm{mod}\ 3-1\,,\qquad\text{(A1)}$$

which we have not been able to find in the literature. The right-hand side is the character of the conjugacy class [111] of elements of finite order in SU(3) on the irreducible representation (p,q), $p\geqslant q$, $p-q=0$ mod 3, as given in Ref. 17, and used in Sec. IV of this paper. We may represent the EFO by R_1R_2, an element of $\mathbf{DT}\subset\mathrm{SU}(3)$ belonging to the **DT** class C_3. Since it has trace 0 on all but the point orbit, its trace for the point orbit is also given by the right-hand side of (A1). We show below that it is also given by the left-hand side of (A1).

The zero-weight space $V_{(p,q)}(0,0)$ is of dimension $q+1$. It is spanned by the $q+1$ vectors which can be written[19] as

$$|x\rangle=(\eta\eta^*)^x(\xi\xi^*)^{q-x}(\eta\xi\zeta)^{(p-q)/3},\quad x=0,1,\ldots,q\,,\qquad\text{(A2)}$$

where η,ξ,ζ are the three weight vectors of the SU(3) representation (1,0) of weights (1,0), $(-1,1)$, $(0,-1)$, respectively; η^*,ξ^*,ζ^* are the weight vectors of the representation (0,1) with weights $(-1,0)$, $(1,-1)$, (0,1), respectively. We eliminate ζ^* of weight (0,1) by means of the syzygy $\eta\eta^*+\xi\xi^*+\zeta\zeta^*=0$ (the scalar $\eta\eta^*+\xi\xi^*+\zeta\zeta^*$ never appears in these states). The action of R_1R_2 is to permute $\eta\xi\zeta$ and $\eta^*\xi^*\zeta^*$ cyclically. Thus (A2) becomes

$$\begin{aligned}R_1R_2|x\rangle&=(\xi\xi^*)^x(-\eta\eta^*-\xi\xi^*)^{q-x}(\eta\xi\zeta)^{(p-q)/3}\\&=(-1)^{q-x}(\eta\xi\zeta)^{(p-q)/3}\sum_{\alpha=0}^{q-x}(\xi\xi^*)^{q-\alpha}\\&\quad\times(\eta\eta^*)^\alpha\frac{(q-x)!}{\alpha!(q-\alpha-x)!}\\&=(-1)^{q-x}\sum_\alpha|\alpha\rangle\frac{(q-x)!}{\alpha!(q-\alpha-x)!}\,.\end{aligned}\qquad\text{(A3)}$$

The contribution of $|x\rangle$ to the trace is the coefficient of $|x\rangle$ on the right-hand side of (A3) and the complete trace is hence the left-hand side of (A1).

[1]M. Demazure and A. Grothendieck, "Séminaire de géometrie algébrique. Schémas en groupes," Notes multigraphiés, IHES 1963–64.
[2]J. Tits, "Normalizateurs de tores. 1. Groupes de Coxeter étendues," J. Algebra 4, 96 (1966).
[3]R. V. Moody and J. Patera, "General charge conjugation operators in simple Lie groups," J. Math. Phys. 25, 2838 (1984).
[4]I. G. Koh, J. Patera, and C. Rousseau, "Clebsch–Gordan coefficients for SU(5)⊆SU(3)×SU(2)×U(1) theories," J. Math. Phys. 24, 1955 (1983).
[5]I. G. Koh, J. Patera, and C.Rousseau, "Clebsch–Gordan coefficients for E(6) and SO(10) models of grand unification," J. Math. Phys. 25, 2863 (1984).
[6]M. del Olmo, J. Patera, M. Rodriguez, and C. Rousseau, "Clebsch–Gordan coefficients for SU(5) grand unified theories," J. Math. Phys. 28, 258 (1987).
[7]R. V. Moody and J. Patera, "Characters of elements of finite order in sim-

ple Lie groups," SIAM J. Algebraic Discrete Methods **5**, 359 (1984).
[8]R. V. Moody and J. Patera, "Computation of character decompositions of class functions on compact semisimple Lie groups," Math. Comput. **48**, 799 (1987).
[9]W. G. McKay, R. V. Moody, and J. Patera, "Tables of E_8 characters and decomposition of plethysms," in *Lie Algebras and Related Topics* (Am. Math. Soc., Providence, RI, 1986).
[10]W. G. McKay, R. V. Moody, and J. Patera, "Decomposition of tensor products of E_8 representations," Algebras Groups Geom. **3**, 286 (1986).
[11]A. Speiser, *Die theorie der Gruppen von endlichen Ordnung* (Vollständigen Gruppen, Birkhäuser, 1956), §42.
[12]R. V. Moody and J. Patera, "Fast recursion formula for weight multiplicities," Bull. Am. Math. Soc. **7**, 237 (1982).
[13]M. Bremner, R. V. Moody, and J. Patera, *Tables of Dominant Weight Multiplicities for Representation of Simple Lie Algebras* (Dekker, New York, 1985).
[14]A. J. Coleman, "The state labeling problem—A universal solution," J. Math. Phys. **27**, 1933 (1986).
[15]Equations (4.7) and (5.1) are the first published generating functions for multiplicities of dominant weights. They were derived using methods described fully in R. Gaskell, J. Patera, and R. T. Sharp, *Generating Functions in Group Representation Theory* (in preparation).
[16]J. Patera, R. T. Sharp, and P. Winternitz, "Polynomial irreducible tensors for point groups," J. Math. Phys. **19**, 2362 (1978).
[17]R. V. Moody, J. Patera, and R. T. Sharp, "Character generators for elements of finite order in simple Lie groups A_1, A_2, A_3, B_2, and G_2," J. Math. Phys. **24**, 2387 (1983).
[18]R. Gaskell and R. T. Sharp, "Generating functions for G_2 characters and subgroup branching rules," J. Math. Phys. **22**, 2736 (1981).
[19]R. T. Sharp and C. S. Lam, "Internal labeling problem," J. Math. Phys. **10**, 2033 (1969).

Superalgebras with Grassman-algebra-valued structure constants from superfields

José A. de Azcárraga
Departamento de Física Teórica, Facultad de Física, and IFIC, Centro Mixto Universidad de Valencia–C.S.I.C., E-46100-Burjasot, Valencia, Spain
Jerzy Lukierski[a)]
International Centre for Theoretical Physics, Trieste, Italy

(Received 22 September 1987; accepted for publication 9 December 1987)

Generalized Lie algebras and superalgebras with generators and structure constants taking values in a Grassmann algebra are introduced. Such algebraic structures describe the equal time algebras in the superfield formalism. As an example, the equal time commutators and anticommutators among bilinears made out of the $D = 1$ quantum superfields describing the supersymmetric harmonic oscillator are considered.

I. INTRODUCTION

Let us introduce the Grassmann algebra $\mathscr{G}_N$ generated by $\theta = (\theta_1,\ldots,\theta_N)$, where $\{\theta_i,\theta_j\} = 0$ $(i,j = 1,\ldots,N)$. The multiplet $J^a(\theta)$ $(a = 1,\ldots,m)$ of superfields is described by the mapping J^a: $\theta \in \mathscr{G}_N \rightarrow J^a \in \mathscr{G} = \mathscr{G}_F \otimes \mathscr{G}_N$, where $\mathscr{G}_F$ describes the Z_2-graded algebra of quantized bosonic and fermionic components, with even (odd) grading for bosonic (fermionic) ones. The expression of a superfield J^a in its components takes the following explicit form:

$$J^a(\theta) = j^a + j^a_{(i)}\theta_i + \cdots + j^a_{(i_1\cdots i_k)}\theta_{i_1}\cdots\theta_{i_k} + \tilde{j}^a\theta_1\cdots\theta_N. \tag{1}$$

The grading (parity) of $J^a(\theta)$ is defined as the product of parities in $\mathscr{G}_F$ and $\mathscr{G}_N$; we assume that the superfields $J^a(\theta)$ have definite parities. Thus we have (i) bosonic superfields, with $j^a_{(i_1\cdots i_k)}$ even (odd) if k is even (odd); or (ii) fermionic superfields, with $j^a_{(i_1\cdots i_k)}$ even (odd) if k is odd (even). A superfield $J^a(\theta)$ is bosonic (fermionic) if j^a is even (odd).

In a superfield description of supersymmetric quantum mechanics (QM) the components $j^a_{(i_1\cdots i_k)}$ are simply functions of time. If we consider the algebra of $D = 1$ superfields at fixed time (e.g., $t = 0$), the algebra $\mathscr{G}_F$ is finite.

The aim of this paper is to study the algebraic structures for the superfields $J^a(\theta)$ closed under commutation or anticommutation. For the sake of simplicity we shall suppress the space-time arguments of the component fields $j^a_{(i_1\cdots i_k)}$, which corresponds to (a) considering superfield algebras in supersymmetric QM at fixed time; and (b) considering $j^a_{(i_1\cdots i_k)}$ as component fields $j^a_{(i_1\cdots i_k)}(\mathbf{x},t)$ smeared out with a test function $f(\mathbf{x},t)$. In particular one can formally consider the improper limit, $f(\mathbf{x},t) \rightarrow \delta(t)$, in which the $J^a(\theta)$'s describe at fixed time the superfield extension of the conserved local current, integrated over $(D-1)$-dimensional space coordinates. In such a way one can introduce the notion of superfield of charges usually with only one component describing the "conventional" conserved global charge.

II. ALGEBRAIC STRUCTURES FOR CLOSED FINITE-DIMENSIONAL SUPERFIELD ALGEBRAS

The simplest Lie algebraic structure for bosonic superfields $J^a(\theta)$ can be written as follows (see also Mansouri[1]):

$$[J^a(\theta),J^b(\theta)] = f^{ab}{}_c J^c(\theta), \tag{2}$$

where $f^{ab}{}_d f^{dc}{}_e + \text{cycl}(a,b,c) = 0$, so that the $f^{ab}{}_c$ are the Lie algebra structure constants, and the components j^a in (1) describe a conventional finite-dimensional Lie algebra.

An essential feature of the algebra (2) is that it is generated by the commutators of operators depending on *the same* Grassmann coordinates. In contrast, the analog of a local current algebra in Grassmann space would be

$$[J^a(\theta),J^b(\theta')] = \delta^{(N)}(\theta-\theta')f^{ab}{}_c J^c(\theta'), \tag{3}$$

where

$$\delta^{(N)}(\theta-\theta') = \prod_{i=1}^{N}(\theta_i - \theta'_i).$$

Assuming, by analogy with the local current algebra for the quark model (see, e.g., Ref. 2), that the $J^a(\theta)$'s are bilinear in some basic superfields $\Phi_i(\theta)$, the local algebra (3) could be derived if the superfield description would allow for the canonical formalism in superspace. In such a case one would have

$$J^a(\theta) \sim \Phi_i(\theta)\lambda^a_{ij}\Pi_j, \tag{4}$$

where

$$\begin{aligned}[]
[\Phi_i(\theta),\pi_j(\theta')] &= i\delta^{(N)}(\theta-\theta')\cdot\delta_{ij},\\
[\Phi_i(\theta),\Phi_j(\theta')] &= [\Pi_i(\theta),\Pi_j(\theta')] = 0.
\end{aligned} \tag{5}$$

However, it is known[3–6] that superfields do not honor the canonical formalism, i.e., the superfields $\Phi_i(\theta)$ cannot be accompanied with canonical supermomenta $\Pi_i(\theta)$, satisfying the relations (5). As a result, the relation (3) does not seem to be useful in supersymmetric quantum-field theory (QFT), and one has to consider the following generalization:

[a)] On leave of absence from the Institute for Theoretical Physics, University of Wroclaw, 50-205 Wroclaw, Poland.

 0022-2488/88/040797-05$02.50

$$[J^a(\theta),J^b(\theta')]=\int d^N\eta\, f^{ab}{}_c(\theta,\theta',\eta)J^c(\eta), \tag{6}$$

where the integration over Grassmann variables is understood in the sense of Berezin,[7] and

$$\int d^N\eta\, f^{ab}{}_d(\theta,\theta',\eta)f^{dc}{}_e(\eta,\theta'',\eta') + \mathrm{cycl}[(a,\theta),(b,\theta'),(c,\theta'')]=0. \tag{7}$$

The algebra (6) can be extended to a superalgebra, provided that we supplement the even operators $J^a(\theta)$ with odd ones $S^\alpha(\theta)$. A first guess would be the following extension of (6):

$$\begin{aligned}[J^a(\theta),S^\alpha(\theta')]&=\int d^N\eta\, f^{a\alpha}{}_\beta(\theta,\theta',\eta)S^\beta(\eta),\\ \{S^\alpha(\theta),S^\beta(\theta')\}&=\int d^N\eta\, f^{\alpha\beta}{}_a(\theta,\theta',\eta\}J^a(\eta).\end{aligned} \tag{8}$$

However, in the presence of odd operators $S^\alpha(\theta)$, the relations (6) and (8) are not the most general ones, because they imply definite parity properties for $f^{ab}{}_c$, $f^{a\alpha}{}_\beta$, and $f^{\alpha\beta}{}_a$ in $\mathscr{G}^N$ (even for even N, odd for odd N). The relations (6) and (8) can be generalized as follows:

$$[J^a(\theta),J^b(\theta')]=\int d^N\eta[f^{ab}{}_c(\theta,\theta',\eta)J^c(\eta) + h^{ab}{}_\alpha(\theta,\theta',\eta)S^\alpha(\eta)], \tag{9a}$$

$$[J^a(\theta),S^\alpha(\theta')]=\int d^N\eta[f^{a\alpha}{}_\beta(\theta,\theta',\eta)S^\beta(\eta) + h^{a\alpha}{}_c(\theta,\theta',\eta)J^c(\eta)], \tag{9b}$$

$$\{S^\alpha(\theta),S^\beta(\theta')\}=\int d^N\eta[f^{\alpha\beta}{}_c(\theta,\theta',\eta)J^c(\eta) + h^{\alpha\beta}{}_\gamma(\theta,\theta',\eta)S^\gamma(\eta)], \tag{9c}$$

where the three different types of f's (h's), which have an even (odd) number of Greek indices are even (odd) functions of the Grassmann variables if N is even; if N is odd, their parity is the opposite. The position of the Greek and Latin indices should be noticed: for instance, $f^{a\alpha}{}_\beta$ and $f^{\alpha\beta}{}_c$ have the same parity, but they multiply the generators with opposite gradings in (9b) and (9c), respectively. The f's and the h's satisfy the obvious relation, e.g.,

$$f^{uv}{}_w(\theta,\theta',\eta)=-(-1)^{\deg u\cdot\deg v}f^{vu}{}_w(\theta',\theta,\eta), \tag{10}$$

where $\deg u$ is the grading of the index $u=a,b,\ldots$ or $\alpha,\beta,\ldots$ [$\deg u=0$ (1) Latin (Greek) indices].

Using graded Jacobi identities

$$[[J^a(\theta),J^b(\theta')],J^c(\theta'')] + \mathrm{cycl}\{(a\theta),(b\,\theta'),(c\theta'')\}=0, \tag{11a}$$

$$[[S^\alpha(\theta),J^b(\theta')],J^c(\theta'')] + \mathrm{cycl}\{(\alpha\theta),(b\,\theta'),(c\theta'')\}=0, \tag{11b}$$

$$[\{S^\alpha(\theta),S^\beta(\theta')\},S^\alpha(\theta'')] + \mathrm{cycl}\{(\alpha\theta),(\beta\,\theta'),(\beta\,\theta'')\}=0, \tag{11c}$$

and

$$[\{S^\alpha,S'^\beta\},J''^c]+\{[J''^c,S^\alpha],S'^\beta\}-\{[S'^\beta,J''^c],S^\alpha\}=0, \tag{11d}$$

and assuming irreducibility for J^a and S^α, one obtains, respectively, the following identities for the structure functions $f^{uv}{}_w$ and $h^{uv}{}_w$:

$$\left.\begin{aligned}&\int d^N\eta[f^{ab}{}_d(\theta,\theta',\eta)f^{dc}{}_e(\eta,\theta'',\eta') + h^{ab}{}_\epsilon(\theta,\theta',\eta)h^{\epsilon c}{}_e(\eta,\theta'',\eta')] + \mathrm{cycl}\{(a\theta),(b\,\theta'),(c\theta'')\}=0,\\ &\int d^N\eta[f^{ab}{}_d(\theta,\theta',\eta)h^{dc}{}_\beta(\eta,\theta'',\eta') + h^{ab}{}_\epsilon(\theta,\theta',\eta)f^{\epsilon c}{}_\beta(\eta,\theta'',\eta')] + \mathrm{cycl}\{(a\theta),(b\,\theta'),(c\theta'')\}=0;\end{aligned}\right\} \tag{12a}$$

$$\left.\begin{aligned}&\int d^N\eta[f^{ab}{}_u(\theta,\theta',\eta)f^{uc}{}_\gamma(\eta,\theta'',\eta') + h^{ab}{}_u(\theta,\theta',\eta)h^{uc}{}_\gamma(\eta,\theta'',\eta')] + \mathrm{cycl}\{(\alpha\theta),(b\,\theta'),(c\theta'')\}=0,\\ &\int d^N\eta[f^{ab}{}_u(\theta,\theta',\eta)h^{uc}e(\eta,\theta'',\eta') + h^{ab}{}_u(\theta,\theta',\eta)f^{uc}e(\eta,\theta'',\eta')] + \mathrm{cycl}\{(\alpha\theta),(b\,\theta'),(c\theta'')\}=0;\end{aligned}\right\} \tag{12b}$$

$$\left.\begin{aligned}&\int d^N\eta[f^{\alpha\beta}{}_d(\theta,\theta',\eta)f^{d\gamma}{}_\xi(\eta,\theta'',\eta') + h^{\alpha\beta}{}_\epsilon(\theta,\theta',\eta)h^{\epsilon\gamma}{}_\xi(\eta,\theta'',\eta')] + \mathrm{cycl}\{(\alpha\theta),(\beta\,\theta'),(\gamma\theta'')\}=0,\\ &\int d^N\eta[f^{\alpha\beta}{}_d(\theta,\theta',\eta)h^{d\gamma}{}_e(\eta,\theta'',\eta') + h^{\alpha\beta}{}_\epsilon(\theta,\theta',\eta)f^{\epsilon\gamma}{}_e(\eta,\theta'',\eta')] + \mathrm{cycl}\{(\alpha\theta),(\beta\,\theta'),(\gamma\theta'')\}=0;\end{aligned}\right\} \tag{12c}$$

$$\left.\begin{aligned}&\int d^N\eta[f^{\alpha\beta}{}_u(\theta,\theta',\eta)f^{uc}{}_e(\eta,\theta'',\eta') + h^{\alpha\beta}{}_u(\theta,\theta',\eta)h^{uc}{}_e(\eta,\theta'',\eta')] + \mathrm{grad\ cycl}\{(\alpha\theta),(\beta\,\theta'),(c\theta'')\}=0,\\ &\int d^N\eta[f^{\alpha\beta}{}_u(\theta,\theta',\eta)h^{uc}{}_\gamma(\eta,\theta'',\eta') + h^{\alpha\beta}{}_u(\theta,\theta',\eta)f^{uc}{}_\gamma(\eta,\theta'',\eta')] + \mathrm{grad\ cycl}\{(\alpha\theta),(\beta\,\theta'),(c\theta'')\}=0;\end{aligned}\right\} \tag{12d}$$

where η and η' indicate the integration variables and (i) in formulas (12b) and (12d) the index u is Greek or Latin depending on the permutation, in such a way that the f's (h's) have always even (odd) number of Greek indices; and (ii) in formulas (12d) graded cyclic means that $(\alpha\theta)$, $(\beta\theta')$, $(c\theta'')$ and $(c\theta'')$, $(\alpha\theta)$, $(\beta\theta')$ have a $+$ sign and $(\beta\theta'),(c\theta''),(\alpha\theta)$ a $-$ one (β jumps over α).

It turns out that by calculating the equal time (E.T.) commutators of bilinear products of free superfields the superalgebra (9a)–(9c) emerges provided that we consider products *bilocal* in the Grassmann variables. Therefore in superfield applications the Grassmann algebra $\mathscr{G}_N$ is described by a graded tensor product of *two copies* of Grassman algebras describing the anticommuting superspace coordinates.

III. E.T. ALGEBRAS FOR SUPERFIELDS DESCRIBING THE SUPERSYMMETRIC HARMONIC OSCILLATOR

We shall realize below the superalgebra (9a)–(9c) in terms of the bilocal bilinears of $D=1$ superfields, which describe the SUSY harmonic oscillator with complex supersymmetries.[3,8,9] The action for Witten's supersymmetric QM in superspace has the form

$$I=\int dt\, d\bar{\theta}\, d\theta \times\left[-\tfrac{1}{4}(\bar{D}\Phi_i\, D\Phi_i - D\Phi_i\, \bar{D}\Phi_i) - V(\Phi_i^2)\right], \tag{13}$$

where

$$D=\frac{\partial}{\partial\theta}-i\bar{\theta}\frac{\partial}{\partial t},\quad \bar{D}=\frac{\partial}{\partial\bar{\theta}}-i\theta\frac{\partial}{\partial t}, \tag{14}$$

and the Φ_i are real superfields ($i=1,...,m$)

$$\Phi_i(t,\theta,\bar{\theta})=X_i(t)+i\bar{\theta}\psi_i(t)+i\theta\bar{\psi}_i(t)+\theta\bar{\theta}A_i(t). \tag{15}$$

Putting $V(\Phi_i^2)=\frac{1}{2}k\Phi_i^2$ one gets the action for supersymmetric oscillator, and Φ_i satisfies the equation

$$[\tfrac{1}{2}(\bar{D}D-D\bar{D})-k]\Phi_i(t,\theta,\bar{\theta})=0. \tag{16}$$

Using the identity

$$[\tfrac{1}{2}(\bar{D}D-D\bar{D})-k][\tfrac{1}{2}(\bar{D}D-D\bar{D})+k] \times(\partial_t^2+k^2)^{-1}=1, \tag{17}$$

one gets the following formula for the supercommutator[9]:

$$\begin{aligned}\Delta_{ij}(t-t';\theta,\bar{\theta},\theta',\bar{\theta}') &= (1/i)[\Phi_i(t,\theta,\bar{\theta}),\Phi_j(t,\theta',\bar{\theta}')]\\ &= \delta_{ij}[\tfrac{1}{2}(\bar{D}D-D\bar{D})+k]\{[\sin k(t'-t)]/k\}\delta(\theta-\theta')\delta(\bar{\theta}-\bar{\theta}')\\ &= \delta_{ij}(\{[\sin k(t'-t)]/k\}[1+k(\theta-\theta')(\bar{\theta}-\bar{\theta}')+k^2(\theta\bar{\theta}\theta'\bar{\theta}')]+\cos k(t'-t)(\theta\bar{\theta}'+\bar{\theta}\theta')).\end{aligned} \tag{18}$$

From (18) one gets the following nonvanishing E.T. graded commutators of Φ_i, $D\Phi_i$, and $\bar{D}\Phi_i$:

$$\begin{aligned}&[\Phi_i(t,\theta,\bar{\theta}),\Phi_j(t,\theta',\bar{\theta}')]=\delta_{ij}(\theta\bar{\theta}'+\bar{\theta}\theta'),\\ &[\bar{D}\Phi_i(t,\theta,\bar{\theta}),\Phi_j(t,\theta',\bar{\theta}')]\\ &\quad=\delta_{ij}[-(\theta-\theta')-k(\theta\theta'\bar{\theta}'+\theta'\theta\bar{\theta})],\\ &[D\Phi_i(t,\theta,\bar{\theta}),\Phi_j(t,\theta',\bar{\theta}')]\\ &\quad=\delta_{ij}[-(\bar{\theta}-\bar{\theta}')-k(\bar{\theta}\theta'\bar{\theta}'+\bar{\theta}'\theta\bar{\theta})],\\ &\{\bar{D}\Phi_i(t,\theta,\bar{\theta}),D\Phi_j(t,\theta',\bar{\theta}')\}\\ &\quad=\delta_{ij}[-1+k(\theta\bar{\theta}+\theta'\bar{\theta}'-2\theta\bar{\theta}')-k^2\theta\bar{\theta}\theta'\bar{\theta}'].\end{aligned} \tag{19}$$

The E.T. superfield graded commutators (supercommutators) (19) are fully equivalent to the canonical E.T. relations for the fields $X_i(t)$, $\psi_i(t)$, and $\bar{\psi}_i(t)$; the fields $A_i(t)=-kX_i(t)$ are auxiliary. Formulas (19) describe an extension of the Heisenberg algebra in the spirit of our generalization, i.e., with the numerical constants describing quantization conditions represented by the elements of Grassmann algebra ("Grassmann numbers"). The E.T. supercommutators (19) have the following new features with respect to the conventional canonical formalism.

(a) They are "nonlocal" in the Grassmann variables (i.e., they are *not* proportional to Dirac deltas in Grassmann variables).

(b) The E.T. odd commutators $[J,S]$ among bosonic ($J=\Phi_i$) and fermionic ($S=D\Phi_i,\bar{D}\Phi_i$) superfields do not vanish. In contrast, the E.T. commutators of a bosonic and a fermionic canonical variable (e.g., $[X_i(t),\psi_i(t)]$) are all zero.

The nonlocality of E.T. supercommutators (19) implies that the superalgebra (9a)–(9c) can only be realized in terms of bilinears in superfields that are *nonlocal* in the Grassman variables. The property $[J,S]\neq 0$ implies that the Grassmann-valued odd structure constants h are different from zero.

From the superfields Φ_i, $D\Phi_i$, and $\bar{D}\Phi_i$ the following examples of closed algebraic structures may be constructed.

(1) Let us consider

$$J(t,\theta,\bar{\theta},\theta',\bar{\theta}')=\Phi_i(t,\theta,\bar{\theta})\Phi_i(t,\theta',\bar{\theta}'). \tag{20}$$

The E.T. supercommutators of the operators (20) provide an example of the algebra (6).

(2) The E.T. supercommutators of the bilinear products,

$$\begin{aligned}J_+(t,\theta,\bar{\theta},\theta',\bar{\theta}')&=D\Phi_i(t,\theta,\bar{\theta})D\Phi_i(t,\theta',\bar{\theta}'),\\ J_0(t,\theta,\bar{\theta},\theta',\bar{\theta}')&=D\Phi_i(t,\theta,\bar{\theta})\bar{D}\Phi_i(t,\theta'\bar{\theta}'),\\ J_-(t,\theta,\bar{\theta},\theta',\bar{\theta}')&=\bar{D}\Phi_i(t,\theta,\bar{\theta})\bar{D}\Phi_i(t,\theta',\bar{\theta}'),\end{aligned} \tag{21}$$

generate another example of the algebra (6).

(3) If we consider the commutators $[J,J_+]$ we obtain an example of relation (9a) with $f^{ab}{}_c=0$ and $h^{ab}{}_\alpha\neq 0$, where the odd operators are described by

$$\begin{aligned}S_+(t,\theta,\bar{\theta},\theta',\bar{\theta}')&=\Phi_i(t,\theta,\bar{\theta})D\Phi_i(t,\theta',\bar{\theta}'),\\ S_-(t,\theta,\bar{\theta},\theta',\bar{\theta}')&=\Phi_i(t,\theta,\bar{\theta})\bar{D}\Phi_i(t,\theta',\bar{\theta}').\end{aligned} \tag{22}$$

(4) The six bilinears (20)–(22) provide an example of E.T. algebra described by the set of relations (9a)–(9c). In particular the E.T. supercommutators $[S_\pm, J_0]$ provide examples of the relation (9b) with $f\neq 0$ and $h\neq 0$, and the superanticommutators $\{S_\pm,S_\pm\}$ and $\{S_\pm,S_\mp\}$ provide an example of the relations (9c), also with $f\neq 0$ and $h\neq 0$.

For simplicity in the bilinears (20)–(22) we have ignored the possible matrix insertions, which could generate a more complicated form of the algebra (9). For example, we could replace (20) by

$$J^a(t,\theta,\bar{\theta},\theta',\bar{\theta}')=\Phi_i(t,\theta,\bar{\theta})\lambda^a_{ij}\Phi_j(t,\theta',\bar{\theta}'), \tag{23}$$

where, for instance, the 3×3 matrices λ^a describe the generators of $U(3)$, closed under matrix multiplication, i.e., $\lambda^a\lambda^b=c^{abd}\lambda^d$, where c^{abd} are real.

IV. THREE DIFFERENT FORMS OF THE SUPERALGEBRA (9a)–(9c)

A. The algebra of local products of superfields

We see from the example in Sec. III that the basic superalgebra (9a)–(9c) is obtained if J^a and S^α are the products of superfields *bilocal* in the Grassmann sector [see, e.g., (23)]. One can use, however, a Taylor expansion in the Grassmann variables, e.g.,

$$\Phi_i(t,\theta',\bar\theta') = \Phi_i(t,\theta,\bar\theta) + (\theta'-\theta)\frac{\partial}{\partial\theta}\Phi_i(t,\theta,\bar\theta) + (\bar\theta'-\bar\theta)\frac{\partial}{\partial\bar\theta}\Phi_i(t,\theta,\bar\theta) + (\bar\theta'-\bar\theta)(\theta'-\theta)\frac{\partial}{\partial\theta}\frac{\partial}{\partial\bar\theta}\Phi_i(t,\theta,\bar\theta), \tag{24}$$

and write

$$J^a(t,\theta,\bar\theta,\theta',\bar\theta') = J^a_{(0,0)}(t,\theta,\bar\theta) + (\theta'-\theta)J^a_{(1,0)}(t,\theta,\bar\theta) + (\bar\theta'-\bar\theta)J^a_{(0,1)}(t,\theta,\bar\theta) + (\bar\theta'-\bar\theta)(\theta'-\theta)J^a_{(1,1)}(t,\theta,\bar\theta), \tag{25}$$

where $(i,j=0,1)$

$$J^a_{(i,j)}(t,\theta,\bar\theta) = \left(\frac{\partial}{\partial\theta'}\right)^i\left(\frac{\partial}{\partial\bar\theta'}\right)^j J^a(t,\theta,\bar\theta,\theta',\bar\theta')\Big|_{\substack{\theta'=\theta\\ \bar\theta'=\bar\theta}}. \tag{26}$$

In such a way one can replace, e.g., the relation (6) for the *partly bilocal* product (23) by the superalgebra (9a)–(9c) for the *local* products (26), with $J^a_{(0,0)}$ and $J^a_{(1,1)}$ describing bosonic and $J^a_{(1,0)}$, $J^a_{(0,1)}$ describing fermionic local superfields. It is easy to check that in the general case of N anticommuting coordinates in superspace every bilocal superfield $J(\theta,\theta')$ is described by 2^N local superfields.

B. The generalization of the algebra (2)

Let us consider the following generalization of the algebra (2):

$$[J^a_{\hat i}(\boldsymbol\theta),J^b_{\hat j}(\boldsymbol\theta)] = F^{ab}{}_c(\hat i,\hat j,\hat k;\boldsymbol\theta)J^c_{\hat k}(\boldsymbol\theta) + H^{ab}{}_\alpha(\hat i,\hat j,\hat r;\boldsymbol\theta)S^\alpha_{\hat r}(\boldsymbol\theta), \tag{27a}$$

$$[J^a_{\hat i}(\boldsymbol\theta),S^\alpha_{\hat r}(\boldsymbol\theta)] = F^{a\alpha}{}_\beta(\hat i,\hat r,\hat s;\boldsymbol\theta)S^\beta_{\hat s}(\boldsymbol\theta) + H^{a\alpha}{}_c(\hat i,\hat r,\hat k;\boldsymbol\theta)J^c_{\hat k}(\boldsymbol\theta), \tag{27b}$$

$$\{S^\alpha_{\hat r}(\boldsymbol\theta),S^\beta_{\hat s}(\boldsymbol\theta)\} = F^{\alpha\beta}{}_c(\hat r,\hat s,\hat k;\boldsymbol\theta)J^c_{\hat k}(\boldsymbol\theta) + H^{\alpha\beta}{}_\gamma(\hat r,\hat s,\hat t;\boldsymbol\theta)S^\gamma_{\hat t}(\boldsymbol\theta). \tag{27c}$$

Expanding, $J^a(\boldsymbol\theta')$, $S^\alpha(\boldsymbol\theta')$ from (9a)–(9c) in the powers of $\theta'_i-\theta_i$, e.g.,

$$J^a(\boldsymbol\theta') = \sum_{k=0}^N \frac{1}{k!}\sum_{i_1\cdots i_k}(\theta'_{i_1}-\theta_{i_1})\cdots(\theta'_{i_k}-\theta_{i_k}) \times J^a_{(i_1\cdots i_k)}(\boldsymbol\theta), \tag{28a}$$

where

$$J^a_{(i_1\cdots i_k)}(\boldsymbol\theta) = \frac{\partial}{\partial\theta_{i_1}}\cdots\frac{\partial}{\partial\theta_{i_k}}J^a(\boldsymbol\theta) \tag{28b}$$

one can replace the superalgebra (9a)–(9c) by the finite superalgebra (27a)–(27c) of the superfields $J^a_{(i_1\cdots i_k)}(\boldsymbol\theta)$, $S^\alpha_{(i_1\cdots i_k)}(\boldsymbol\theta)$, where

$$J^a_{\hat i}(\boldsymbol\theta) = \{J^a_{(i_1\cdots i_k)}(\boldsymbol\theta)(k \text{ even}), S^\alpha_{(i_1\cdots i_k)}(\boldsymbol\theta)(k\text{ odd})\},$$
$$S^\alpha_{\hat r}(\boldsymbol\theta) = \{J^a_{(i_1\cdots i_k)}(\boldsymbol\theta)(k\text{ odd}), S^\alpha_{(i_1\cdots i_k)}(\boldsymbol\theta)(k\text{ even})\}. \tag{29}$$

By differentiating the relations (9a)–(9c) with respect to the θ'_i variables and substituting the Taylor expansion of $J^c(\boldsymbol\eta)$, $S^\alpha(\boldsymbol\eta)$ around the point $\boldsymbol\eta=\boldsymbol\theta$ [as in (28a)] one can express the structure constants $F^{uv}{}_w$, $H^{uv}{}_w$ in (27a)–(27c) in terms of the $f^{uv}{}_w$, $h^{uv}{}_w$ taken from (9a)–(9c) [$u,v,w = (a,\alpha)$].

C. Conventional form of the superalgebra

First we observe that the algebra (6) can be reexpressed as a large superalgebra in the components $j^a_{(i_1\cdots i_k)}$ [see (1)]. Indeed, using expansion (1) for the Grassmann-valued structure constants, e.g.,

$$f^{ab}{}_c(\boldsymbol\theta,\boldsymbol\theta',\boldsymbol\eta) = \sum_{l,m,n=0}\sum_{i_1\cdots i_l}\sum_{j_1\cdots j_m}\sum_{k_1\cdots k_n} f^{ab}{}_c(i_1\cdots i_l;j_1\cdots j_m;k_1\cdots k_n) \times\theta_{i_1}\cdots\theta_{i_l}\cdot\theta'_{j_1}\cdots\theta'_{j_m}\cdot\eta_{k_1}\cdots\eta_{k_n}, \tag{30}$$

and the formula

$$\int d^N\theta\,\theta_{i_1}\cdots\theta_{i_k}J^a(\boldsymbol\theta) = \epsilon_{i_1\cdots i_N}j^a_{(i_{k+1}\cdots i_N)}, \tag{31}$$

one gets from (6) the following conventional superalgebra with numerical structure constants:

$$[j^a_{(i_1\cdots i_l)},j^b_{(j_1\cdots j_m)}]_\pm = \sum_{n=0}^N\sum_{k_1\cdots k_n} f^{ab}{}_c(i_1\cdots i_l,j_1\cdots j_m,k_1\cdots k_n) \times\epsilon_{k_1\cdots k_nk_{n+1}\cdots k_N}j^c_{(k_{n+1}\cdots k_N)}, \tag{32}$$

where anticommutators are assumed if both l and m are odd, and commutators otherwise. The Jacobi identities (7) can be written as the following set of bilinear relations for the numerical structure constants defined by the relation (30):

$$\sum_{n=0}^N f^{ab}{}_d(i_1\cdots i_l,j_1\cdots j_m,k_1\cdots k_n)\epsilon_{k_1\cdots k_nk_{n+1}\cdots k_N}f^{dc}{}_e \times(k_{n+1}\cdots k_N,s_1\cdots s_r,u_1\cdots u_t) + \text{cycl}\left[\begin{pmatrix}a\\ i_1\cdots i_l\end{pmatrix},\begin{pmatrix}b\\ j_1\cdots j_m\end{pmatrix},\begin{pmatrix}c\\ s_1\cdots s_r\end{pmatrix}\right]=0. \tag{33}$$

An analogous rewriting of the superalgebra (9a)–(9c) and the graded Jacobi identities (12a)–(12d) in its component form is tedious but straightforward.

V. FINAL REMARKS

Our example of the superalgebra (9a)–(9c) is constructed out of the E.T. bilinears of $D=1$ *free* superfields. In the presence of *interactions* the basic relations (19) are modified—some Grassmann numbers are replaced by the operator terms (the second derivatives of the superpotential $V(\Phi_i^2)$ [see (13)]). In such a case the E.T. supercommutators of the bilinears (20)–(22) cease to form a closed superalgebraic system (9a)–(9c).

In the presence of interactions describing an asymptotically free theory, one can introduce the fully bilocal products of superfields (also bilocal in the space-time coordinates). If $D > 1$, one can consider their graded supercommutators for the differences of superspace coordinates lying on the superlight cone. The postulate that such a superalgebra closes leads to the supersymmetric extension of $D = 4$ Fritzsch–Gell-Mann algebra for bilocal internal symmetry currents.[10]

ACKNOWLEDGMENTS

One of the authors (J. L.) would like to acknowledge the hospitality at The Department of Theoretical Physics of Valencia University. He would also like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste.

The research of the other author (J. A. A.) was partially supported by a CICYT grant (#AE-0011-C03-02).

[1]F. Mansouri, J. Math. Phys. **18**, 2395 (1977).
[2]B. Renner, *Current Algebra and their Applications* (Pergamon, New York, 1968).
[3]J. A. de Azcárraga, J. Lukierski, and P. Vindel, Mod. Phys. Lett. A **1**, 293 (1986).
[4]J. Barcelos-Neto and Ashok Das, Phys. Rev. D **33**, 2863 (1986).
[5]J. Barcelos-Neto, A. Das, and W. Scherer, Phys. Rev. D **34**, 1108 (1986).
[6]J. A. de Azcárraga, J. Lukierski, and P. Vindel, Fortschr. Phys., in press.
[7]F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966); Sov. J. Nucl. Phys. **30**, 605 (1979).
[8]E. Witten, Nucl. Phys. B **188**, 513 (1981).
[9]J. A. de Azcárraga, J. Lukierski, and P. Vindel, *Fields and Geometry 1986*, Proceedings of XXII Karpacz Winter School and Workshop, February 1986, edited by A. Jadczyk (World Scientific, Singapore, 1986), p. 646.
[10]J. A. de Azcárraga, J. Lukierski, and P. Vindel, Phys. Lett. B **195**, 195 (1987).

On the spectrum of a third-order SO(3) scalar in the enveloping algebra of SO(6)

J. Vanthournout, J. Van der Jeugt,[a] H. De Meyer,[b] and G. Vanden Berghe
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit Gent, Krijgslaan 281-gebouw S9, B9000 Gent, Belgium

(Received 29 September 1987; accepted for publication 25 November 1987)

With the aid of previously derived expressions for the SO(3) reduced matrix elements of the SO(6) generators, which were obtained by considering an intermediate SO(6)↓SU(2) ⊗ SU(2) reduction, a method is set up to evaluate analytic expressions for the eigenvalues of a third-order scalar operator belonging to the integrity basis of SO(3) scalar operators in the enveloping algebra of SO(6).

I. INTRODUCTION

In two previous papers[1,2] it has been shown that within the SU(3) limit, symmetry-conserving higher-order interaction terms influence the energy spectrum within the interacting boson model (IBM) quite seriously. Three- and four-body interactions introduced in the IBM Hamiltonian by Vanden Berghe *et al.*[1] gave rise to a much better approximation of the energy spectra as well as to the removal of the degeneracy, which originally existed for members of the so-called β and γ bands. By this success it is tempting to investigate the analogous problem in the two other IBM limits, SO(6) and SU(5). In this scope it is natural to ask for SO(3) scalars that preserve both mentioned symmetries. It has been proved recently[3] that, when applied to symmetric irreps, there exists in the integrity basis of the corresponding algebras only two functionally independent third-order SO(3) scalars. In this paper special attention will be paid to the SO(6) limit and to one of these scalars in particular.

Denoting in the SO(3) basis the SO(6) Lie algebra by the SO(3) basis elements $l_0, l_{\pm 1}$, together with the components p_μ ($\mu \in [-2,2]$) of a five-dimensional irreducible SO(3) tensor operator, and the components q_μ ($\mu \in [-3,3]$) of a seven-dimensional SO(3) tensor operator, the third-order functionally independent SO(3) scalars can be expressed as[3]

$$\Gamma_1 = ((pp)^2 p)^{(0)} \tag{1.1}$$

and

$$\Gamma_2 = ((pl)^1 l)^{(0)}. \tag{1.2}$$

It is the aim of the present paper to derive for several classes of SO(6) physical basis states closed formulas for the Γ_2 eigenvalues by making use of a previously derived expression[4] for the reduced matrix elements of p. Because of the complex structure of the operator Γ_1, closed formulas are very difficult to derive. As a by-product we shall discuss the possibility of using the p-reduced matrix elements for the derivation of compact expressions of $B(E2)$ values between SO(6) basis states.

II. BASIC FORMULAS

Using the explicit expression (1.2) for the operator Γ_2, applying the Wigner–Eckart theorem, and introducing the analytic expression for the occurring $3j$ symbol, it is easy to verify that the matrix elements of Γ_2, with respect to the SO(6) totally symmetric irreps $|\sigma\tau\nu l\rangle$ can be written as

$$\langle \sigma\tau'\nu' lm|\Gamma_2|\sigma\tau\nu lm\rangle = [l(l+1)(2l+3)(2l-1)/2.3.5(2l+1)]^{1/2} \times \langle \sigma\tau'\nu' l\,\|p\|\sigma\tau\nu l\rangle. \tag{2.1}$$

Herein we have denoted for the SO(6) symmetric irrep $[\sigma,0,0]$ the states that constitute the physical basis as

$$|\sigma\tau\nu lm\rangle, \tag{2.2}$$

with $m = -l, -l+1,\ldots,l$. The labels σ, τ, l, and m are, respectively, related to the Casimir operator eigenvalues of the algebras SO(6), SO(5), SO(3), and SO(2). Let us remember that the SO(5) label τ takes on the values

$$\tau = \sigma, \sigma-1, \sigma-2,\ldots,0, \tag{2.3}$$

while the extra label ν is given by

$$\nu = 0,1,2,\ldots,[\tau/3]. \tag{2.4}$$

For a given ν value, the SO(3) content is specified by

$$l = 2K, 2K-2, 2K-3,\ldots,K+1, K, \tag{2.5}$$

with

$$K = \tau - 3\nu. \tag{2.6}$$

It should also be remarked that the states (2.2) are orthogonal in all labels except for the ν label. More explicitly,

$$\langle \sigma\tau' l' m'|\sigma\tau lm\rangle = A_l^\tau(\nu',\nu)\delta_{\tau'\tau}\delta_{l'l}\delta_{m'm}. \tag{2.7}$$

Several equivalent formulas that permit one to evaluate the overlap integrals $A_l^\tau(\nu',\nu)$ have been derived by Williams and Pursey.[5] For further discussion, we mention one of the expressions here:

[a] Senior Research Assistant N. F. W. O. (Belgium). Present address: Department of Mathematics, University of Southampton, Southampton SO9 5NH, England.
[b] Senior Research Associate N. F. W. O. (Belgium).

$$A_l^\tau(\nu',\nu)=2^{\nu'-\nu}[(2l+1)(3\nu'-3\nu)!]^{-1}[(\tau-\nu)!(\tau-\nu')!\nu!\nu'!(l+3\nu'-\tau)!$$
$$\times(l+\tau-3\nu)!/(l+\tau-3\nu')!(l+3\nu-\tau)!]^{1/2}\sum_{\alpha,\beta}(-4)^{\nu+\alpha-\beta}(3\nu'-3\beta+\alpha)!(2\tau-2\nu-2\nu'+2\beta)!$$
$$\times[(\tau-\nu-\nu'+\beta-\alpha)!(\nu'-\beta)!(\nu-\beta)!\alpha!\beta!(2\tau+\nu'-2\nu+\alpha-\beta+1)!]^{-1}$$
$$\times{}_3F_2(\tau-3\nu-l,\tau+l-3\nu+1,3\nu'-3\beta+\alpha+1;3\nu'-3\nu+1,2\tau+\nu'-2\nu-\beta+\alpha+2;1). \tag{2.8}$$

The reduced matrix element of p with respect to the physical basis states (2.2) has been the subject of another paper.[4] For the sake of self-containment we mention the result here:

$$(2l'+1)^{-1/2}\langle\sigma\tau'\nu'l'\|p\|\sigma\tau\nu l\rangle$$
$$=\delta_{\tau',\tau+1}[(\sigma-\tau)(\sigma+\tau+4)/(2\tau+5)]^{1/2}\{-[l'(l'+1)/3(2\tau+3)][\nu/(\tau-\nu+1)(\tau-\nu+2)]^{1/2}$$
$$\times\langle lK22|l'K+2\rangle\langle l'K+4\,1\;-1|l'K+3\rangle\langle l'K+3\,1\;-1|l'K+2\rangle A_{l'}^{\tau+1}(\nu',\nu-1)+(\tau-\nu+1)^{-1/2}$$
$$\times[-[l'(l'+1)/2]^{1/2}(2\tau-2\nu+3)/(2\tau+3)\langle lK22|l'K+2\rangle\langle l'K+1\,1\,1|l'K+2\rangle$$
$$+(\tau-\nu+1)\langle lK21|l'K+1\rangle+[l'(l'+1)/3]^{1/2}\langle lK20|l'K\rangle\langle l'K+1\,1\;-1|l'K\rangle$$
$$+(l'(l'+1)/3(2\tau+3))\langle lK2\;-1|l'K-1\rangle\langle l'K+1\,1\;-1|l'K\rangle\langle l'K1\;-1\,l'K-1\rangle]A_{l'}^{\tau+1}(\nu',\nu)$$
$$+(\nu+1)^{1/2}[-[2l'(l'+1)]^{1/2}/(2\tau+3)\langle lK2\;-1|l'K-1\rangle\langle l'K-2\,1\,1|l'K-1\rangle$$
$$+\langle lK2\;-2|l'K-2\rangle]A_{l'}^{\tau+1}(\nu',\nu+1)\}+\delta_{\tau',\tau-1}[(\sigma-\tau+1)(\sigma+\tau+3)/(2\tau+3)]^{1/2}$$
$$\times\{\nu^{1/2}\langle lK22|l'K+2\rangle A_{l'}^{\tau-1}(\nu',\nu-1)-(\tau-\nu)^{1/2}\langle lK2\;-1|l'K-1\rangle A_{l'}^{\tau-1}(\nu',\nu)\}. \tag{2.9}$$

It is clear that these reduced matrix elements only differ from zero when $\tau'=\tau\pm1$.

III. Γ_2 EIGENVALUES

Before discussing the evaluation of the different eigenvalues it is worthwhile to define some classes in which the SO(6) physical basis states (2.2), with a fixed but large enough σ value, can be subdivided with respect to their l contents. For this classification we make use of the restrictions imposed by Eqs. (2.3)–(2.6) on the occurring labels.

Class (a): l nondegenerated states: $l=2\sigma,2\sigma-3$ ($\tau=\sigma$, $\nu=0$).
Class (b): l doubly degenerated states: $l=2\sigma-2,2\sigma-5$ ($\tau=\sigma$, $\nu=0$; $\tau=\sigma-1$, $\nu=0$).
Class (c): l triple-degenerated states: $l=2\sigma-4,2\sigma-7$ ($\tau=\sigma$, $\nu=0$; $\tau=\sigma-1$, $\nu=0$; $\tau=\sigma-2$, $\nu=0$).
Class (d): l quintuple-degenerated states: $l=2\sigma-6,2\sigma-9$ ($\tau=\sigma$, $\nu=0$; $\tau=\sigma$, $\nu=1$; $\tau=\sigma-1$, $\nu=0$; $\tau=\sigma-2$, $\nu=0$; $\tau=\sigma-3$, $\nu=0$).

All other possible states are at least sevenfold degenerated. This has as a consequence that the secular equation, giving rise to the eigenvalues of Γ_2, becomes a polynomial equation of sixth degree or higher. This means that an analytic treatment for these cases is excluded; a numerical solution is, however, still possible.

A. Γ_2 eigenvalues for states of class (a)

Since these states are uniquely described by one physical basis state of the type (2.2), the Γ_2 eigenvalue is, because of the relation (2.1), linearly correlated with the reduced matrix element of p. Since the reduced matrix elements (2.9) vanish for $\tau'=\tau$, it is clear that for all σ,

$$\Gamma_2(l=2\sigma)=\Gamma_2(l=2\sigma-3)=0. \tag{3.1}$$

Hereafter, we denote the eigenvalue of Γ_2, corresponding to an eigenstate with angular momentum l, as $\Gamma_2(l)$.

B. Γ_2 eigenstates for states of class (b)

For this class of doubly degenerated states, we have at our disposal two basis states of the type (2.2), i.e., the one with $\tau=\sigma$, $\nu=0$ and the other with $\tau=\sigma-1$, $\nu=0$. These states are not normalized and are not orthogonal to each other. Let us, as an example, demonstrate how the secular equation for this specific class can be derived.

An eigenvector $|lm\rangle\rangle$ of the operator Γ_2 will be a linear combination of the two basis states available, i.e.,

$$|l=(2\sigma-2\text{ or }2\sigma-5)m\rangle\rangle$$
$$=a|\sigma\,\tau=\sigma\,\nu=0\,lm\rangle+b\,|\sigma\,\tau=\sigma-1\,\nu=0\,lm\rangle, \tag{3.2}$$

with

$$\Gamma_2|lm\rangle\rangle=\Gamma_2(l)|lm\rangle\rangle. \tag{3.3}$$

From (3.2), (3.3), and (2.7) it is easy to derive the secular equation

$$\begin{vmatrix}-\Gamma_2(l)A_l^\sigma(0,0) & Q(\sigma0;\sigma-1\,0)\\ Q(\sigma-1\,0;\sigma0) & -\Gamma_2(l)A_l^{\sigma-1}(0,0)\end{vmatrix}=0,$$

or (3.4)

$$\Gamma_2(l)^2A_l^\sigma(0,0)A_l^{\sigma-1}(0,0)$$
$$=Q(\sigma0;\sigma-1\,0)Q(\sigma-1\,0;\sigma0),$$

with $Q(\tau'\nu';\tau\nu)$ a shorthand notation for $\langle\sigma\tau'\nu'l|\Gamma_2|\sigma\tau\nu l\rangle$.

Taking into account (2.1), and after a lengthy evaluation of the occurring reduced matrix elements of p, where special properties of the hypergeometric functions present (see Ref. 6) are used, one arrives at the following eigenvalues,

$$\Gamma_2(l) = \pm\,[\tfrac{2}{15}(2\sigma-1)^2(2l+3)(l-\sigma+1)]^{1/2},$$
$$l=2\sigma-2,2\sigma-5. \tag{3.5}$$

C. Γ_2 eigenvalues for states of class (c)

For these triple-degenerated states, one can derive the secular equation in a way analogous to that of the class (b) states. Because of the fact that the p-reduced matrix elements only differ from zero when $\tau'=\tau\pm 1$, a few zero elements occur in the determinant representation of the secular equation, which reads

$$\begin{vmatrix} -\Gamma_2(l)A_l^\sigma(0,0) & Q(\sigma 0;\sigma-1\,0) & 0 \\ Q(\sigma-1\,0;\sigma 0) & -\Gamma_2(l)A_l^{\sigma-1}(0,0) & Q(\sigma-1\,0;\sigma-2\,0) \\ 0 & Q(\sigma-2\,0;\sigma-1\,0) & -\Gamma_2(l)A_l^{\sigma-2}(0,0)\end{vmatrix}=0.$$

In this way one immediately observes that one eigenvalue takes on a zero value and that the two others are the solutions of the following quadratic equation:

$$\Gamma_2(l)^2A_l^\sigma(0,0)A_l^{\sigma-1}(0,0)A_l^{\sigma-2}(0,0)$$
$$=A_l^{\sigma-2}(0,0)Q(\sigma-1\,0;\sigma 0)Q(\sigma 0;\sigma-1\,0)+A_l^\sigma(0,0)Q(\sigma-2\,0;\sigma-1\,0)Q(\sigma-1\,0;\sigma-2\,0). \tag{3.6}$$

After a tedious evaluation of the matrix elements occurring, one obtains the following eigenvalues for states of class (c):

$$\Gamma_2(l)=0,\pm\,[\tfrac{4}{15}(2\sigma-3)(2l^3+16l^2+65l+81-22\sigma l-45\sigma)]^{1/2},\quad l=2\sigma-4,2\sigma-7. \tag{3.7}$$

D. Eigenvalues for states of class (d)

For this class, it is the first time that a basis state with a ν value of 1 occurs. This makes the evaluation of the p-reduced matrix elements much more involved. It is not possible anymore to always base upon known relations between hypergeometric functions as discussed in Ref. 6. Therefore separate formulas will be given for each of the l values present. Nevertheless the determinant giving rise to the secular equation can still be given in a general way, i.e.,

$$\begin{vmatrix} \Gamma_2(l)A_l^\sigma(0,0) & \Gamma_2(l)A_l^\sigma(1,0) & -Q(\sigma 0;\sigma-1\,0) & 0 & 0 \\ \Gamma_2(l)A_l^\sigma(1,0) & \Gamma_2(l)A_l^\sigma(1,1) & -Q(\sigma 1;\sigma-1\,0) & 0 & 0 \\ Q(\sigma-1\,0;\sigma 0) & Q(\sigma-1\,0;\sigma 1) & -\Gamma_2(l)A_l^{\sigma-1}(0,0) & Q(\sigma-1\,0;\sigma-2\,0) & 0 \\ 0 & 0 & Q(\sigma-2\,0;\sigma-1\,0) & -\Gamma_2(l)A_l^{\sigma-2}(0,0) & Q(\sigma-2\,0;\sigma-3,0) \\ 0 & 0 & 0 & Q(\sigma-3\,0;\sigma-2\,0) & -\Gamma_2(l)A_l^{\sigma-3}(0,0)\end{vmatrix}=0. \tag{3.8}$$

Again one eigenvalue takes on a zero value and the four other eigenvalues are the solutions of a biquadratic equation in $\Gamma_2(l)$. After a tedious evaluation of the occurring matrix elements, the following results are obtained:

$$\Gamma_2(l=2\sigma-6)$$
$$=0,\pm\,[\tfrac{2}{15}(80\sigma^4-820\sigma^3+3050\sigma^2-4992\sigma$$
$$+3069\pm D_1^{1/2})]^{1/2}, \tag{3.9}$$

with

$$D_1=4096\sigma^8-83\,968\sigma^7+742\,656\sigma^6-3706\,240\sigma^5$$
$$+11\,434\,180\sigma^4-22\,359\,876\sigma^3+27\,087\,561\sigma^2$$
$$-18\,592\,956\sigma+5536\,836, \tag{3.10}$$

and

$$\Gamma_2(l=2\sigma-9)$$
$$=0,\pm\,[\tfrac{2}{15}(80\sigma^4-1180\sigma^3+5930\sigma^2-12\,606\sigma$$
$$+9873\pm D_2^{1/2})]^{1/2}, \tag{3.11}$$

with

$$D_2=4096\sigma^8-120\,832\sigma^7+1498\,368\sigma^6$$
$$-10\,228\,864\sigma^5+42\,251\,620\sigma^4$$
$$-108\,571\,308\sigma^3+169\,914\,861\sigma^2$$
$$-148\,231\,026\sigma+55\,194\,129. \tag{3.12}$$

In these and the following formulas plus and minus signs must be combined in all possible ways.

E. Γ_2 eigenvalues for states belonging to the physical basis (2.2) but with small σ values

It is evident that for small σ values the above discussion does not hold anymore. Certain basis states defined in classes (a)–(d) do not occur for small σ values due to the restrictions (2.2)–(2.6). Hereafter, we summarize the derived eigenvalues for $\sigma\in\{0,1,\ldots,6\}$. If necessary, we give (between brackets) the number of times the same eigenvalue occurs. The evaluation of these results proceeds in a way analogous to the general case, with the exception that for many l values the representation space is smaller:

$$\sigma=0:\quad \Gamma_2(0)=0,$$

$$\sigma=1:\quad \Gamma_2(2)=0,\quad \Gamma_2(0)=0,$$

$$\sigma=2:\quad \Gamma_2(4)=0,\quad \Gamma_2(2)=\pm\sqrt{\tfrac{42}{5}},\quad \Gamma_2(0)=0,$$

$$\sigma=3:\quad \Gamma_2(6)=0,\quad \Gamma_2(4)=\pm 2\sqrt{\tfrac{55}{3}},\quad \Gamma_2(3)=0,$$

$$\Gamma_2(2)=\pm 4\sqrt{\tfrac{6}{5}},\quad \Gamma_2(0)=0(2),$$

$$\sigma=4:\quad \Gamma_2(8)=0,\quad \Gamma_2(6)=\pm 7\sqrt{6},\quad \Gamma_2(5)=0,$$

$$\Gamma_2(4)=0,\quad \pm 2\sqrt{\tfrac{193}{3}},$$

$$\Gamma_2(3)=0,\quad \Gamma_2(2)=0,\pm 9\sqrt{\tfrac{2}{5}},\quad \Gamma_2(0)=0,$$

$$\sigma=5:\quad \Gamma_2(10)=0,\quad \Gamma_2(8)=\pm 18\sqrt{\tfrac{38}{15}},\quad \Gamma_2(7)=0,$$

$$\Gamma_2(6)=0,\pm 6\sqrt{\tfrac{154}{5}},\quad \Gamma_2(5)=\pm 9\sqrt{\tfrac{26}{15}},$$

$$\Gamma_2(4)=\pm\sqrt{\tfrac{1}{15}(2\cdot 11\cdot 13^2\pm 2\sqrt{11(285\,371)})},$$

$$\Gamma_2(3)=0,\quad \Gamma_2(2)=\pm\sqrt{\tfrac{78}{5}},\pm 4\sqrt{3},$$

$$\Gamma_2(0)=0(2),$$

$$\sigma=6:\quad \Gamma_2(12)=0,\quad \Gamma_2(10)=\pm 11\sqrt{\tfrac{46}{3}},\quad \Gamma_2(9)=0,$$

$$\Gamma_2(8)=0,\pm 126/\sqrt{5},\quad \Gamma_2(7)=\pm 22\sqrt{\tfrac{17}{15}},$$

$$\Gamma_2(6)=0,\pm 3\sqrt{\tfrac{2}{5}(351\pm 16\sqrt{331})},$$

$$\Gamma_2(5)=\pm 6\sqrt{\tfrac{42}{5}},$$

$$\Gamma_2(4)=0,\pm 2\sqrt{\tfrac{1697}{15}\pm\tfrac{4}{5}\sqrt{8761}},\quad \Gamma_2(3)=0(2),$$

$$\Gamma_2(2)=\pm 2\sqrt{\tfrac{42}{5}},\pm\sqrt{66},\quad \Gamma_2(0)=0(3).$$

IV. CALCULATION OF B (E2) VALUES IN THE SO(6) LIMIT

In the original paper of Arima and Iachello[7] treating the SO(6) limit, the derivation of analytic expressions for B (E2) values is rather lengthy and quite involved. This follows from the fact that for constructing compact expressions for the matrix elements of the quadrupole operator, and all operators other than the Hamiltonian, the wave functions belonging to the group chain SU(6)⊃SO(6)⊃SO(5)⊃SO(3) have to be expanded in terms of the wave functions belonging to the group chain SU(6)⊃SU(5)⊃SO(5)⊃SO(3). The expansion coefficients occurring can only be obtained by making use of rather complex integral representations of certain sums.

Since the most general "SO(6) invariant" quadrupole operator used by Arima and Iachello[7] can be denoted as

$$T_\mu^{E2}=\tilde{q}_2 p_\mu, \tag{4.1}$$

and since a general expression for the p-reduced matrix element is at our disposal (2.9), the B (E2) values can be derived much more easily. Since we are working in a nonorthonormalized basis the classical definition (see, for example, Ref. 8) for the B (E2) values cannot be used immediately. Instead, the following definition is appropriate:

$$B(\mathrm{E2};\sigma\tau'v'l'\to\sigma\tau vl)=\tilde{q}_2^2\,\frac{1}{(2l'+1)}(-1)^{l'-l}\frac{1}{A_{l'}^{\tau'}(v',v')}\,\frac{1}{A_l^{\tau}(v,v)}\langle\sigma\tau'v'l'\|p\|\sigma\tau vl\rangle\langle\sigma\tau vl\|p\|\sigma\tau'v'l'\rangle. \tag{4.2}$$

In this way the results of Arima and Iachello[7] [formulas (5.14), (5.18), and (5.25)–(5.28)] are reproduced in a more elegant way. As an example of a more complex result, let us give a B (E2) value not derived previously,

$$B(\mathrm{E2};\sigma=N\;\tau\;v'=1\;l'=2\tau-6\to\sigma=N\;\tau-1\;v=0\;l=2\tau-5)$$
$$=\tilde{q}_2^2\,\frac{32(2\tau-3)(4\tau-5)(4\tau-7)}{(2\tau+3)(2\tau-5)(4\tau-11)}(N-\tau+1)(N+\tau+3)$$
$$\times\frac{(\tau^2+27\tau-88)^2}{(18\tau^6-1169\tau^5+29118\tau^4-117347\tau^3+83592\tau^2+204628\tau-230160)}.$$

V. CONCLUSIONS

We have succeeded in deriving closed expressions for the eigenvalues of one of the third-order SO(3) scalar operators occurring in the SO(6) limit of the IBM. The method discussed has been applied to four typical classes of SO(6) physical basis states. It is, however, straightforward to extend the method in a numerical way to other cases. The derived formulas can now be used for a systematic study of the influence of higher-order interaction terms of the type Γ_2 on the theoretical spectra and transition rates of so-called SO(6) nuclei. We hope to report on this in the near future.

[1] G. Vanden Berghe, H. De Meyer, and P. Van Isacker, Phys. Rev. C **32**, 1049 (1985).

[2] G. Vanden Berghe and H. De Meyer, *Proceedings of the 14th ICGTMP*, edited by Y. M. Cho (World Scientific, Singapore, 1986), p. 346.

[3] J. Van der Jeugt and H. De Meyer, "Generating functions for higher-order interaction terms in the IBA-Hamiltonian," J. Phys. A **20**, 5045 (1987).

[4] J. Vanthournout, J. Van der Jeugt, H. De Meyer, and G. Vanden Berghe, J. Math. Phys. **28**, 2529 (1987).

[5] S. A. Williams and D. L. Pursey, J. Math. Phys. **9**, 1230 (1968).

[6] G. Vanden Berghe and H. De Meyer, J. Math. Phys. **22**, 2326 (1981).

[7] A. Arima and F. Iachello, Ann. Phys. (NY) **123**, 468 (1979).

[8] A. De-Shalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963).

New classes of symmetries for partial differential equations

George W. Bluman and Gregory J. Reid
Department of Mathematics, University of British Columbia, Vancouver, British Columbia V6T 1Y4, Canada

Sukeyuki Kumei
Faculty of Textile Science, Shinshu University, Ueda, Nagano-ken 386, Japan

(Received 6 August 1987; accepted for publication 18 November 1987)

New classes of symmetries for partial differential equations are introduced. By writing a given partial differential equation S in a conserved form, a related system T with potentials as additional dependent variables is obtained. The Lie group of point transformations admitted by T induces a symmetry group of S. *New symmetries* may be obtained for S that are neither point nor Lie–Bäcklund symmetries. They are determined by a completely algorithmic procedure. Significant new symmetries are found for the wave equation with a variable wave speed and the nonlinear diffusion equation.

I. INTRODUCTION

In this paper we introduce new classes of symmetries for partial differential equations (PDE's). We present an algorithm to find such symmetries. In general, they are not determined by a direct application, to the given PDE, of Lie's method for finding point symmetries and Lie–Bäcklund symmetries. These new symmetries significantly extend the applicability of group analysis to differential equations.

A symmetry group of a differential equation is a group that maps solutions to other solutions of the differential equation.

Lie considered groups of point transformations depending on continuous parameters, acting on the space of independent and dependent variables of a given differential equation. Unlike the case for a discrete group, Lie showed that the continuous group of point transformations admitted by a differential equation can be found by an explicit algorithm (cf. Refs. 1–3 for recent accounts). Such a group is completely characterized in terms of its infinitesimal generators, which depend on the independent and dependent variables of the given differential equation. Lie extended his work to groups of contact transformations that act on the space of independent and dependent variables and first derivatives of the dependent variables of the given differential equation.

Noether[4] recognized the possibility of generalizing Lie's infinitesimals by allowing them to depend on derivatives of the dependent variables up to any finite order. Such generalized symmetries, commonly called Lie–Bäcklund transformations, came to fruition in Ref. 5. Lie–Bäcklund symmetries lead directly to the infinity of conservation laws arising in the study of the Korteweg–de Vries, sine–Gordon, nonlinear Schrödinger, and other nonlinear differential equations exhibiting soliton behavior and are computed by a simple extension of Lie's algorithm.[1,6,7]

In our approach we obtain new classes of symmetries by computing Lie groups of point transformations whose infinitesimals act on a different space than the space of independent variables, dependent variables, and their derivatives, of the given differential equation. In terms of the variables of the given differential equation, our new symmetries are neither point symmetries nor Lie–Bäcklund symmetries.

Our approach can be applied to a system S of PDE's with independent variables x and dependent variables u, written in a conserved form with respect to some choice of these variables. Through the conserved form we naturally introduce potentials ϕ. The resulting system T of PDE's has as its variables the independent variables x, the dependent variables u of S, plus new dependent variables ϕ. We find the Lie group G_T of point transformations, of this enlarged space of variables (x,u,ϕ), admitted by system T.

Any transformation in G_T maps solutions of T into other solutions of T and hence maps solutions of S into other solutions of S. Consequently, G_T is a symmetry group of S. A transformation in G_T is a *new symmetry for S* if the infinitesimal of the transformation, corresponding to any of the variables (x,u), depends explicitly on ϕ. We show that a *new symmetry* is neither a point symmetry nor a Lie–Bäcklund symmetry of S.

Our new symmetries are nonlocal symmetries that are realized as local (point) symmetries in the space (x,u,ϕ). Thus they can be found by Lie's algorithm.

Special types of nonlocal symmetries have been studied by other authors.[8–10] Their works give no explicit algorithms for finding nonlocal symmetries. In general, our nonlocal symmetries do not belong to the types considered by these authors.

In Sec. II we present our method for obtaining new symmetries admitted by PDE's. By way of example, we find new symmetries for the wave equation in Sec. III and the nonlinear diffusion equation in Sec. IV.

II. METHOD FOR FINDING NEW SYMMETRIES

Consider a PDE S of order m written in a conserved form,

$$\sum_{i=1}^{n} \frac{\partial}{\partial x_i} F_i(x,u,\partial u,\partial^2 u,\ldots,\partial^{m-1}u) = 0, \qquad (2.1)$$

with $n \geqslant 2$ independent variables $x=(x_1,x_2,\ldots,x_n)$ and a single dependent variable u; $\partial^j u$ represents all jth-order partials of u with respect to x. (For simplicity we consider a single PDE—the generalization to a system of PDE's in a conserved form is straightforward.)

 0022-2488/88/040806-06\$02.50

We remark that if a given PDE is not written in a conserved form, there are a number of ways of attempting to put it in a conserved form. As discussed in Sec. V, these include a change of variables (*dependent as well as independent*), an application of Noether's theorem, and combinations of the above.

Since Eq. (2.1) is in a conserved form, there is an $(n-1)$-exterior differential form F such that Eq. (2.1) can be written as $dF=0$. It follows that there is an $(n-2)$-form Φ[11],

$$F=d\Phi. \tag{2.2}$$

In terms of components, Eq. (2.2) implies that there exist $\frac{1}{2}n(n-1)$ "potentials" Ψ_{ij}, components of an antisymmetric tensor, such that

$$\begin{aligned}F_i(x,u,\partial u,...,\partial^{m-1}u)&=\sum_{i<j}(-1)^j\frac{\partial\Psi_{ij}}{\partial x_j}+\sum_{j<i}(-1)^{i-1}\frac{\partial\Psi_{ji}}{\partial x_j},\\ i,j&=1,2,...,n.\end{aligned} \tag{2.3}$$

Equation (2.3) is a system of PDE's with $1+\frac{1}{2}n(n-1)$ dependent variables u, Ψ_{ij} $(i<j)$. Thus (2.3) is underdetermined for $n\geqslant 3$. We can impose suitable constraints (effectively, a choice of gauge) on the potentials Ψ_{ij} to make system (2.3) into a determined system. A natural way to do this is to impose the conditions

$$\Psi_{ij}=0,\quad |i-j|\neq 1. \tag{2.4}$$

In this case, letting

$$\phi_i=\Psi_{i,i+1},\quad i=1,2,...,n-1, \tag{2.5}$$

system (2.3) becomes the determined system T,

$$\begin{aligned}F_1&=\frac{\partial\phi_1}{\partial x_2},\\ F_\ell&=(-1)^{\ell-1}\left[\frac{\partial\phi_\ell}{\partial x_{\ell+1}}+\frac{\partial\phi_{\ell-1}}{\partial x_{\ell-1}}\right],\quad 1<\ell<n,\\ F_n&=(-1)^{n+1}\frac{\partial\phi_{n-1}}{\partial x_{n-1}}.\end{aligned} \tag{2.6}$$

If $n=2$, let $x_1=x$, $x_2=t$, $F_1=F$, and $F_2=-G$, so that S becomes

$$\frac{\partial F}{\partial x}-\frac{\partial G}{\partial t}=0. \tag{2.7}$$

Let the potential $\Psi_{12}=\phi_1=\phi$. Consequently, T is

$$\frac{\partial\phi}{\partial t}=F(x,t,u,\partial u,...,\partial^{m-1}u), \tag{2.8a}$$

$$\frac{\partial\phi}{\partial x}=G(x,t,u,\partial u,...,\partial^{m-1}u). \tag{2.8b}$$

If $n=4$, let $x_1=x$, $x_2=y$, $x_3=z$, $x_4=t$, $F_1=F$, $F_2=G$, $F_3=H$, and $F_4=I$, so that S becomes

$$\frac{\partial F}{\partial x}+\frac{\partial G}{\partial y}+\frac{\partial H}{\partial z}+\frac{\partial I}{\partial t}=0. \tag{2.9}$$

The corresponding determined system T is

$$\frac{\partial\phi_1}{\partial y}=F(x,y,z,t,u,\partial u,...,\partial^{m-1}u), \tag{2.10a}$$

$$-\left[\frac{\partial\phi_1}{\partial x}+\frac{\partial\phi_2}{\partial z}\right]=G(x,y,z,t,u,\partial u,...,\partial^{m-1}u), \tag{2.10b}$$

$$\frac{\partial\phi_2}{\partial y}+\frac{\partial\phi_3}{\partial t}=H(x,y,z,t,u,\partial u,...,\partial^{m-1}u), \tag{2.10c}$$

$$-\frac{\partial\phi_3}{\partial z}=I(x,y,z,t,u,\partial u,...,\partial^{m-1}u). \tag{2.10d}$$

Now assume that a system T admits a one-parameter (ϵ) Lie group of point transformations

$$x^*=f(x,u,\phi;\epsilon)=x+\epsilon\xi_T(x,u,\phi)+O(\epsilon^2), \tag{2.11a}$$

$$u^*=g(x,u,\phi;\epsilon)=u+\epsilon\eta_T(x,u,\phi)+O(\epsilon^2), \tag{2.11b}$$

$$\phi^*=h(x,u,\phi;\epsilon)=\phi+\epsilon\zeta_T(x,u,\phi)+O(\epsilon^2), \tag{2.11c}$$

where ξ_T, η_T, and ζ_T are the infinitesimals of x, u, and ϕ, respectively, of the group. This group maps a solution of T into another solution of T and hence induces a mapping of a solution of S into another solution of S. Thus the group (2.11) is a symmetry group of PDE S. *This one-parameter symmetry group of* PDE *S is a new symmetry group of S if and only if either ξ_T or η_T depends explicitly on ϕ.* A new symmetry of S is neither a point symmetry nor a Lie–Bäcklund symmetry of S since ϕ, as defined by system (2.6), appears only in derivative form. Hence this new symmetry cannot be expressed as a function of $(x,u,\partial u,...,\partial^k u)$, for any finite k. Clearly, from its form, a new symmetry of S is a nonlocal symmetry of S. We let G_T denote the group of all point transformations admitted by T.

A one-parameter Lie group of point transformations admitted by S, in terms of its given variables, is of the form

$$x^*=x+\epsilon\xi_S(x,u)+O(\epsilon^2), \tag{2.12a}$$

$$u^*=u+\epsilon\eta_S(x,u)+O(\epsilon^2). \tag{2.12b}$$

Let G_S denote the group of point transformations of the form (2.12) admitted by S. It is important to note that the transformations belonging to G_S with infinitesimals $\xi_S(x,u)$ and $\eta_S(x,u)$ may not belong to G_T in the following sense: there exist no transformations in G_T with infinitesimals $\xi_T(x,u,\phi)$, $\eta_T(x,u,\phi)$, and $\zeta_T(x,u,\phi)$ such that

$$\xi_T(x,u,\phi)\equiv\xi_S(x,u), \tag{2.13a}$$

$$\eta_T(x,u,\phi)\equiv\eta_S(x,u). \tag{2.13b}$$

Say S is a linear PDE and T is a linear system of PDE's. In this case, ξ_S and ξ_T depend only on x. Here we distinguish two types of new symmetries arising from a new symmetry in G_T with an infinitesimal $\xi_T(x)$.

(i) *A linear partial differential equation S is said to have a new symmetry of type I if it has a new symmetry for which there is no infinitesimal in G_S such that $\xi_S(x)\equiv\xi_T(x)$.*

(ii) *A linear partial differential equation S is said to have a new symmetry of type II if it has a new symmetry for which there is some infinitesimal in G_S such that $\xi_S(x)\equiv\xi_T(x)$.*

For a new symmetry of type II, the similarity variables (group invariants depending only on x) are identical to those for some symmetry in G_S. This is not the case for a new symmetry of type I.

There are many ways of expressing a given PDE S as a system. However, the symmetries of such a system may not induce nonlocal symmetries for S. For example, the "usual" way to find a system $\widehat{T}$ related to S is to introduce new dependent variables $v_i = \partial u/\partial x_i$, $1 \leqslant i \leqslant n$. A point symmetry admitted by $\widehat{T}$, namely,

$$x^* = x + \epsilon\hat{\xi}(x,u,v) + O(\epsilon^2), \tag{2.14a}$$

$$u^* = u + \epsilon\hat{\eta}(x,u,v) + O(\epsilon^2), \tag{2.14b}$$

$$v^* = v + \epsilon\hat{\zeta}(x,u,v) + O(\epsilon^2), \tag{2.14c}$$

always induces a local symmetry of S that is either a point symmetry or a Lie–Bäcklund symmetry of S.

III. EXAMPLES OF NEW SYMMETRIES FOR THE WAVE EQUATION

Consider the wave equation S:

$$c^2(x)\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial t^2} = 0. \tag{3.1}$$

Equation (3.1) can be expressed in a conserved form,

$$\frac{\partial F}{\partial x} - \frac{\partial G}{\partial t} = 0, \tag{3.2}$$

where

$$F = \frac{\partial u}{\partial x}, \tag{3.3a}$$

$$G = \frac{1}{c^2(x)}\frac{\partial u}{\partial t}. \tag{3.3b}$$

The associated system T is

$$\frac{\partial \phi}{\partial t} = \frac{\partial u}{\partial x}, \tag{3.4a}$$

$$\frac{\partial \phi}{\partial x} = \frac{1}{c^2(x)}\frac{\partial u}{\partial t}. \tag{3.4b}$$

Let G_S and G_T be the Lie groups of point transformations admitted by S [Eq. (3.1)] and T [Eqs. (3.4)], respectively. These groups depend on the form of the wave speed $c(x)$ and were derived in Ref. 12. The results in that paper can be broadly summarized in terms of Theorems 1–5 following. [A prime denotes differentiation with respect to x; we exclude the case $c(x) = (\alpha x + \beta)^2$, with $\{\alpha,\beta\}$ arbitrary constants, for which G_S is an ∞-parameter group.]

Theorem 1: The wave equation (3.1) admits a four-parameter Lie group of point transformations G_S if and only if the wave speed $c(x)$ satisfies the fifth-order ODE

$$\left\{c^2\left[\frac{H'''}{2H'+H^2} + 3\,\frac{[2(H')^3 - 2HH'H'' - (H'')^2]}{[2H'+H^2]^2}\right]\right\}' = 0, \tag{3.5}$$

where

$$H = c'/c. \tag{3.6}$$

Theorem 2: G_T is a four-parameter Lie group of point transformations if and only if the wave speed $c(x)$ satisfies the fourth-order ODE

$$[cc'(c/c')'']' = 0. \tag{3.7}$$

Theorem 3: For any wave speed $c(x)$ satisfying ODE (3.7), there exists a new symmetry of the wave equation (3.1).

Theorem 4: The new symmetries of the wave equation (3.1) arising from G_T are new symmetries of type II if and only if the wave speed $c(x)$ satisfies the third-order ODE

$$(c/c')'' = 0. \tag{3.8}$$

The general solution of (3.8) is

$$c(x) = (\alpha x + \beta)^\gamma, \tag{3.9}$$

where $\{\alpha,\beta,\gamma\}$ are arbitrary constants.

Theorem 5: The new symmetries of the wave equation (3.1) arising from G_T are new symmetries of type I if and only if the wave speed $c(x)$ satisfies the ODE

$$cc'(c/c')'' = \text{const} \neq 0. \tag{3.10}$$

The following theorem was proved in Ref. 13.

Theorem 6: A wave speed $c(x)$ simultaneously satisfies (3.10) and (3.5) if and only if either

$$\sqrt{c} - \arctan \gamma\sqrt{c} = \alpha x + \beta, \tag{3.11a}$$

or

$$2\sqrt{c} + \log|(\sqrt{c} - \gamma)/(\sqrt{c} + \gamma)| = \alpha x + \beta, \tag{3.11b}$$

where $\{\alpha,\beta,\gamma\}$ are arbitrary constants.

From the above follows this corollary.

Corollary 1: Both of the groups G_T and G_S are four-parameter groups if and only if the wave speed $c(x)$ satisfies (3.8), (3.11a), or (3.11b). The family of wave speeds (3.8) yields new symmetries of type II and no new symmetries of type I. The families of wave speeds (3.11a) and (3.11b) yield new symmetries of type I and no new symmetries of type II.

The following representative examples illustrate the above theorems.

1. $c(x) = \sqrt{1 + e^x}$. In this case, G_S is a two-parameter group and G_T is a four-parameter group. Infinitesimal generators of their Lie algebras are

$$G_S:\ L_1 = u\frac{\partial}{\partial u},\quad L_2 = \frac{\partial}{\partial t};$$

$$G_T:\ \widetilde{L}_1 = u\frac{\partial}{\partial u} + \phi\frac{\partial}{\partial \phi},\quad \widetilde{L}_2 = \frac{\partial}{\partial t},$$

$$\widetilde{L}_3 = e^t\left\{2(1+e^{-x})\frac{\partial}{\partial x} - (2e^{-x}+1)\frac{\partial}{\partial t} + (u-\phi)\left[(1+e^{-x})\frac{\partial}{\partial u} - e^{-x}\frac{\partial}{\partial \phi}\right]\right\},$$

$$\widetilde{L}_4 = e^{-t}\left\{2(1+e^{-x})\frac{\partial}{\partial x} + (2e^{-x}+1)\frac{\partial}{\partial t} + (u+\phi)\left[(1+e^{-x})\frac{\partial}{\partial u} - e^{-x}\frac{\partial}{\partial \phi}\right]\right\}.$$

The generators $\widetilde{L}_3$ and $\widetilde{L}_4$ are new symmetries of type I for the corresponding wave equation (3.1).

2. $c(x) = 1 - x^2$. In this case, G_T is a two-parameter group and G_S has four parameters. Infinitesimal generators of their Lie algebras are

$$G_S:\ L_1 = u\frac{\partial}{\partial u},\quad L_2 = \frac{\partial}{\partial t},$$

$$L_3 = (1-x^2)\frac{\partial}{\partial x} - xu\frac{\partial}{\partial u},$$

$$L_4 = t(1-x^2)\frac{\partial}{\partial x} + \frac{1}{2}\log\frac{|x+1|}{|x-1|}\frac{\partial}{\partial t} - xtu\frac{\partial}{\partial u};$$

$$G_T\colon\ \tilde{L}_1 = u\frac{\partial}{\partial u} + \phi\frac{\partial}{\partial \phi},\quad \tilde{L}_2 = \frac{\partial}{\partial t}.$$

3. $c(x) = x$. Here both of the groups G_S and G_T have four parameters, and there is a new symmetry of type II for S. Infinitesimal generators of their Lie algebras are

$$G_S\colon\ L_1 = u\frac{\partial}{\partial u},\quad L_2 = \frac{\partial}{\partial t},\quad L_3 = x\frac{\partial}{\partial x} + u\frac{\partial}{\partial u},$$

$$L_4 = 2xt\frac{\partial}{\partial x} + 2\log|x|\frac{\partial}{\partial t} + tu\frac{\partial}{\partial u};$$

$$G_T\colon\ \tilde{L}_1 = u\frac{\partial}{\partial u} + \phi\frac{\partial}{\partial \phi},\quad \tilde{L}_2 = \frac{\partial}{\partial t},$$

$$\tilde{L}_3 = x\frac{\partial}{\partial x} + u\frac{\partial}{\partial u},$$

$$\tilde{L}_4 = 2xt\frac{\partial}{\partial x} + 2\log|x|\frac{\partial}{\partial t} + (tu - x\phi)\frac{\partial}{\partial u} - (x^{-1}u + t\phi)\frac{\partial}{\partial \phi}.$$

The infinitesimal generator $\tilde{L}_4$ of G_T is a new symmetry of type II for S.

4. $2\sqrt{c} + \log|(\sqrt{c}-1)/(\sqrt{c}+1)| = x$. In this case, both of the groups G_S and G_T have four parameters, and there are new symmetries of type I for S. Infinitesimal generators of their Lie algebras are

$$G_S\colon\ L_1 = u\frac{\partial}{\partial u},\quad L_2 = \frac{\partial}{\partial t},$$

$$L_3 = e^{t/2}(c-1)^{-1/2}\times\left[c^{3/2}\frac{\partial}{\partial x} - \frac{\partial}{\partial t} + \frac{(c-1)}{2}u\frac{\partial}{\partial u}\right],$$

$$L_4 = e^{-t/2}(c-1)^{-1/2}\times\left[c^{3/2}\frac{\partial}{\partial x} + \frac{\partial}{\partial t} + \frac{(c-1)}{2}u\frac{\partial}{\partial u}\right];$$

$$G_T\colon\ \tilde{L}_1 = u\frac{\partial}{\partial u} + \phi\frac{\partial}{\partial \phi},\quad \tilde{L}_2 = \frac{\partial}{\partial t},$$

$$\tilde{L}_3 = \frac{e^t}{c-1}\left\{4c^{3/2}\frac{\partial}{\partial x} - 2(c+1)\frac{\partial}{\partial t} + [(3c-1)u - 2c^{3/2}\phi]\frac{\partial}{\partial u} + [(3-c)\phi - 2c^{-1/2}u]\frac{\partial}{\partial \phi}\right\},$$

$$\tilde{L}_4 = \frac{e^{-t}}{c-1}\left\{4c^{3/2}\frac{\partial}{\partial x} + 2(c+1)\frac{\partial}{\partial x} + [(3c-1)u + 2c^{3/2}\phi]\frac{\partial}{\partial u} + [(3-c)\phi + 2c^{-1/2}u]\frac{\partial}{\partial \phi}\right\}.$$

Any linear combination of $\tilde{L}_3$ and $\tilde{L}_4$ is a new symmetry of type I for S.

As these examples clearly demonstrate, our method enables one to discover systematically new symmetries of (3.1) that cannot be found by a direct application of Lie's algorithm to (3.1).

Ovsiannikov[14] recognized the difference between the groups admitted by an equation equivalent to (3.1) and by a corresponding system equivalent to (3.4). He made some cursory remarks about these differences and went no further.

IV. EXAMPLES OF NEW SYMMETRIES FOR THE NONLINEAR DIFFUSION EQUATION

Consider the nonlinear diffusion equation S,

$$\frac{\partial}{\partial x}\left[K(u)\frac{\partial u}{\partial x}\right] - \frac{\partial u}{\partial t} = 0. \tag{4.1}$$

As it is written, Eq. (4.1) is already in a conserved form,

$$\frac{\partial F}{\partial x} - \frac{\partial G}{\partial t} = 0, \tag{4.2}$$

where

$$F = K(u)\frac{\partial u}{\partial x}, \tag{4.3a}$$

$$G = u. \tag{4.3b}$$

The associated system T is

$$\frac{\partial \phi}{\partial t} = K(u)\frac{\partial u}{\partial x}, \tag{4.4a}$$

$$\frac{\partial \phi}{\partial x} = u. \tag{4.4b}$$

The group G_S of (4.1) depends on the form of the conductivity $K(u)$ and is derived in Refs. 2, 3, and 15. The results are summarized as follows.

1. $K(u)$ arbitrary. Equation (4.1) always admits a three-parameter group with infinitesimal generators

$$L_1 = \frac{\partial}{\partial t},\quad L_2 = \frac{\partial}{\partial x},\quad L_3 = x\frac{\partial}{\partial x} + 2t\frac{\partial}{\partial t}. \tag{4.5}$$

2. $K(u) = \lambda(u+\kappa)^\nu$, $\{\nu(\neq -\frac{4}{3}),\lambda,\kappa\}$ arbitrary constants. Here G_S is a four-parameter group with infinitesimal generators L_1, L_2, and L_3 given by (4.5), and

$$L_4 = x\frac{\partial}{\partial x} + \frac{2}{\nu}(u+\kappa)\frac{\partial}{\partial t}. \tag{4.6}$$

A limiting case is $K(u) = \lambda e^{\nu u}$.

3. $K(u) = \lambda(u+\kappa)^{-4/3}$, $\{\lambda,\kappa\}$ arbitrary constants. Here G_S is a five-parameter group with infinitesimal generators L_1, L_2, and L_3 given by (4.5), L_4 given by (4.6) with $\nu = -\frac{4}{3}$, and

$$L_5 = x^2\frac{\partial}{\partial x} - 3x(u+\kappa)\frac{\partial}{\partial u}. \tag{4.7}$$

The group G_T of system (4.4) also depends on the form of the conductivity $K(u)$. This group is presented here for the first time. The results are summarized as follows.

1. $K(u)$ arbitrary. Equations (4.4) always admit a four-parameter group with infinitesimal generators

$$\tilde{L}_0=\frac{\partial}{\partial\phi},\quad \tilde{L}_1=\frac{\partial}{\partial t},\quad \tilde{L}_2=\frac{\partial}{\partial x},$$
$$\tilde{L}_3=x\frac{\partial}{\partial x}+2t\frac{\partial}{\partial t}+\phi\frac{\partial}{\partial\phi}. \tag{4.8}$$

2. $K(u)=\lambda(u+\kappa)^\nu$, $\{\nu(\neq -2),\lambda,\kappa\}$ arbitrary constants. Here G_T is a five-parameter group with infinitesimal generators $\tilde{L}_0$, $\tilde{L}_1$, $\tilde{L}_2$, and $\tilde{L}_3$ given by (4.8), and

$$\tilde{L}_4=x\frac{\partial}{\partial x}+\frac{2}{\nu}(u+\kappa)\frac{\partial}{\partial u}+\left(1+\frac{2}{\nu}\right)\phi\frac{\partial}{\partial\phi}. \tag{4.9}$$

3. $K(u)=\lambda(u+\kappa)^{-2}$, $\{\lambda,\kappa\}$ arbitrary constants. Here G_T is an ∞-parameter group with infinitesimal generators $\tilde{L}_0$, $\tilde{L}_1$, $\tilde{L}_2$, and $\tilde{L}_3$ given by (4.8), $\tilde{L}_4$ given by (4.9), and

$$\tilde{L}_2=-x\phi\frac{\partial}{\partial x}+(u+\kappa)[\phi+x(u+\kappa)]\frac{\partial}{\partial u}+2t\frac{\partial}{\partial\phi}, \tag{4.10a}$$

$$\tilde{L}_5=-x(\phi^2+2t)\frac{\partial}{\partial x}+4t^2\frac{\partial}{\partial t}+(u+\kappa)$$
$$\times[\phi^2+6t+2x\phi(u+\kappa)]\frac{\partial}{\partial u}+4t\phi\frac{\partial}{\partial\phi}, \tag{4.10b}$$

$$\tilde{L}_\infty=\Theta(\phi,t)\frac{\partial}{\partial x}-u^2\frac{\partial\Theta(\phi,t)}{\partial\phi}\frac{\partial}{\partial u}, \tag{4.10c}$$

where $v=\Theta(\phi,t)$ is an arbitrary solution of the linear differential equation

$$\frac{\partial^2 v}{\partial\phi^2}-\frac{\partial v}{\partial t}=0. \tag{4.11}$$

4. $$K(u)=\frac{1}{u^2+pu+q}\exp\left[r\int\frac{du}{u^2+pu+q}\right].$$

In this example, $\{p,q,r\}$ are arbitrary constants *not* satisfying either of the relationships

(a) $r=\pm 2$, $p^2-4q>0$,

(b) $r=0$, $p^2-4q=0$.

The cases (a) and (b) belong to 3.

Here G_T is a five-parameter group with infinitesimal generators $\tilde{L}_0$, $\tilde{L}_1$, $\tilde{L}_2$, and $\tilde{L}_3$ given by (4.8), and

$$\tilde{L}_4=\phi\frac{\partial}{\partial x}+(r-p)t\frac{\partial}{\partial t}-(u^2+pu+q)\frac{\partial}{\partial u}$$
$$-(qx+p\phi)\frac{\partial}{\partial\phi}. \tag{4.12}$$

A comparison of the groups G_S and G_T leads to the following theorem.

Theorem 7: The nonlinear diffusion equation (4.1) has a new symmetry, arising from $G_{T'}$ if and only if

$$K(u)=\frac{1}{u^2+pu+q}\exp\left[r\int\frac{du}{u^2+pu+q}\right],$$

with arbitrary constants $\{p,q,r\}$.

Ovsiannikov[1,16] expressed the nonlinear diffusion equation (4.1) as a system,

$$v=K(u)\frac{\partial u}{\partial x}, \tag{4.13a}$$

$$\frac{\partial v}{\partial x}=\frac{\partial u}{\partial t}. \tag{4.13b}$$

Our remarks at the end of Sec. II show that a point symmetry of system (4.13) is always a local symmetry of the single equation (4.1). In particular, as Ovsiannikov found in his complete point group classification of system (4.13), the group is G_S when restricted to (x,t,u) space.

V. CONCLUDING REMARKS

(1) There are various ways of attempting to put a PDE S into a conserved form. One way is to find a change of variables $\bar{x}=X(x,u)$, $\bar{u}=U(x,u)$, if possible, so that S becomes a conserved form,

$$\sum_{i=1}^{n}\frac{\partial\bar{F}_i}{\partial\bar{x}_i}(\bar{x},\bar{u},\partial\bar{u},\ldots,\partial^{m-1}\bar{u})=0, \tag{5.1}$$

where $\partial^k\bar{u}$ denotes all k th-order partials of $\bar{u}$ with respect to $\bar{x}$.

Another way depends on S being represented as the Euler–Lagrange equation for some Lagrangian density L. Each one-parameter Lie group of point transformations that leaves the action integral invariant leads to a conserved form for S through an application of Noether's theorem.

The following two examples illustrate other ways of obtaining conserved forms.

Consider the Schrödinger equation S:

$$-\frac{\partial^2 u}{\partial x^2}+V(x)u=i\frac{\partial u}{\partial t}. \tag{5.2}$$

We can reexpress (5.2) in the form

$$\frac{\partial F}{\partial x}-\frac{\partial G}{\partial t}=0, \tag{5.3}$$

where

$$F=\omega(x)\frac{\partial u}{\partial x}-\omega'(x)u, \tag{5.4a}$$

$$G=-i\omega(x)u, \tag{5.4b}$$

with

$$V(x)=\omega''(x)/\omega(x). \tag{5.5}$$

The corresponding system T is

$$\frac{\partial\phi}{\partial t}=\omega(x)\frac{\partial u}{\partial x}-\omega'(x)u, \tag{5.6a}$$

$$\frac{\partial\phi}{\partial x}=-i\omega(x)u. \tag{5.6b}$$

In a future paper we will show that for a class of potentials $V(x)$, the group G_T of system (5.6) generates new symmetries for the Schrödinger equation (5.2).

For our second example we consider the nonlinear wave equation

$$c^2\left(x,t,\frac{\partial u}{\partial x}\right)\frac{\partial^2 u}{\partial x^2}-\frac{\partial^2 u}{\partial t^2}=0. \tag{5.7}$$

We differentiate (5.7) with respect to x and let $v=\partial u/\partial x$ so that (5.7) becomes the PDE S,

$$\frac{\partial}{\partial x}\left[c^2(x,t,v)\frac{\partial v}{\partial x}\right]-\frac{\partial}{\partial t}\left(\frac{\partial v}{\partial t}\right)=0. \tag{5.8}$$

The corresponding system T is

$$\frac{\partial \phi}{\partial t} = c^2(x,t,v)\,\frac{\partial v}{\partial x}, \tag{5.9a}$$

$$\frac{\partial \phi}{\partial x} = \frac{\partial v}{\partial t}. \tag{5.9b}$$

If (v,ϕ) solves T, then $u(x,t)$, defined by

$$\frac{\partial u}{\partial x} = v, \tag{5.10a}$$

$$\frac{\partial u}{\partial t} = \phi, \tag{5.10b}$$

solves the nonlinear wave equation (5.7). Hence the symmetry group G_T is a symmetry group of (5.7). From the form of (5.10) we see that new symmetries may arise for (5.7).

(2) A new symmetry leads to invariant solutions of T, which, in turn, lead to solutions of S. If S and T are linear and the new symmetry is of type I, then these solutions cannot be obtained by applying the infinitesimal operators of G_S to the invariant solutions of S arising from G_S.

If a new symmetry arising from G_T has ξ_T depending only on x, then it can be used to solve boundary value (initial value) problems explicitly. New symmetries have been used to solve initial value problems for wave equations (3.1) for a class of wave speeds with a smooth transition.[17]

(3) Since the choice of conserved form is not necessarily unique, various new groups could be admitted by a given differential equation. For any conserved form the symmetries of the related system are computed by the standard Lie algorithm. The work presented in this paper, when combined with recent advances using symbolic manipulation to execute Lie's algorithm,[18] offers considerable promise for applying group methods to much wider classes of differential equations.

[1] P. J. Olver, *Applications of Lie Groups to Differential Equations* (Springer, New York, 1986).
[2] L. V. Ovsiannikov, *Group Analysis of Differential Equations* (Academic, New York, 1982).
[3] G. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer, New York, 1974).
[4] E. Noether, Nachr. Akad. Wiss. Goettingen Math. Phys. Kl. 2 **1918**, 234.
[5] R. L. Anderson, S. Kumei, and C. E. Wulfman, Phys. Rev. Lett. **28**, 988 (1972).
[6] R. L. Anderson and N. H. Ibragimov, *Lie–Bäcklund Transformations in Applications* (SIAM, Philadelphia, 1979).
[7] N. H. Ibragimov, *Transformation Groups Applied to Mathematical Physics* (Reidel, Dordrecht, 1985).
[8] B. G. Konopelchenko and V. G. Mokhnachev, Sov. J. Nucl. Phys. **30**, 288 (1979); J. Phys. A: Math. Gen. **13**, 3113 (1980).
[9] S. Kumei, Ph.D. thesis, University of British Columbia, 1981.
[10] O. V. Kapcov, Sov. Math. Dokl. **25**, 173 (1982).
[11] W. Slebodzinski, *Exterior Forms and Their Applications* (PWN, Warsaw, 1970).
[12] G. Bluman and S. Kumei, J. Math. Phys. **28**, 307 (1987).
[13] G. Bluman and S. Kumei, IAM Technical Report No. 87-6, University of British Columbia, 1987, to appear in J. Math. Anal. Appl.
[14] L. V. Ovsiannikov, *Group Properties of Differential Equations* (Novosibirsk, 1962), translated by G. Bluman, unpublished.
[15] G. Bluman, Ph.D. thesis, California Institute of Technology, 1967.
[16] L. V. Ovsiannikov, Dokl. Akad. Nauk SSSR **125**, 492 (1959).
[17] G. Bluman and S. Kumei, J. Math. Phys. **29**, 86 (1988).
[18] F. Schwarz, Computing **34**, 91 (1985).

Composite variational principles and the determination of conservation laws

G. Caviglia
Department of Mathematics, Via L. B. Alterti 4, 16132 Genova, Italy

(Received 12 March 1986; accepted for publication 9 December 1987)

Sufficiently regular systems of partial differential equations can always be derived from a composite variational principle via the introduction of a suitable set of adjoint variables. By applying an appropriate formulation of Noether's theorem to such composite variational principles, an operative procedure leading to the determination of the conservation laws of the given system is determined. As an application, conservation laws for a viscous incompressible fluid and an inviscid barotropic compressible fluid are examined in detail. In particular, new conservation laws for Euler flows are found.

I. INTRODUCTION

It is commonly believed that conservation laws could be beneficial in the application and understanding of physical theories. Moreover, conservation laws can be used in proving the existence of solutions and in analyzing scattering and integrability properties of the given system. Thus considerable attention has always been paid to the question of the determination and classification of conservation laws for given systems of partial differential equations.[1-3]

In this connection, a systematic method for finding conserved currents in terms of isovectors of closed ideals of exterior forms[4,5] and the use of the Green's identity has recently been proposed.[6] In this paper we describe an alternative approach leading to the determination of the complete set of conserved vectors for any system of partial differential equations, without any need for additional assumptions on the structure of the given equations, e.g., that they are expressed in Hamiltonian form or may be derived from a variational principle. The present formulation is based on the application of a suitable version of Noether's theorem to an appropriately defined composite variational principle.[7] The resulting systematic procedure reduces the operative determination of conservation laws to the problem of seeking solutions to largely underdetermined linear systems of partial differential equations. Besides these practical advantages, it is to be emphasized that the algorithm based on composite variational principles exhausts the class of allowable conservation laws, yields a proper extension of the approach proposed in Ref. 6, and also transfers to the framework of generally nonlinear field theories some results holding for mechanical systems.[8,9]

Specific applications of this approach to Navier–Stokes equations and compressible Euler equations are then discussed. In particular, this allows a detailed comparison with some pertinent results already discussed in the literature, and leads to the determination of new conservation laws for the Euler equations.

The plan of the paper is as follows. The construction of composite variational principles is outlined briefly in Sec. II. The approach to conservation laws is described in Sec. III. Sections IV and V deal with the equations modeling the motion of a viscous imcompressible fluid (Navier–Stokes equations) and of an inviscid barotropic compressible fluid (Euler equations), respectively. Additional comments are found in Sec. VI.

II. ESSENTIALS ON COMPOSITE VARIATIONAL PRINCIPLES

Consider the following system of partial differential equations:

$$F_a(x^\alpha,\phi^c,\phi^c_\alpha\phi^c_{\alpha\beta})=0 \tag{2.1}$$

in the unknowns $\phi^c(x^\alpha)$, where Greek (Latin) indices vary from 1 to n (m), $\phi^c_\alpha=\partial\phi^c/\partial x^\alpha$, $\phi^c_{\alpha\beta}=\partial^2\phi^c/\partial x^\alpha\,\partial x^\beta$, and the F_a's are sufficiently regular functions of their arguments. To save writing we omit all hypotheses concerning continuity, differentiability, etc., whenever it is clear from the context what these ought to be. It is well known that the system (2.1) admits a variational formulation if and only if it is self-adjoint,[7,10,11] which means that a set of quite stringent conditions involving partial derivations of the F's must be identically satisfied. However, one can always look for a variational formulation of (2.1) in terms of a so-called composite variational principle.[7,8,12] This requires the introduction of a suitable set of additional variables $\tilde\eta^a(x^\alpha)$, added to the original field variables ϕ.

Namely, consider the following action functional:

$$A(\phi,\tilde n)=\int_V F_a\tilde\eta^a\,dx^1\cdots dx^n, \tag{2.2}$$

where V is a suitable domain in R^n. The related stationarity conditions under arbitrary variations of $\tilde\eta$ and ϕ imply the original field equations (2.1)—not involving $\tilde\eta$—and

$$\widetilde M_c(\tilde\eta):=\frac{\partial F_a}{\partial\phi^c}\tilde\eta^a-D_\alpha\left(\frac{\partial F_a}{\partial\phi^c_\alpha}\tilde\eta^a\right)+D_{\alpha\beta}\left(\frac{\partial F_a}{\partial\phi^c_{\alpha\beta}}\tilde\eta^a\right)=0, \tag{2.3}$$

where D_α denotes the total derivative with respect to x^α and $D_{\alpha\beta}=D_\alpha\circ D_\beta$.

Clearly, the previous process can be adapted straightforwardly to deal with arbitrary, not necessarily second-order, systems. Thus it follows that any system of partial differential equations may be embedded in a set of Euler–Lagrange equations yielding the stationarity conditions for the functional (2.2). In particular, Eq. (2.3) determines the additional variables $\tilde\eta$ and should also provide their physical interpretation. In fact, a similar approach to purely mechanical systems has led to the introduction of an oscillator with "negative" friction (cf. Ref. 12, p. 298); later, the

 0022-2488/88/040812-05$02.50

additional variables and corresponding equations have been related to one-forms invariant along the trajectories of the given system and thus play a natural role as generators of constants of motion.[13] It will be shown that this interpretation can be extended to general field theories.

III. OPERATIVE APPROACH TO CONSERVATION LAWS

Consider an n-dimensional vector with local components $I^\alpha = I^\alpha(x^\beta, \phi^b, \phi^b_\beta, \phi^b_{\beta\sigma}, ...)$; notice that dependence on higher derivatives of the ϕ's is allowed explicitly. A *conservation law* for the system (2.1) is an equation of the form

$$D_\alpha I^\alpha \doteq 0, \tag{3.1}$$

where the $\doteq$ symbol means that equality holds on solutions to (2.1). Whenever (3.1) holds the vector I^α is said to be conserved or divergence-free. In this section a suitable formulation of Noether's theorem is applied to the functional (2.2), with the aim of finding conservation laws—and the associated conserved vectors—related to the additional variables $\tilde{\eta}$.

Accordingly, we consider a local one-parameter group of transformations of the form

$$\bar{x}^\alpha = x^\alpha + \epsilon\tau^\alpha, \quad \bar{\phi}^c = \phi^c + \epsilon\xi^c, \quad \bar{\tilde{\eta}}^c = \tilde{\eta}^c + \epsilon\lambda^c, \tag{3.2}$$

where ϵ denotes as usual an infinitesimal parameter and the generators τ, ξ, and λ are allowed to depend on x and ϕ. Expressing invariance of the functional A under the transformation (3.2) to within the integral of a divergence field of the form $\epsilon D_\alpha W^\alpha(x^\beta, \phi^b, \phi^b_\beta)$ and considering first-order terms in the parameter ϵ, we find the invariance condition[14]

$$M_a(\eta)\tilde{\eta}^a + \tau^\alpha D_\alpha F_a \tilde{\eta}^a + F_a(\lambda^a + \tilde{\eta}^a D_\alpha \tau^\alpha) - D_\alpha W^\alpha \doteq 0, \tag{3.3}$$

where

$$M_a(\eta) := \eta^c \frac{\partial F_a}{\partial \phi^c} + D_\alpha \eta^c \frac{\partial F_a}{\partial \phi^c_\alpha} + D_{\alpha\beta}\eta^c \frac{\partial F_a}{\partial \phi^c_{\alpha\beta}}, \tag{3.4}$$

with η^c defined by

$$\eta^c = \xi^c - \phi^c_\sigma \tau^\sigma. \tag{3.5}$$

Taking into account (3.4) and (3.5) and recalling the definition (2.3) of $\widetilde{M}_c(\tilde{\eta})$, the lhs of (3.3) can be transformed to read as

$$\eta^c \widetilde{M}_c(\tilde{\eta}) + \tau^\alpha D_\alpha F_c \tilde{\eta}^c + F_a(\lambda^a + \tilde{\eta}^a D_\alpha \tau^\alpha) + D_\alpha J^\alpha - D_\alpha W^\alpha \doteq 0, \tag{3.6}$$

with

$$J^\alpha := \eta^c \frac{\partial F_a}{\partial \phi^c_\alpha}\tilde{\eta}^a + D_\beta \eta^c \frac{\partial F_a}{\partial \phi^c_{\alpha\beta}} - \eta^c D_\beta\left(\frac{\partial F_a}{\partial \phi^c_{\beta\alpha}}\tilde{\eta}^a\right). \tag{3.7}$$

On recalling that $\widetilde{M}_c(\tilde{\eta}) \doteq 0$, $D_\alpha F_c \doteq 0$, and $F_\alpha \doteq 0$, it turns out that (3.6) is written in the form of a conservation law, with related conserved current $I^\alpha = J^\alpha - W^\alpha$. However, (3.6) is simply a reformulation of the invariance condition (3.3) and so we are led to conclude that $J^\alpha - W^\alpha$ yields a Noether-type conservation law pertaining to the composite variational principle (2.2). In addition, the dependence of J^α on $\tilde{\eta}$ shows that such a conservation law is related to the additional variables.

To illustrate the operative features of this approach, it is convenient to analyze in detail Eq. (3.3). In principle, the variables $\tilde{\eta}$ are to be determined as solutions to the stationarity conditions (2.1) and (2.3); however, on observing that (2.1) does not involve the $\tilde{\eta}$'s, it is found that the $\tilde{\eta}$'s are only required to satisfy the restriction of (2.3) to solutions of (2.1). Thus, in actual practice, no explicit representation for ϕ is required in solving (2.3): simply, we let $\tilde{\eta}^a = \tilde{\eta}^a(x^\alpha, \phi^c, \phi^c_\alpha, ...)$; we substitute into (2.3); and then we recall that, under rather general assumptions,[3] the condition $\widetilde{M}_c(\tilde{\eta}) \doteq 0$ reduces to the fact that $\widetilde{M}_c(\tilde{\eta})$ is a linear combination of (2.1) and possibly of its total derivatives. This suffices to find the explicit form of the $\tilde{\eta}$'s. Now (3.3) can be imposed and the expressions of η and W will follow. Notice that η and W may be allowed to depend on higher derivatives of the ϕ's, that the simplifying conditions $D_\alpha F_a \doteq 0$ and $F_a \doteq 0$ have to be taken into account, and that λ^a is multiplied by F_a and hence does not enter the conserved vector. We have thus proved the following theorem.

Theorem: Consider $\tilde{\eta}$, η, and W as functions of x^α, $\phi^c, \phi^c_\alpha, \phi^c_{\alpha\beta}, ...$ satisfying

$$\widetilde{M}_c(\tilde{\eta}) \doteq 0, \tag{3.8a}$$

$$M_a(\eta)\tilde{\eta}^a - D_\alpha W^\alpha \doteq 0. \tag{3.8b}$$

Then the vector $I^\alpha = J^\alpha - W^\alpha$, with J^α given by (3.7), is conserved.

Several remarks are in order now. First it is convenient to observe that condition (4.8a) corresponds to Eq. (5.52) in Ref. 3, thus showing that the $\tilde{\eta}$'s are solutions to the conditions necessarily obeyed by characteristics of conservation laws for Eq. (2.1) (see Proposition 5.33 in Ref. 3). Thus the theorem shows that such functions $\tilde{\eta}$ can be used effectively to construct conservation laws through the knowledge of solutions to (3.8b), that is, in particular through the knowledge of generalized symmetries of Eq. (2.1), as clarified by the subsequent discussion. It is also worth pointing out that any conserved vector may be represented in the form $I^\alpha = J^\alpha - W^\alpha$, provided W^α is suitably chosen. Notice also that the presence of the term $D_\alpha W^\alpha$ in (3.8b) makes this equation largely underdetermined and may lead to considerable simplifications in the integration procedure.

A further comparison with pertinent results already available can be made if we examine the case $W^\alpha = 0$. Then a sufficient condition for the validity of (3.8b) is simply given by

$$M_a(\eta) \doteq 0. \tag{3.9}$$

On comparing with (3.4) and (3.5) it is found that (3.9) identifies the generalized symmetries of the system (2.1) (see Sec. 5.1 in Ref. 3). If in addition $\tilde{\eta}$, τ, and ξ are not allowed to depend on the derivatives of the ϕ's we recover the results of Ref. 6.

Equation (3.8a) is the formal adjoint[11] to (3.9), i.e., to the definition of generalized symmetry transformation. This observation suggests a possible interpretation of the meaning of the additional variables $\tilde{\eta}$ as solutions to the adjoint of the definition of generalized symmetries. Moreover, this connection between additional variables and symmetries accounts for the role played by the $\tilde{\eta}$'s as generators of conservation laws. Let us also recall that if the definition (3.9) of generalized symmetries is self-adjoint then the system (2.1) is derivable from a variational principle,[6,7,10,11] just as it happens in the case of purely mechanical systems[8,9]; in such a case the added variables identify generalized symmetries and the theorem shows that a conservation law is always associated with pairs of generalized symmetries.

The method described thus far leads to the construction of conservation laws without any need for additional restrictions on the structure of the given equations (2.1), e.g., that they come from a variational principle or may be represented in Hamiltonian form. Moreover, the previous procedure may be modified straightforwardly to deal with systems of partial differential equations of any order. However, we are not concerned with those conservation laws that are *trivial*, which means that they are generated by a vector I satisfying $D_\alpha I^\alpha = 0$ identically, or $I^\alpha \doteq 0$. Consequently we regard as *equivalent* any two conserved currents that can be combined linearly to give rise to a trival conservation law; two conserved currents that are not equivalent are said to be *independent*.[3] In the following we are only concerned with independent conservation laws. Moreover, we also make use of the fact that equivalent conservation laws have the same physical interpretation.

IV. NAVIER–STOKES EQUATIONS

We are now in a position to apply the algorithms described in Sec. III in order to find conservation laws. To illustrate the method we consider the system modeling the motion of a viscous incompressible fluid under the assumption that body forces may be neglected. Following the Eulerian description, the equations satisfied by the velocity field $\mathbf{v} = u\mathbf{e}_1 + v\mathbf{e}_2 + w\mathbf{e}_3$ and the pressure p read as

$$\mathbf{v}_t + (\mathbf{v}\cdot\nabla)\mathbf{v} + \nabla p - \nu\nabla^2\mathbf{v} = 0, \tag{4.1a}$$

$$\nabla\cdot\mathbf{v} = 0, \tag{4.1b}$$

where ν is the (constant) kinematic viscosity; $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ are the unit vectors of the given Cartesian axes; ∇ denotes the gradient operator; and the subscript t denotes partial differentiation with respect to t, that is, $\mathbf{v}_t = \partial\mathbf{v}/\partial t$.

The independent variables x^α ($\alpha = 1,...,4$) are identified with the spatial coordinates x, y, and z and the time t. The field equations yield the definitions of $F_1,...,F_4$. Finally, we let the ϕ^c's ($c = 1,...,4$) coincide with u, v, w, and p, respectively. To adhere to the usual terminology the t component of a conserved current is referred to as a conserved *density*, whereas the space components yield the corresponding *flux*.[1,3,15] The meaning of the conserved currents is discussed on the basis of an analysis of the associated densities. In so doing we also take into account the fact that conserved densities are defined up to a divergence taken over the space variables.

The generators of symmetry transformations for Navier–Stokes equations (4.1) have already been determined under the assumption that τ and ξ do not depend on the derivatives of the field variables[16]; moreover, the equations yielding the additional variables $\tilde{\eta}$ have been written down and solved explicitly elsewhere.[17] Accordingly, we only reproduce the general solutions and devote the remaining part of this section to a detailed analysis of the related conservation laws, which was not given in the already cited references.

Explicitly, we have

$$\begin{aligned}
&\tau^1 = a_2x - a_3y - a_4z + f, \quad \tau^2 = a_2y + a_3x - a_5z + g,\\
&\tau^3 = a_2z + a_4x + a_5y + h, \quad \tau^4 = a_1 + 2a_2t,\\
&\xi^1 = -a_2u - a_3v - a_4w + f_t,\\
&\xi^2 = -a_2v + a_3u - a_5w + g_t,\\
&\xi^3 = -a_2w + a_4u + a_5v + h_t,\\
&\xi^4 = -2a_2p + j - xf_{tt} - yg_{tt} - zh_{tt},
\end{aligned} \tag{4.2}$$

where $a_1,...,a_5$ are five arbitrary parameters, and f, g, h, and j are arbitrary functions of t. If all but one of these is set equal to zero, the remaining one describes an independent generator of the group of symmetry transformations. The physical interpretation of the generators has been discussed elsewhere.[16] Thus we concentrate our attention on the added variables. The most general family of functions $\tilde{\eta}^c(x,y,...,p)$ satisfying the system $\widetilde{M}_a(\tilde{\eta}) \doteq 0$ reads as[17]

$$\begin{aligned}
&\tilde{\eta}^1 = b_1y + b_2z + k, \quad \tilde{\eta}^2 = -b_1x + b_3z + m,\\
&\tilde{\eta}^3 = -b_2x - b_3y + n,\\
&\tilde{\eta}^4 = \tilde{\eta}^1u + \tilde{\eta}^2v + \tilde{\eta}^3w - (xk_t + ym_t + zn_t) + o,
\end{aligned} \tag{4.3}$$

where b_1, b_2, and b_3 are arbitrary parameters; k, m, n, and o are arbitrary functions of the time t.

With the aim of finding conserved densities, let us observe preliminarily that in view of (4.1) expression (3.7) for J^t reduces to

$$J^t = \tilde{\eta}^1\eta^1 + \tilde{\eta}^2\eta^2 + \tilde{\eta}^3\eta^3. \tag{4.4}$$

Recalling that $\tilde{\eta}$ is given by (4.3) and that η is formed by comparison of (3.5) and (4.2), substitution into (4.4) yields the required conserved densities. After long and cumbersome calculations, which involve repeated use of the fact that J^t is defined up to a spatial divergence, we find the following six independent conserved densities:

$$\begin{aligned}
&J_1^t = zv - yw, \quad J_2^t = xw - zu, \quad J_3^t = xv - yu,\\
&J_4^t = q(t)u, \quad J_5^t = r(t)v, \quad J_6^t = s(t)w,
\end{aligned} \tag{4.5}$$

where q, r, and s are arbitrary functions of t. The fact that each of the above densities gives rise to a conservation law can also be proved directly. The physical interpretation of these conserved densities is easily obtained: J_1^t, J_2^t, and J_3^t correspond to conservation of angular momentum; and J_4^t, J_5^t, and J_6^t yield conservation of momentum provided q, r, and s are set equal to 1.

In principle, further conservation laws can be generated by application of the symmetry group generators described in (4.2) to the conserved currents $J_1,...,J_6$ (see Ref. 3, Sec. 5.3). However, rather long calculations show that these procedures do not give rise to new independent conservation laws. It is also to be pointed out that J_4, J_5, and J_6 identify an infinite number of independent conservation laws, e.g., by taking $q_n(t) = t^n$, where n goes from 0 to infinity. One may wonder whether an additional infinite set of conservation laws can be generated by multiplication of the components of the angular momentum by an arbitrary function of time. It turns out that this result can really be achieved, but the corresponding conservation law is nonlocal. Specifically, if one considers the density $Q(zv - yw)$, which is obtained when J_1^t is multiplied by an arbitrary function of time $Q(t)$, one finds that the components of the corresponding conserved flux are given by

$$\begin{aligned}
&-Q_t \int^x (zv - yw)\,d\lambda \\
&\quad - Q\,[y(uw - vw_x) - z(uv - vv_x)], \\
&-Q\,[y(vw - vw_y) - z(v^2 + p - vv_y) + vw], \\
&-Q\,[y(w^2 + p - vw_z) - z(vw - vv_z) - vv],
\end{aligned}$$

where the definite integral is taken over the space variable x.

By making use of the condition $\nabla\cdot\mathbf{v} = 0$ we can form another infinite set of conservation laws, where the conserved density is

$$-n \int^t (c_1 x^{n-1} u + c_2 y^{n-1} v + c_3 z^{n-1} w)\,d\lambda$$

and the corresponding flux reads as $(c_1 x^n + c_2 y^n + c_3 z^n)\mathbf{v}$; here c_2, c_2, and c_3 are arbitrary constants; n goes from 1 to infinity; and the integral is evaluated with respect to the time variable.

V. EULER EQUATIONS

Consider now an inviscid, barotropic, compressible fluid. The balance laws of momentum and mass may be written in the usual form:

$$\begin{aligned}
&\mathbf{v}_t + (\mathbf{v}\cdot\nabla)\mathbf{v} + P\,\nabla\rho = 0, \\
&\rho_t + \nabla\cdot(\rho\mathbf{v}) = 0,
\end{aligned} \tag{5.1}$$

where ρ is the mass density; P is given by $P = (1/\rho)dp/d\rho$, with $p = p(\rho)$; and we have adopted the conventions of Sec. IV, with the only exception that ϕ^4 is to be identified with the density ρ.

The generators of symmetry transformations are given by[18]

$$\begin{aligned}
&\tau^1 = a_1 x - a_2 y - a_3 z + a_4 t + a_5, \\
&\tau^2 = a_2 x + a_1 y + a_6 z + a_7 t + a_8, \\
&\tau^3 = a_3 x - a_6 y + a_1 z + a_9 t + a_{10}, \quad \tau^4 = a_1 t + a_{11}, \\
&\xi^1 = -a_2 v - a_3 w + a_4, \quad \xi^2 = a_2 u + a_6 w + a_7, \\
&\xi^3 = a_3 u - a_6 v + a_9, \quad \xi^4 = 0,
\end{aligned} \tag{5.2}$$

where $a_1,\ldots,a_{11}$ are arbitrary parameters. The most general solution of the form $\tilde\eta^c = \tilde\eta^c(x,\ldots,\rho)$ to the system (3.8a) reads as[18]

$$\begin{aligned}
\tilde\eta^1 &= (b_1 u - b_2 y - b_3 z + b_4 t + b_5)\rho, \\
\tilde\eta^2 &= (b_1 v + b_2 x + b_6 z + b_7 t + b_8)\rho, \\
\tilde\eta^3 &= (b_1 w + b_3 x - b_6 y + b_9 t + b_{10})\rho, \\
\tilde\eta^4 &= \frac{1}{2} b_1\left(u^2 + v^2 + w^2 + 2\int^\rho P(\lambda)\,d\lambda\right) \\
&\quad + b_5 u + b_8 v + b_{10} w + b_{11} + (b_2 v + b_3 w - b_4)x \\
&\quad - (b_2 u + b_6 w + b_7)y + (b_6 v - b_3 u - b_9)z \\
&\quad + (b_4 u + b_7 v + b_9 w)t,
\end{aligned} \tag{5.3}$$

with $b_1,\ldots,b_{11}$ arbitrary constants.

We will now study in detail the conserved densities generated by (5.2) and (5.3). In this case the expression of J^t reduces to

$$J^t = \tilde\eta^1\eta^1 + \tilde\eta^2\eta^2 + \tilde\eta^3\eta^3 + \tilde\eta^4\eta^4. \tag{5.4}$$

In view of (5.2) and (5.3) it may be shown that (5.4) gives rise to the following list of independent conserved densities:

$$\begin{aligned}
&J_1^t = \rho u, \quad J_2^t = \rho v, \\
&J_3^t = \rho w, \quad J_4^t = (zv - yw)\rho, \\
&J_5^t = (xw - zu)\rho, \quad J_6^t = (xu - yv)\rho, \\
&J_7^t = (ut - x)\rho, \quad J_8^t = (vt - y)\rho, \\
&J_9^t = (wt - z)\rho, \quad J_{10}^t = \rho, \\
&J_{11}^t = \frac{1}{2}(u^2 + v^2 + w^2) + \int^\rho d\lambda \int^\lambda P(\tau)\,d\tau.
\end{aligned}$$

The physical interpretation of these conserved densities can be derived by observing that J_1, J_2, and J_3 correspond to conservation of momentum; J_4, J_5, and J_6 describe conservation of angular momentum; and J_{10} and J_{11} yield conservation of mass and energy, respectively. In order to discuss the meaning of the remaining densities, it suffices to examine one of them, say J_7^t. The form of this density shows that it is related to the motion of the center of mass[19] provided we recall that here x is to be evaluated along the trajectories of the particles of the fluid. This interpretation is understood most easily if we observe that under the assumption that u, v, and w behave suitably at the boundary, the integral of J_7^t over the space variables is constant in time and coincides with the value of the x coordinate of the center of mass at the time $t = 0$. Notice also that a counterpart for J_7^t in the Lagrangian description has already been discussed in Ref. 14.

It seems that J_7^t, J_8^t, and J_9^t give rise to new conservation laws for perfect fluid motion, whereas the remaining densities are already well known. In fact, it has been shown that energy, momentum, angular momentum, and the so-called total helicity, with density[20] $\mathbf{v}\cdot(\nabla\wedge\mathbf{v})$, exhaust the class of local invariants depending on $\mathbf{v}$ and its first-order derivatives[21] in the case of an incompressible fluid[22]; in this connection, the previous conclusions show that this result cannot be extended to the model of compressible fluids, if dependence on position is allowed.

Finally, to recover conservation of total helicity within the present framework, we have to consider the possibility that additional variables depend on the derivatives of $\mathbf{v}$. Then it is to be verified that

$$\tilde{\eta}^1 = w_y - v_z, \quad \tilde{\eta}^2 = u_z - w_x, \quad \tilde{\eta}^3 = v_x - u_y, \quad \tilde{\eta}^4 = 0$$

yield a solution to (2.3) in the case of a compressible fluid. Thus substitution into (5.4) under the assumption that $a_1 = 1$ and the remaining constants vanish leads to conservation of total helicity. When $a_2,\ldots,a_{11}$ in turn are set equal to 1 no other significant conserved density is obtained.

VI. COMMENTS AND CONCLUSIONS

In this paper we have described an algorithm leading to the practical construction of conservation laws for any given system of partial differential equations. Our procedure does not require *a priori* restrictions either on the form of the given equations or on the number of dependent and independent variables, besides the usual differentiability assumptions. Indeed, the present approach seems quite simple in the sense that it depends on the application of an appropriate version of Noether's theorem to a suitably defined composite variational principle.

First, our procedure requires the determination of the additional variables $\tilde{\eta}$ as functions of the independent variables, of the field functions, and, possibly, of their higher order derivatives; in practice, the $\tilde{\eta}$'s are found by integration of the linear system (3.8a), which is a subset of the stationarity conditions for the already mentioned composite variational principle. Second, we need η and W as solutions to Eq. (3.8b). It is worth noting that this condition may be fulfilled when the η's are related to the generators of generalized symmetry transformations for the original system of differential equations provided we introduce the simplifying assumption $W = 0$. In this case a considerable simplification is achieved because the infinitesimal symmetry transformations of a great number of significant equations have already been determined in connection with the problem of finding the so-called similarity solutions.[3,23] Third, the final expression of the conserved vectors is given as $J^\alpha - W^\alpha$, with J^α defined by Eq. (3.7).

The algorithm used to find conservation laws can work even though a limited amount of information is available. For instance, one can construct an infinite number of independent conservation laws by using an infinite class of symmetry transformations depending on higher order derivatives of the unknowns and a single solution of Eq. (3.8a), as in the case of the Korteweg–de Vries equation.[24] Alternatively, one can make use of an infinite class of solutions to the adjoint equation (3.8a) in order to combine them with a generator of symmetry transformation and thus find an infinite class of conservation laws: this possibility has been dealt with in the analysis of the heat equation; the corresponding conserved densities yield the initial values of the moments of the solution, in terms of which the time development of the solution can be recovered.[24]

In most cases, however, the problem of finding $\tilde{\eta}$, η, and W as functions of the higher order derivatives is rather cumbersome, as in the case of the Euler and Navier–Stokes equations. Thus we have introduced the simplifying assumptions that W vanishes and that the symmetry generators and additional variables are independent of the derivatives; this requirement has been relaxed only in the derivation of conservation of total helicity. In spite of these stringent conditions, we have found an infinite set of conserved currents of Navier-Stokes equations and we have explored the possible existence of further infinite families of nonlocal divergence-free vectors. As to compressible inviscid fluids, our approach has led to the determination of a new family of conserved currents that are related to the motion of the center of mass.

ACKNOWLEDGMENTS

The author thanks an anonymous referee for many helpful suggestions.

This paper has been partially supported by the National Group for Mathematical Physics of the Italian Research Council and by the Italian Ministry of the Public Education through the research project "Problemi di evoluzione nei fluidi e nei solidi."

[1] R. M. Miura, SIAM Rev. **18**, 412 (1976).
[2] H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. **16**, 1 (1975); P. J. Olver, Math. Proc. Camb. Philos. Soc. **94**, 529 (1983); M. Wadati, H. Sanuki, and K. Konno, Prog. Theor. Phys. **53**, 419 (1975); P. D. Lax, Commun. Pure Appl. Math. **21**, 467 (1968); E. S. Suhubi, Int. J. Eng. Sci. **22**, 119 (1984); W. A. Strauss, in *Invariant Wave Equations*, edited by G. Velo and A. S. Wightman (Springer, New York, 1978), p. 197.
[3] P. J. Olver, *Applications of Lie Groups to Differential Equations* (Springer, New York, 1986).
[4] B. K. Harrison and F. B. Estabrook, J. Math. Phys. **12**, 653 (1971).
[5] D. G. B. Edelen, *Applied Exterior Calculus* (Wiley, New York, 1985).
[6] G. Caviglia, J. Math. Phys. **27**, 972 (1986).
[7] R. W. Atherton and G. M. Homsy, Stud. Appl. Math. **54**, 31 (1975).
[8] G. Caviglia, Int. J. Theor. Phys. **25**, 147 (1986).
[9] G. Caviglia, Inv. Prob. **1**, L13 (1985).
[10] F. Bampi and A. Morro, J. Math. Phys. **23**, 2312 (1982).
[11] R. M. Santilli, Ann. Phys. (NY) **103**, 354, 409 (1977); **105**, 227 (1977).
[12] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. 1.
[13] See Ref. 8. An alternative interpretation has been proposed by F. Cantrijn, M. Crampin, and W. Sarlet, "Evading the inverse problem for second order ordinary differential equations by using additional variables," preprint.
[14] G. Caviglia and A. Morro, J. Math. Phys. **28**, 1056 (1987). A detailed discussion of this point is omitted because it follows straightforwardly from Sec. II of the cited reference.
[15] T. B. Benjamin and P. J. Olver, J. Fluid Mech. **125**, 653 (1982); M. D. Kruskal, R. M. Miura, and C. S. Gardner, J. Math. Phys. **11**, 952 (1970).
[16] S. P. Lloyd, Acta Mech. **38**, 85 (1981); R. E. Boisvert, W. F. Ames, and U. N. Srivastava, J. Eng. Math. **17**, 201 (1983).
[17] G. Caviglia, Int. J. Eng. Sci. **24**, 1295 (1986).
[18] M. Benati and G. Caviglia, Int. J. Eng. Sci. **25**, 1 (1987).
[19] E. L. Hill, Rev. Mod. Phys. **23**, 253 (1951).
[20] H. K. Moffatt, J. Fluid Mech. **35**, 117 (1969).
[21] D. Serre, C. R. Acad. Sci. Paris **289**, 267 (1979).
[22] P. J. Olver, J. Math. Anal. Appl. **89**, 233 (1982).
[23] W. F. Ames, *Differential Equations in Engineering* (Academic, New York, 1972), Vol. 2; L. V. Ovsjannikov, *Group Analysis of Differential Equations* (Academic, New York, 1982).
[24] G. Caviglia, in *Atti VIII Congresso Nazionale AIMETA*, Vol. 2, Torino, 29 September–3 October 1986, pp. 743–747.

Fourier transforms and the dilatation group

P. Hofstee and A. Tip
FOM-Institute for Atomic and Molecular Physics, Kruislaan 407, 1098 SJ Amsterdam, The Netherlands

(Received 17 June 1987; accepted for publication 21 October 1987)

The Fourier transform $g(k)$ of a square integrable function $f(x)$, vanishing for $x<0$, is analytic in the upper half plane, so that, replacing k by $k\exp\zeta$, $k\geqslant 0$, $0<\operatorname{Im}\zeta<\pi$, it can be associated with an operator $K(\zeta)$ in $\mathscr{H}_+ = L^2((0,\infty),dx)$. The operator $K(\zeta)$ can be expressed in terms of the generator D of the dilatation group on $\mathscr{H}_+$ and it can be shown that it is analytic in the strip $0<\operatorname{Im}\zeta<\pi$ with strong limits as $\operatorname{Im}\zeta\downarrow 0$ and $\uparrow\pi$. The Laplace transform ($\zeta = i\pi/2$) is an analytic vector for D. It is also found that D is not a spectral operator of scalar type on $L^p((0,\infty),dx)$, $1\leqslant p<\infty$, $p\neq 2$. Applying the results obtained here to the time-evolution operator for a one-dimensional Sommerfeld model for the interaction between an electron and a metal, it is found that this operator has a complex-dilated analytic extension.

I. INTRODUCTION

The Fourier transform of a square integrable function $f(x)$ vanishing on a half-line, i.e.,

$$g(z) = (2\pi)^{-1/2}\int_0^\infty dx\exp[izx]f(x),\quad \operatorname{Im} z\geqslant 0, \tag{1.1}$$

is one of the matters considered by Paley and Wiener in their classic treatise on Fourier transforms in the complex domain.[1] Thus they showed that $g(z)=g(u+iv)$ is analytic in the open upper half plane, is square integrable in u for fixed v, and has a limit in the strong L^2 sense as $v\downarrow 0$. Instead of considering the square integrability as a function of z restricted to a line parallel to $\mathbb{R}$ in $\mathbb{C}$, we can also investigate the square integrability of g as z runs through a ray in the upper half plane originating in zero. Thus we put

$$g(k,\zeta) = (2\pi)^{-1/2}\exp\left[\frac{\zeta}{2}\right]\int_0^\infty dx \times\exp[ie^{\zeta}kx]f(x) = (K(\zeta)f)(k), \tag{1.2}$$

where $k\geqslant 0$ and $\zeta = \vartheta + i\psi$, $\vartheta\in\mathbb{R}$, $\psi\in[0,\pi]$, and consider $K(\zeta)$ as an operator that maps $\mathscr{H}_+$ ($\mathscr{H}_\pm = L^2(\mathbb{R}_\pm,dx)$, $\mathbb{R}_+ = [0,\infty]$, $\mathbb{R}_- = (-\infty,0]$, $\mathscr{H} = \mathscr{H}_+\oplus\mathscr{H}_- = L^2(\mathbb{R},dx)$) into itself. Recalling that the dilatation transformation is defined as the map

$$(U(\vartheta)f) = (\exp[i\vartheta D]f)(x) = \exp[\vartheta/2]f(e^{\vartheta}x),\quad \vartheta\in\mathbb{R}, \tag{1.3}$$

we see that $g(\zeta)$ and $g(\zeta')$ are connected by a complex dilatation transformation. We note further that $g(i\pi/2)$ is, up to a factor, the Laplace transform. In the next section, where we give a precise meaning to $K(\vartheta)$, we show that $g(i\pi/2)$ is an analytic vector for D, the generator of dilatations, and that the Fourier transforms $g(0)$ and $g(i\pi)$ are strong limits of $g(\zeta)$ as $\zeta\to 0$ and $\zeta\to i\pi$, respectively. In fact we obtain an expression for $K(\zeta)$ in terms of D, which can be extended to the Fourier operator F:

$$(Ff)(k) = (2\pi)^{-1/2}\int_{-\infty}^{+\infty}dx\exp[ikx]f(x),\quad f\in\mathscr{H}. \tag{1.4}$$

At this point we note in passing that another representation of F is

$$F = \exp[-i\pi/2H] = R\exp[i\pi/2H], \tag{1.5}$$

where R is the reflection or parity operator, $(Rf)(x) = f(-x)$, and $H = \frac{1}{2}(p^2+x^2-1)$ is the self-adjoint closure of $\frac{1}{2}(-\partial_x^2+x^2-1)$ acting in $\mathscr{H}$.

Equation (1.5) follows from the observation that the Hermite functions $u_n(x)$, which are the eigenfunctions of H with corresponding eigenvalue n ($n=0,1,2,\ldots$), constitute an orthonormal basis for $\mathscr{H}$ and have the property $(Fu_n)(k) = (i)^n u_n(k) = i^n u_n(-k)$. It does not have an obvious extension to other $L^p(\mathbb{R},dx)$ spaces, whereas ours apparently has one. But, as discussed in Sec. III, this is not the case. In fact we find that D, appropriately defined, is not a spectral operator of scalar type.[2] In Sec. IV we present an application. We consider the time-evolution operator $\exp[-iHt]$ associated with

$$H = p^2 - \mu\theta(x),\quad \mu>0, \tag{1.6}$$

the closure of $-\partial_x^2 - \mu\theta(x)$, $\theta(x)=0$, $x<0$, $\theta(x)=1$, $x\geqslant 0$ (H is the one-dimensional version of the Sommerfeld Hamiltonian describing the interaction of an electron with a metal filling a half space). We show that under the dilatation transformation, restricted to $x\geqslant 0$, $\exp[-iHt]$ transforms into $\exp[-iH(\zeta)t]$, analytic in ζ. This result is not covered by the cases considered by Kato[3] since H does not generate a holomorphic semigroup.

II. THE OPERATOR $K(\zeta)$

On $L^p = L^p(\mathbb{R},dx)$, $1\leqslant p<\infty$, the dilatation group is the strongly continuous group of isometric operators defined by

$$(U(\vartheta)f)(x) = (\exp[iD\vartheta]f)(x) = \exp[\vartheta/p]f(e^{\vartheta}x),\quad f\in L^p. \tag{2.1}$$

We note that $L^p_\pm = L^p(\mathbb{R}_\pm,dx)$ are invariant subspaces under the action of $U(\vartheta)$. We now consider its restriction to $L^p_\pm$, again denoted by $U(\vartheta)$. Let T be the 1–1 isometric map from L^p_+ to L^p defined by

 0022-2488/88/040817-06$02.50

$$\phi(u) = (Tf)(u) = \exp[u/p]\, f(e^u). \tag{2.2}$$

Then $V(\vartheta)$ defined by $V(\vartheta)\phi = TU(\vartheta)f$ acts according to

$$(V(\vartheta)g)(u) = \phi(u+\vartheta), \tag{2.3}$$

i.e., $\{V(\vartheta)\,|\,\vartheta\in\mathbb{R}\}$ is the group of translations of L^p. Thus there exists a 1–1 correspondence between $U(\vartheta)$ on L^p and $V(\vartheta)$ on $L^p \oplus L^p$. The generator of $V(\vartheta)$ on $\mathscr{H} = L^2(\mathbb{R},dx)$ is the closure of $-i\partial_x$, the momentum operator, which has spectrum $\mathbb{R}$ and has purely absolutely continuous spectrum. Thus we obtain the following.

Proposition 1: D acting in either $\mathscr{H}$, $\mathscr{H}_+$, or $\mathscr{H}_-$ has spectrum $\sigma(D) = \mathbb{R}$, and $\sigma(D)$ is purely absolutely continuous, $\sigma(D) = \sigma_{\text{ac}}(D)$.

We now continue with the discussion of $K(\zeta)$ on $\mathscr{H}^+$. For later convenience we introduce the operator L,

$$(Lf)(x) = x^{-1}f(x^{-1}), \quad f\in\mathscr{H}^+. \tag{2.4}$$

The operator L is bounded, self-adjoint, and unitary (since $L^2 = 1$) and has ± 1 as its spectrum. Thus

$$L = P_+ - P_- = \exp[i\pi P_-], \tag{2.5}$$

where P_+ and P_- are the eigenprojectors associated with the eigenvalues $+1$ and -1, respectively. We also have

$$LU(\vartheta) = U(-\vartheta)L. \tag{2.6}$$

The functions $f_n(x)$, $n = 0,1,2,\ldots$, are defined by

$$f_n(x) = x^n \exp[-x]. \tag{2.7}$$

They constitute a fundamental set in $\mathscr{H}_+$ since the Laguerre basis consists of finite linear combinations of the f_n's. The action of $K(\zeta)$ on f_n is readily calculated:

$$\begin{aligned} g_n(k,\zeta) &= (K(\zeta)f_n)(k) \\ &= n!(2\pi)^{-1/2}e^{\zeta/2}[1 - ie^{\zeta}k]^{-(n+1)}, \end{aligned} \tag{2.8}$$

which is in $\mathscr{H}_+$ for $\zeta = \vartheta + i\psi\in\mathbb{C}$, $\psi\in[0,\pi]$. Thus $K(\zeta)$ is densely defined. [It is also easily checked that $K(\zeta)f\in\mathscr{H}_+$ for the dense set of continuous $f\in\mathscr{H}_+$ with compact support bounded away from zero.]

Proposition 2: $K(\zeta)$ defined by (2.8) can be extended to a bounded linear operator on $\mathscr{H}_+$, which is analytic in ζ in the strip $\psi = \operatorname{Im}\zeta\in(0,\pi)$ and has strong limits $K(\vartheta)$ and $K(\vartheta + i\pi)$ as $\psi\downarrow 0$ and $\psi\uparrow\pi$, respectively. These limits are strongly continuous in ϑ.

Proof: Let $f = \sum_{n=0}^{N} a_n f_n$, $a_n\in\mathbb{C}$ and f_n given by (2.7). Then for $\psi\in(0,\pi)$,

$$\begin{aligned} g(k,\zeta) &= (2\pi)^{-1/2}e^{\zeta/2}\int_0^\infty dy\,\exp[ie^{\zeta}y]k^{-1}f(yk^{-1}) \\ &= (2\pi)^{-1/2}e^{\zeta/2}\int_{-\infty}^{\infty} d\eta\, e^{\eta}\exp[ie^{\zeta+\eta}]k^{-1}f(e^{\eta}k^{-1}) \\ &= (2\pi)^{-1/2}e^{\zeta/2}\int_{-\infty}^{\infty} d\eta\,\exp[ie^{\zeta+\eta}]\exp\left[\frac{\eta}{2}\right] \\ &\quad\times(LU(\eta)f)(k). \end{aligned} \tag{2.9}$$

Decomposing

$$D = \int \lambda\, dE_\lambda, \tag{2.10}$$

and defining

$$|D| = \int |\lambda|\, dE_\lambda, \tag{2.11}$$

we note that we can write

$$f = \exp[-\alpha|D|]h, \quad \alpha\in(0,\pi/2), \quad h\in\mathscr{H}_+. \tag{2.12}$$

Now

$$\begin{aligned} g(\zeta) &= (2\pi)^{-1/2}e^{\zeta/2}L\int_{-\infty}^{+\infty} d\eta\,\exp[ie^{\zeta+\eta}]\exp\left[\left(\frac{1}{2}+iD\right)\eta\right]\exp[-\alpha|D|]h \\ &= (2\pi)^{-1/2}e^{\zeta/2}L\int_{-\infty}^{+\infty} d\eta\,\exp[ie^{\zeta+\eta}]\int_\lambda \exp\left[\left(\frac{1}{2}+i\lambda\right)\eta\right]\exp[-\alpha|\lambda|]dE_\lambda h \\ &= (2\pi)^{-1/2}e^{\zeta/2}L\int_\lambda\int_{-\infty}^{+\infty} d\eta\,\exp[ie^{\zeta+\eta}]\exp\left[\left(\frac{1}{2}+i\lambda\right)\eta\right]\exp[-\alpha|\lambda|]dE_\lambda\, h \\ &= (2\pi)^{-1/2}e^{\zeta/2}L\int_\lambda\int_{-\infty+i(\psi-\pi/2)}^{+\infty+i(\psi-\pi/2)} d\eta\,\exp[-e^{\eta}]\exp\left[\left(\frac{1}{2}+i\lambda\right)\left(\eta-\zeta+\frac{i\pi}{2}\right)\right]\cdot\exp[-\alpha|\lambda|]dE_\lambda\, h \\ &= (2\pi)^{-1/2}e^{\zeta/2}L\int_\lambda \Gamma\left(\frac{1}{2}+i\lambda\right)\exp\left[\left(\frac{1}{2}+i\lambda\right)\left(-\zeta+\frac{i\pi}{2}\right)\right]\exp[-\alpha|\lambda|]dE_\lambda\, h \\ &= (2\pi)^{-1/2}e^{i\pi/4}L\exp[(-i\zeta-\pi/2)D]\Gamma(\tfrac{1}{2}+iD)f. \end{aligned} \tag{2.13}$$

Since

$$|\Gamma(\tfrac{1}{2}+iu)|^2 = 2\pi[\exp[\pi u] + \exp[-\pi u]]^{-1}, \tag{2.14}$$

it follows that

$$\begin{aligned} K(\zeta) &= (2\pi)^{-1/2}e^{i\pi/4}L \\ &\quad\times\exp[(-i\zeta-\pi/2)D]\Gamma(\tfrac{1}{2}+iD) \end{aligned} \tag{2.15}$$

is a bounded operator on $\mathscr{H}_+$ for $\psi\in[0,\pi]$ with

$$\begin{aligned} \|K(\zeta)\| &= \|\exp[(\psi-\pi/2)D][\exp[\pi D] \\ &\quad + \exp[-\pi D]]^{1/2}\| \leqslant 1, \quad \psi\in[0,\pi]. \end{aligned} \tag{2.16}$$

The analyticity of $K(\zeta)$ in the strip $\psi\in(0,\pi)$ is evident from (2.15), whereas the existence of the strong limits for $\psi\downarrow 0$ and $\psi\uparrow\pi$ follows from the uniform estimate (2.16) and the fact that they exist for f as in the beginning of this proof. We thus find

$$K(\eta) = (2\pi)^{-1/2}e^{i\pi/4}L\exp[-i\eta D] \times \exp[-(\pi/2)D]\Gamma(\tfrac{1}{2}+iD), \quad (2.17)$$

$$K(\eta+i\pi) = (2\pi)^{-1/2}e^{i\pi/4}L\exp[-i\eta D] \times \exp[(\pi/2)D]\Gamma(\tfrac{1}{2}+iD), \quad (2.18)$$

which are evidently strongly continuous in η. [Note that $\exp[\pm(\pi/2)D]\Gamma(\tfrac{1}{2}+ID)$ are bounded operators.]

Let now $f\in\mathscr{H}_+$ and

$$g = K(i\pi/2)f = (2\pi)^{-1/2}e^{i\pi/4}L\Gamma(\tfrac{1}{2}+iD)f. \quad (2.19)$$

It follows from (2.14) that g is an analytic vector for D. Thus since g equals, up to a factor, the Laplace transform, we have the following corollary.

Corollary: The Laplace transform of $f\in\mathscr{H}_+$ is an analytic vector for D.

Since $\Gamma(\tfrac{1}{2}+iD)$ has an (unbounded) inverse and g is in its domain, we have the inversion formula

$$f = (2\pi)^{-1/2}e^{-i\pi/4}\Gamma(\tfrac{1}{2}+iD)^{-1}Lg, \quad (2.20)$$

which is an abstract version of a result by Paley and Wiener[1] [p. 39, (13.21)].

It follows from (1.2), (2.17), and (2.18) and the definition of F (1.4) that, for a function $f(x)\in\mathscr{H}$ vanishing for negative x,

$$(Ff)(k) = e^{i\pi/4}(L\exp[(\pi/2)D]\Delta(D)f)(k), \quad k>0,$$
$$(Ff)(k) = e^{-i\pi/4}(RL\exp[(\pi/2)D]\Delta(D)f)(k), \quad k<0, \quad (2.21)$$

where

$$\Delta(D) = (2\pi)^{-1/2}\Gamma(\tfrac{1}{2}+iD), \quad (2.22)$$

and R is the reflection operator $(Rf)(k) = f(-k)$. We can obtain similar expressions for $f(x)$ that vanishes for $x>0$ with the final result

$$F = L\{e^{i\pi/4}e^{-\pi/2D} + Re^{-i\pi/4}e^{\pi/2D}\}\Delta(D). \quad (2.23)$$

Since $L^*L = 1$ we have

$$F^*F = [\{e^{i\pi/4}e^{(\pi/2)D} + Re^{-i\pi/4}e^{(\pi/2)D}\}\Delta(D)]^* \cdot \{e^{i\pi/4}e^{-(\pi/2)D} + Re^{-i\pi/4}e^{(\pi/2)D}\}\cdot\Delta(D)$$
$$= (\exp[-\pi D] + \exp[\pi D])\Delta(D)^*\Delta(D) = 1. \quad (2.24)$$

It follows from (2.6) that, for $\Phi\in L^\infty(\mathbb{R},dx)$,

$$L\Phi(D) = \Phi(-D)L, \quad (2.25)$$

so that

$$F^2 = L(e^{i\pi/4}e^{-(\pi/2)D} + Re^{-i\pi/4}e^{(\pi/2)D})\Delta(D) \times L(e^{i\pi/4}e^{-(\pi/2)D} + Re^{-i\pi/4}e^{(\pi/2)D})\Delta(D)$$
$$= R\{\exp[\pi D] + \exp[-\pi D]\}\Delta(-D)\Delta(D) = R. \quad (2.26)$$

Thus the two major properties of F are verified.

Let us finally consider the map $M(\zeta)$ on $\mathscr{H}$, corresponding with $K(\zeta)$ on $\mathscr{H}_+$, as discussed at the beginning of this section. Thus let

$$\Psi(u) = \exp[u/2]g(e^u), \quad \phi(v) = \exp[v/2]f(e^v). \quad (2.27)$$

Then (2.9) translates into

$$\Psi(u) = (2\pi)^{-1/2}e^{\zeta/2}\int dv\exp[ie^{\zeta+v}]\exp\left[\frac{v}{2}\right]\phi(v-u)$$
$$= (2\pi)^{-1/2}e^{\zeta/2}\int dv\exp[ie^{\zeta+v}] \times\left(R\exp\left[\left(\frac{1}{2}+iP\right)v\right]\phi\right)(u)$$
$$= (M(\zeta)\phi)(u), \quad (2.28)$$

where R is again the reflection operator and $P = (-i\,\partial_u)^c$, the closure of $-i\,\partial_u$, is the generator of the translation group (the momentum operator). Following the same procedure as before, we obtain

$$M(\zeta) = (2\pi)^{-1/2}e^{i\pi/4}R \times\exp[(-i\zeta-\pi/2)P]\Gamma(\tfrac{1}{2}+iP). \quad (2.29)$$

III. THE L^p CASE, $p\neq 2$

It is known[4,5] that P, the generator of translations on $L^p(\mathbb{R},dx)$, $1<p<\infty$, $p\neq 2$, is not a spectral operator of scalar type. In the following we show this in an elementary way by deriving a contradiction. The corresponding statement for the generator of dilatations, D, follows immediately.

Let $f(x)\in L^p_+$, $1<p\leqslant 2$, and let

$$g(k,\zeta) = (2\pi)^{-1/2}e^{\zeta/q}\int_0^\infty dx\exp[ie^\zeta kx]f(x),$$
$$p^{-1}+q^{-1} = 1, \quad \operatorname{Im}\zeta\in(0,\pi). \quad (3.1)$$

For the set $\{f_n\}$ given by (2.7), $g(k,\zeta)\in L^q_+$, and, since this set is fundamental in L^p_+, (3.1) defines an operator $K(\zeta)$ with dense domain. The factor $\exp[\zeta/q]$ has been chosen in such a way that for $\zeta=\vartheta$, ϑ real, $g(k,\vartheta) = (U(\vartheta)g)(k)$. Because of the somewhat simpler notation, we change to the equivalent map from L^p to L^q by putting

$$\Psi(u) = \exp[u/q]g(e^u), \quad \phi(v) = \exp[v/p]f(e^v), \quad (3.2)$$

so that

$$\Psi(u) = (2\pi)^{-1/2}e^{\zeta/q}\int dv\exp[ie^{\zeta+v}]\exp\left[\frac{v}{q}\right]\phi(v-u)$$
$$= (2\pi)^{-1/2}e^{\zeta/q}\int dv\exp[ie^{\zeta+v}] \times\left(R\exp\left[\left(\frac{1}{q}+iP\right)v\right]\phi\right)(u)$$
$$= (M(\zeta)\phi)(u), \quad (3.3)$$

where P is now the generator of translations on L^p. Its spectrum is $\mathbb{R}$. This can be seen in an elementary way by noting that for $\operatorname{Im} z>0$,

$$f(x,z) = ([z-P]^{-1}f)(x)$$
$$= -i\int_0^\infty dt(\exp[i(z-P)t]f)(x)$$
$$= -i\int_0^\infty dt\exp[izt]f(x-t), \quad f\in L^p. \quad (3.4)$$

Assuming that $z_0\in\mathbb{R}$ is in the resolvent set of P, we can continue $f(z)$ analytically to $f(z_0)$. Taking $f(x) = \exp[-|x|]$ we find that $f(z_0)$ is not in L^p, so that z_0 must be in $\sigma(P)$.

Recalling that $L^q = (L^p)^*$, $1 \leqslant p < \infty$, we can write a general bounded linear functional on L^p

$$\Phi\chi(\phi) = \langle \phi,\chi \rangle = \int_{-\infty}^{+\infty} dv\, \phi(v)\chi(v),$$

$$\phi \in L^p, \quad \chi \in L^q, \quad p^{-1} + q^{-1} = 1. \tag{3.5}$$

Suppose now that P, defined above, is a spectral operator of scalar type. Then, since $\sigma(P) = \mathbb{R}$, there is a spectral measure $E(\cdot)$ defined on the Borel sets of $\mathbb{R}$ such that (for these notions see Dunford and Schwartz[2])

$$\langle P\phi,\chi \rangle = \int_{-\infty}^{+\infty} \lambda\, d\langle E(\lambda)\phi,\chi \rangle,$$

$$\phi \in \mathscr{D}(P) \subset L^p, \quad \chi \in L^q, \tag{3.6}$$

and also (Berkson,[6] Theorem 3.1)

$$\langle \exp[+iPt]\phi,\chi \rangle = \int_{-\infty}^{+\infty} \exp[-i\lambda t]\, d\langle E(\lambda)\phi,\chi \rangle,$$

$$\phi \in L^p, \quad \chi \in L^q. \tag{3.7}$$

Recalling that $\langle E(\cdot)\phi,\chi \rangle$ has finite total measure, we can use Fubini's theorem to obtain

$$\begin{aligned}
\langle M(\zeta)\phi,\chi \rangle
&= (2\pi)^{-1/2} e^{\zeta/q} \int dv \exp[ie^{\zeta+v}] \exp\left[\frac{v}{q}\right] \\
&\quad \times \langle R \exp[iPv]\phi,\chi \rangle \\
&= (2\pi)^{-1/2} e^{\zeta/q} \int dv \exp[ie^{\zeta+v}] \exp\left[\frac{v}{q}\right] \\
&\quad \cdot \int d\langle E(\lambda)\phi,R\chi \rangle \exp[i\lambda v] \\
&= (2\pi)^{-1/2} e^{\zeta/q} \int d\langle E(\lambda)\phi,R\chi \rangle \int dv \exp[ie^{\zeta+v}] \\
&\quad \times \exp\left[\left(\frac{1}{q} + i\lambda\right)v\right] \\
&= (2\pi)^{-1/2} \exp\left[\frac{i\pi}{2q}\right] \int d\langle E(\lambda)\phi,R\chi \rangle \\
&\quad \times \exp\left[\left(-i\zeta - \frac{\pi}{2}\right)\lambda\right] \Gamma\left(\frac{1}{q} + i\lambda\right).
\end{aligned} \tag{3.8}$$

Since the integrand is a bounded continuous function, analytic in a domain containing $\mathbb{R}$, the operational calculus applies with the result

$$\langle M(\zeta)\phi,\chi \rangle = (2\pi)^{-1/2} \exp\left[\frac{i\pi}{2q}\right] \times \left\langle R \exp\left[\left(-i\zeta - \frac{\pi}{2}\right)P\right] \Gamma\left(\frac{1}{q} + iP\right)\phi,\chi \right\rangle. \tag{3.9}$$

Thus

$$M(\zeta) = (2\pi)^{-1/2} \exp\left[\frac{i\pi}{2q}\right] R \times \exp\left[\left(-i\zeta - \frac{\pi}{2}\right)P\right] \Gamma\left(\frac{1}{q} + iP\right) \tag{3.10}$$

defines a bounded operator. Next we note that [see Ref. 7, p. 47 (6)]

$$\lim_{|y|\to\infty} |\Gamma(x+iy)| \exp[\pi/2|y|]\, |y|^{1/2-x} = (2\pi)^{1/2}, \tag{3.11}$$

which implies that for $1 \leqslant p < 2$, i.e., $0 \leqslant q^{-1} \leqslant \frac{1}{2}$, $M(\zeta)$ is a bounded linear operator on L^p, analytic in ζ in a strip $0 < \operatorname{Im} \zeta < \pi$ with limits continuous in η as in $\zeta = \eta + i\psi$, $\psi \downarrow 0$ and $\psi \uparrow \pi/2$, respectively. But the latter are precisely the Fourier transform, i.e., a bounded 1–1 map from L^p onto L^q, for $1 < p < 2$, and a map from L^1 to a subset of L^∞ containing functions outside L^1, for $p = 1$. Thus the assumption that P is spectral of scalar type is incorrect. Since L^p, $1 < p < \infty$, is reflexive, a duality argument implies the same for P acting in L^p, $2 < p < \infty$. Thus we have the following.

Proposition 3: The generator P of the translation group and, *a fortiori*, the generator D of the dilatation group, are not spectral operators of scalar type on L^p, $1 \leqslant p < \infty$, $p \neq 2$.

IV. AN APPLICATION

In the following we discuss an aspect of a model for the following situation: Consider an atom moving in front of a metal surface. Assuming the existence of the corresponding negative ion, then, depending upon its bound state level and the work function of the metal, the atom may pick up an electron from the metal and emerge as a negative ion, or a negative ion may be stripped of an electron. In the case of an atom frozen close to the metal, the negative ion bound state may now become unstable; it turns into a resonance. A very simple model for such a situation is a one-dimensional one (the Sommerfeld model) where the metal is represented by a step function potential and the atom by an attractive potential with support outside the metal. The corresponding Hamiltonian is then the self-adjoint operator on $\mathscr{H} = L^2(\mathbb{R},dx)$ given by [p^2 is the closure of $-\partial_x{}^2$, $\theta(x) = 1$, $x \geqslant 0$, $\theta(x) = 0$, $x < 0$]

$$H = p^2 - \mu\theta(x) + V(x) = H_0 + V(x), \tag{4.1}$$

where $V(x) \leqslant 0$ is a continuous function of x with support in an interval $[a,b] \subset \mathbb{R}^-$, bounded away from zero. Here H_0 has a purely continuous spectrum consisting of two branches, $[-\mu,\infty)$ and $[0,\infty)$, as can be seen by means of Dirichlet decoupling (adding a Dirichlet condition in $x = 0$) and the use of Krein's formula.[8] Suppose now that $H_1 = p^2 + V(x)$ has a negative eigenvalue ν with $-\mu < \nu < 0$. Then this eigenvalue is unstable under the perturbation $-\mu\theta$.[9] In fact we found for a model problem[10] with $V(x) = -\lambda\,\delta(x+y)$, $0 < \lambda^2/4 < \mu$, $y > 0$ (in which case $\nu = -\lambda^2/4$), that ν turns into a resonance for sufficiently large y but becomes real in the second Riemann sheet for small y (a so-called virtual state). The same generic behavior was found in case the metal was described by a Kronig–Penney model.

A standard procedure to treat resonances is the dilatation-analytic method initiated by Aguilar and Combes.[11] It cannot be used in the present case due to the discontinuity in $\theta(x)$. An adapted version akin to the exterior scaling version of Simon[12] does work, however. Thus we define, for $\vartheta \in \mathbb{R}$,

$$(U_+(\vartheta)f)(x) = \begin{cases} e^{\vartheta/2} f(e^{\vartheta}x), & x \geqslant 0, \\ f(x), & x < 0. \end{cases} \tag{4.2}$$

We note that the $\mathscr{H}_\pm$ reduce $U_+(\vartheta)$ and that

$\{U_+(\vartheta)|\vartheta\in\mathbb{R}\}$ constitutes a strongly continuous group of unitary operators. We have

$$(U_+(\vartheta)Vf)(x) = V(x)(U_+(\vartheta)f)(x),$$
$$(U_+(\vartheta)\theta f)(x) = \theta(e^{\vartheta}x)(U_+(\vartheta)f)(x)$$
$$= \theta(x)(U_+(\vartheta)f)(x),$$

i.e.,

$$V(\vartheta)\equiv U_+(\vartheta)VU_+(\vartheta)^{-1} = V,\qquad \theta(\vartheta)\equiv U_+(\vartheta)\theta U_+(\vartheta)^{-1} = \theta. \tag{4.3}$$

$\mathscr{D}(p^2)$ is not invariant under $U_+(\vartheta)$ [for $f\in\mathscr{D}(p^2)$ with $f(0)\neq 0$, $(U_+(\vartheta)f)(x)$ has a discontinuity in zero for $\vartheta\neq 0$], but we can define $p^2(\vartheta)$ through the resolvent

$$[z-p^2(\vartheta)]^{-1} = U_+(\vartheta)[z-p^2]^{-1}U_+(\vartheta)^{-1}. \tag{4.4}$$

Let $p^2_\pm$ be the closure of $-\partial_x^2$ on $\mathscr{H}_\pm$ with the Dirichlet boundary condition in the origin. Then by Krein's formula ($P_\pm$ are the projectors upon $\mathscr{H}_\pm$),

$$[z-p^2]^{-1} = [z-p^2_+]^{-1}P_+ + [z-p^2_-]^{-1}P_- + (2i\kappa)^{-1}(\ ,\phi(-i\bar\kappa))\phi(i\kappa), \tag{4.5}$$

where $\kappa=\sqrt{z}$ and

$$\phi(x,i\kappa) = \exp[i\kappa|x|]. \tag{4.6}$$

It is easily established that

$$U_+(\vartheta)[z-p^2_+]^{-1}U_+(\vartheta)^{-1} = [z-p^2_+e^{-2\vartheta}]^{-1}$$

and

$$U_+(\vartheta)[z-p^2_-]^{-1}U_+(\vartheta)^{-1} = [z-p^2_-]^{-1}$$

so that

$$[z-p^2(\vartheta)]^{-1} = [z-p^2_+e^{-2\vartheta}]^{-1}\oplus[z-p^2_-]^{-1} + (2i\kappa)^{-1}(\ ,\phi(-i\bar\kappa,\vartheta))\phi(i\kappa,\vartheta), \tag{4.7}$$

where

$$\phi(i\kappa,\vartheta) = U_+(\vartheta)\phi(i\kappa). \tag{4.8}$$

Equation (4.7) can be continued to complex ϑ, i.e., $\vartheta\rightarrow\zeta=\vartheta+i\psi$. For suitably chosen z (negative real, for example),

$$[z-p^2(\zeta)]^{-1} = [z-p^2_+e^{-2\zeta}]^{-1}P_+\ [z-p^2_-]^{-1}P_- + (2i\kappa)^{-1}(\ ,\phi(-i\bar\kappa,\bar\zeta))\phi(i\kappa,\zeta), \tag{4.9}$$

where

$$\phi(i\kappa,\zeta) = \begin{cases} e^{\zeta/2}\exp[i\kappa e^{\zeta}x], & x\geqslant 0,\\ \exp[-i\kappa x], & x<0.\end{cases} \tag{4.10}$$

Equation (4.9) defines a closed operator $p^2(\zeta)$ that converges towards $p^2(\vartheta)$ in strong resolvent sense as $\psi\downarrow 0$. It follows that these properties carry over to

$$H(\zeta) = p^2(\zeta) - \mu\theta(x) + V(x), \tag{4.11}$$

where $H(\zeta)$, which is no longer self-adjoint for nonreal ζ, may possess complex eigenvalues (the resonances mentioned earlier in this section).

In the course of our investigations concerning the present model, connected with the derivation of an adiabatic theorem, we had to determine the analyticity properties of the time evolution generated by $H(\zeta,t)$ given by (4.11) but with V time dependent. Since $-\mu\theta+V(t)$ is a bounded perturbation of $p^2(\zeta)$, this can be handled once we know the properties of

$$W(t,\zeta) = \exp[-ip^2(\zeta)t] \tag{4.12}$$

for the cases $t>0$, $\psi\in(0,\pi/2)$, and $t<0$, $\psi\in(-\pi/2,0)$, respectively. In the standard dilatation case we have $W(t,\zeta)=\exp[-ip^2e^{-2\zeta}t]$, which, for $t>0$, is analytic in ζ, $\psi\in(0,\pi/2)$, due to the sectoriality properties of $p^2e^{-2\zeta}$ (see Kato,[3] Chap. IX). In the present case, $p^2(\zeta)$ does not possess such properties. In fact it is not even dissipative. This difficulty can be circumvented by starting from $\exp[-ip^2t]$, $t>0$ directly. We have (see Kato,[3] Chap. IX, § 1–8)

$$g(x,t) = (\exp[-ip^2t]f)(x) = (4\pi it)^{-1/2}\int dx'\exp\left[\frac{i}{4t}(x-x')^2\right]f(x'),\quad t>0,\ f\in\mathscr{H}, \tag{4.13}$$

or, with $s=\ln\sqrt{2t}$,

$$g(x,t) = \exp\left[-\frac{i\pi}{4}\right]\exp\left[-\frac{s}{2}\right]\exp\left[\frac{i}{2}(e^{-s}x)^2\right]\cdot\frac{1}{\sqrt{2\pi}}\int\exp[-ie^{-s}xx']\exp\left[\frac{i}{2}(x')^2\right]\times\exp\left[\frac{s}{2}\right]f(e^sx'), \tag{4.14}$$

or

$$g(t) = \exp\left[-\frac{i\pi}{4}\right]U(-s)\exp\left[\frac{i}{2}x^2\right]F^{-1}\times\exp\left[\frac{i}{2}x^2\right]U(s)f$$
$$= \exp\left[\frac{-i\pi}{4}\right]U(-\ln\sqrt{2t})\exp\left[\frac{i}{2}x^2\right]F^{-1}\times\exp\left[\frac{i}{2}x^2\right]U(\ln\sqrt{2t})f,\quad t>0, \tag{4.15}$$

where $\exp[(i/2)x^2]$ is interpreted as a multiplication operator. Thus

$$\exp[-ip^2(\vartheta)t] = U_+(\vartheta)\exp[-ip^2t]U_+(\vartheta)^{-1}$$
$$= \exp\left[-\frac{i\pi}{4}\right]U_+(\vartheta)U(-\ln\sqrt{2t})\exp\left[\frac{i}{2}x^2\right]F^{-1}\times\exp\left[\frac{i}{2}x^2\right]U(\ln\sqrt{2t})U_+(\vartheta)^{-1}$$
$$= \exp\left[\frac{-i\pi}{4}\right]U(-\ln\sqrt{2t})\exp\left[\frac{i}{2}x^2(\vartheta)\right]\times U_+(\vartheta)F^{-1}U_+(\vartheta)^{-1}\exp\left[\frac{i}{2}x^2(\vartheta)\right]\cdot U(\ln\sqrt{2t}),\quad t>0, \tag{4.16}$$

where

$$\exp[(i/2)x^2(\vartheta)] = U_+(\vartheta)\exp[(i/2)x^2]U_+(\vartheta)^{-1} \tag{4.17}$$

is reduced by $\mathscr{H}_\pm$ and we have

$$\exp\left[\frac{i}{2}x^2(\vartheta)\right] = \exp\left[\frac{i}{2}e^{2\vartheta}x^2\right]P_+ + \exp\left[\frac{i}{2}x^2\right]P_-. \tag{4.18}$$

Its right-hand side can be continued analytically into the strip $S = \{\zeta\in\mathbb{C}, 0 < \operatorname{Im}\zeta < \pi/2\}$. Thus

$$\exp\left[\frac{i}{2}x^2(\zeta)\right] = \exp\left[\frac{i}{2}e^{2\zeta}x^2\right]P_+ + \exp\left[\frac{i}{2}x^2\right]P_-, \quad \zeta\in S, \tag{4.19}$$

with (4.18) as its strong limit for $\operatorname{Im}\zeta\downarrow 0$. We also have

$$P_+\exp[-ip^2(\vartheta)t]P_+ + P_-\exp[-ip^2(\vartheta)t]P_- = P_+\exp[-ip^2e^{-2\vartheta}t]P_+ + P_-\exp[-ip^2t]P_-, \tag{4.20}$$

which can again be continued into S and is the strong limit of its analytic continuation as $\operatorname{Im}\zeta\downarrow 0$. It remains to consider $P_\pm \exp[-ip^2(\vartheta)t]P_\mp$. We have ($x\leqslant 0$)

$$\begin{aligned}(P_-FU_+(\vartheta)P_+f)(x) &= (2\pi)^{-1/2}\int_0^\infty dx' \exp[ixx']e^{-\vartheta/2}f(e^{-\vartheta}x') \\ &= (2\pi)^{-1/2}e^{-\vartheta/2}\int_0^\infty dx' \exp[-ie^{\vartheta}xx']f(x') \\ &= (K(\vartheta)f)(-x).\end{aligned} \tag{4.21}$$

According to our findings in Sec. II, (4.21) can also be continued into S with (4.21) the strong limit of this continuation, and in view of (4.19) the same result holds for $P_-\exp[-ip^2(\vartheta)t]P_+$. The other case goes similarly and we arrive at the following.

Theorem: $\exp[-ip^2(\vartheta)t]$, $t\geqslant 0$, has an analytic extension $\exp[-ip^2(\zeta)t]$ in the strip $S = \{\zeta\in\mathbb{C}, 0 < \operatorname{Im}\zeta < \pi/2\}$ and is the strong limit of this extension as $\operatorname{Im}\zeta\downarrow 0$. $\{\exp[-ip^2(\zeta)t] \mid t\geqslant 0\}$ is a bounded, strongly continuous semigroup.

Proof: It remains to verify the semigroup properties. The extension of (4.20) is obviously strongly continuous in $t\geqslant 0$, equals the identity operator for $t=0$, and is bounded in norm by 1. The remaining terms are strongly continuous in $t>0$ and, since $\|K(\zeta)\|\leqslant 1$, each is bounded in norm by 1 for all $t>0$. Moreover their strong limits as $t\downarrow 0$ equal zero. In order to see this we note that, in view of the boundedness property above, we only have to show this for a dense set of $f\in\mathscr{H}$. For the latter we take the analytic vectors for D. Now, for instance,

$$\begin{aligned}&P_+U_+(\vartheta)\exp[-ip^2t]U_+(\vartheta)^{-1}P_-f \\ &\quad= P_+\exp[-ip^2e^{-2\vartheta}t]P_-U(\vartheta)f \\ &\quad\xrightarrow{\vartheta\to\zeta} P_+\exp[-ip^2e^{-2\zeta}t]P_-f(\zeta) \xrightarrow{t\downarrow 0} P_+P_-f(\zeta) = 0.\end{aligned} \tag{4.22}$$

Thus $\exp[-ip^2(\zeta)t]$ is strongly continuous in $t\geqslant 0$, approaches 1 as $t\downarrow 0$, and is bounded in norm (by 3). The property

$$\exp[-ip^2(\zeta)(t_1+t_2)] = \exp[-ip^2(\zeta)t_1]\cdot\exp[-ip^2(\zeta)t_2]$$

finally follows from the corresponding property for $\zeta=\vartheta$ real, and the uniqueness of the analytic extension.

Corollary: Let $V(x)$ in (4.1) be time dependent, $V(x) = V(x,t)$ with $|V(x,t)|\leqslant\kappa$, $\kappa>0$, and be strongly continuous in t in the interval $[0,T]$. Then $H = H(t)$ generates a unitary time-evolution operator $U(t,s)$, $T\geqslant t\geqslant s\geqslant 0$, with the usual properties, which has the additional property that $U(t,s,\vartheta) = U(\vartheta)U(t,s)U(\vartheta)^{-1}$ can be continued analytically: $\vartheta\to\zeta$, $\zeta\in S$.

The straightforward proof follows by noting that each term in the Dyson expansion for $U(t,s,\zeta)$ is analytic and that the Dyson series converges uniformly on $[0,T]$. (For the Dyson expansion see Reed and Simon,[13] Sec. X.12, for the self-adjoint case, and Kato,[14] Theorem 4.5, for the general situation.)

ACKNOWLEDGMENTS

A useful conversation with Dr. N. Walet is gratefully acknowledged.

This work is part of the research program of the Stichting voor Fundamenteel Onderzoek der Materie (Foundation for Fundamental Research on Matter) and was made possible by financial support from the Nederlandse Organisatie voor Zuiver-Wetenschappelijk Onderzoek (Netherlands Organization for the Advancement of Pure Research).

[1]R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain* (American Mathematical Society, New York, 1934).

[2]N. Dunford and J. T. Schwartz, *Linear Operators III* (Wiley–Interscience, New York, 1971).

[3]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1984), second corrected printing of the second edition.

[4]T. A. Gillespie, J. London Math. Soc. (2) **11**, 499 (1975).

[5]G. I. Gaudry and W. Ricker, J. Operator Theory **14**, 87 (1985).

[6]E. Berkson, Illinois J. Math. **10**, 345 (1966).

[7]*Higher Transcendental Functions I*, edited by A. Erdélyi (McGraw-Hill, New York, 1953).

[8]N. I. Archieser and I. M. Glasmann, *Theorie der linearen Operatoren im Hilbert-raum* (Akademie, Berlin, 1960).

[9]A. Tip, submitted to J. Math. Phys.

[10]P. Hofstee, H. G. Muller, and A. Tip, submitted to J. Phys. A.

[11]J. Aguilar and J. M. Combes, Commun. Math. Phys. **22**, 269 (1971).

[12]B. Simon, Phys. Lett. A **71**, 211 (1979).

[13]M. Reed and B. Simon, *Methods of Modern Mathematical Physics II, Fourier Analysis, Self-adjointness* (Academic, New York, 1975).

[14]T. Kato, J. Fac. Sci. Imp. Univ. Tokyo Sect. 1 **17**, 241 (1970).

Totally vicious space-times and reflectivity

Yoshihiro Matori
Department of Mathematics, Faculty of Science, Kyushu University 33, Fukuoka 812, Japan

(Received 19 September 1987; accepted for publication 21 October 1987)

A reflecting space-time with a closed timelike curve is proved to be totally vicious. In contrast with this, it is shown that there exists a compact space-time that is not totally vicious.

I. INTRODUCTION

A space-time (M,g) is called totally vicious if $I^+(p)\cap I^-(p) = M$ for every $p\in M$. Tipler[1] has shown that a totally vicious space-time is causally simple, but is not chronological. He has also proved that total viciousness is a stable property of a space-time. Furthermore, Beem and Ehrich[2] have characterized total viciousness by a Lorentzian distance function.

In this paper, we shall investigate the conditions that imply total viciousness. In Sec. II, these conditions, each of which is equivalent to total viciousness, are given. In Sec. III, we prove that a reflecting space-time with a closed timelike curve is totally vicious. This is an improvement of Proposition 4 in Tipler.[1] In Sec. IV, we present the example of a compact space-time that is not totally vicious.

For notations and conventions in this paper, we refer to Beem and Ehrich.[2] In particular, by a space-time (M,g) we mean a connected time-oriented C^∞ Lorentzian manifold of dimension $n+1$ with C^∞ Lorentzian metric g. For $p\in M$, the chronological future $I^+(p)$ [resp. past $I^-(p)$] of p is the set of all points that can be reached from p by a future- (resp. past-) directed timelike curve in M.

II. TOTALLY VICIOUS SPACE-TIMES

Definition 2.1: A space-time (M,g) is called totally vicious if $I^+(p)\cap I^-(p) = M$ for every $p\in M$, where $I^+(p)$ [resp. $I^-(p)$] is the chronological future (resp. past) of p.

We now give the conditions, each of which is equivalent to total viciousness.

Proposition 2.2: The following conditions (A1)–(A4), (B1), and (B2) are equivalent.

(A1) (M,g) is totally vicious.

(A2) For some $p_0\in M$, $I^+(p_0)\cap I^-(p_0) = M$.

(A3) For every $p\in M$, there exists a closed timelike curve through p.

(A4) For every $p,q\in M$ $(p\neq q)$, there exists a timelike curve between p and q.

(B1) For every $p\in M$, $J^+(p)\cap J^-(p) = M$.

(B2) For some $p_0\in M$, $J^+(p_0)\cap J^-(p_0) = M$.

Here $J^+(p)$ [resp. $J^-(p)$] is the causal future (resp. past) of p.

The equivalence of (A1), (A2), and (A3) has been pointed out by Tipler.[1] But it seems that (A4), (B1), and (B2) have not been found in the literature.

Proof: (A1)→(A2). This is obvious.

(A2)→(A1). For every $p,q\in M$, it follows from condition (A2) that $p,q\in I^+(p_0)\cap I^-(p_0)$. In particular, $q\in I^+(p_0)$ and $p\in I^-(p_0)$, so $q\in I^+(p)$. A similar argument shows $q\in I^-(p)$. Hence $q\in I^+(p)\cap i^-(p)$ for every $p\in M$. Since q is arbitrary, $I^+(p)\cap I^-(p) = M$. This is condition (A1).

(A1)→(A3). From condition (A1), $p\in I^+(p)$ for every $p\in M$. This yields condition (A3).

(A3)→(A2). Condition (A3) implies $p\in I^+(p)\cap I^-(p)$ for every $p\in M$; then

$$M = \bigcup_{p\in M} [I^+(p)\cap I^-(p)].$$

From Proposition 6.4.1 of Hawking and Ellis,[3] it follows that this is the disjoint union of the sets $I^+(p)\cap I^-(p)$. Since M is connected and $I^+(p)\cap I^-(p)$ is open for every $p\in M$, there exists a point $p_0\in M$ such that

$$M = I^+(p_0)\cap I^-(p_0).$$

This is condition (A2).

(A1)→(A4). From condition (A1), $q\in I^+(p)$ for every $p,q\in M$. This yields condition (A4).

(A4)→(A3). For every $p\in M$, it follows from condition (A4) that

$$I^+(p)\cup I^-(p) = M \text{ or } M-(p).$$

Here $I^+(p)$ and $I^-(p)$ are open, so $I^+(p)\cap I^-(p)\neq\phi$ since M and $M-(p)$ are connected. If we let $r\in I^+(p)\cap I^-(p)$, then $r\in I^+(p)$ and $p\in I^+(r)$, so $p\in I^+(p)$. This yields condition (A3).

Therefore the equivalence of (A1)–(A4) has been shown. The equivalence of (B1) and (B2) can be proved in a way similar to proving the equivalence of (A1) and (A2). Hence there remains only to show that (A1) and (B1) are equivalent.

(A1)→(B1). This is immediate from the fact that $I^+(p)\subset J^+(p)$ and $I^-(p)\subset J^-(p)$ for every $p\in M$.

(B1)→(A1). Let $p\in M$ and $p'\in I^+(p)$. For every $q\in M$, it follows from condition (B1) that $q\in J^+(p')$. Thus $q\in I^+(p)$. A similar argument shows $q\in I^-(p)$. Since q is arbitrary, $I^+(p)\cap I^-(p) = M$ for every $p\in M$. This is condition (A1).

Remark: In general, the following conditions (B3) and (B4) do not imply total viciousness.

(B3) For every $p\in M$, there exists a closed nonspacelike curve through p.

(B4) For every $p,q\in M$ $(p\neq q)$, there exists a nonspacelike curve between p and q.

For example, let (M,g) be a space-time such that $M = S^1\times\mathbb{R}$ and $g = -dt\,dx$, $(x,t)\in S^1\times\mathbb{R}$, where S^1 is the one-dimensional sphere. Then (M,g) satisfies conditions (B3) and (B4), but is not totally vicious.

 0022-2488/88/040823-02$02.50

III. TOTAL VICIOUSNESS OF A REFLECTING SPACE-TIME

First we shall describe the definition and some properties of a reflecting space-time which were established by Hawking and Sachs.[4]

Definition 3.1: A space-time (M,g) is called reflecting if (M,g) satisfies the following condition: for every $p,q \in M$,

$$I^+(p) \supset I^+(q) \text{ iff } I^-(p) \subset I^-(q).$$

Proposition 3.2: A space-time (M,g) is reflecting iff (M,g) satisfies the following condition: for every $p,q \in M$,

$$q \in \text{Closure}[I^+(p)] \text{ iff } p \in \text{Closure}[I^-(q)].$$

Proof: This proposition derives from the fact that

$$\text{Closure}[I^+(p)] = \{r \in M\colon I^+(p) \supset I^+(r)\}$$

and

$$\text{Closure}[I^-(q)] = \{s \in M\colon I^-(s) \subset I^-(q)\}$$

for every $p,q \in M$.

The following corollary is straightforward from Proposition 3.2.

Corollary 3.3: A causally simple space-time (M,g) is reflecting. [A space-time (M,g) is called causally simple if $J^+(p)$ and $J^-(p)$ are closed in M for every $p \in M$.]

Tipler[1] has proved that a causally simple space-time that has a closed timelike curve is totally vicious. In the following theorem we improve upon this.

Theorem 3.4: A reflecting space-time (M,g) that has a closed timelike curve is totally vicious.

Proof: Let $\gamma\colon \mathbb{R} \to M$ be a closed timelike curve such that

$$\gamma(u) = \gamma(u+1),$$

$$\gamma(v) \in I^+(\gamma(u)) \quad \text{if } v > u \quad (u,v \in \mathbb{R}).$$

For $p_0 = \gamma(0)$, we shall show that $I^+(p_0) = \text{Closure}[I^+(p_0)]$. If we let $r \in \text{Closure}[I^+(p_0)]$, then it follows from Proposition 3.2 that $p_0 \in \text{Closure}[I^-(r)]$, so $I^-(p_0) \subset I^-(r)$. In particular, $\gamma(-\varepsilon) \in I^-(r)$, i.e., $r \in I^+(\gamma(-\varepsilon))$ for every $0 < \varepsilon < 1$. Since $\gamma(-\varepsilon) = \gamma(1-\varepsilon)$,

$$r \in I^+(\gamma(1-\varepsilon)) \subset I^+(p_0).$$

Hence $I^+(p_0) = \text{Closure}[I^+(p_0)]$.

Now $I^+(p_0)$ is always open; this means that $I^+(p_0)$ is open and closed. Since M is connected, $I^+(p_0) = M$. A similar argument shows $I^-(p_0) = M$. Hence it has been proved that (M,g) satisfies condition (A2) of Proposition 2.2, i.e., (M,g) is totally vicious.

For example, let (M,g) be a space-time such that

$$M = \mathbb{R} \times S^1 - \{\text{finite points}\},$$

$$g = -dt^2 + dx^2 \quad [(x,t) \in \mathbb{R} \times S^1].$$

Here S^1 is the one-dimensional sphere. Obviously, (M,g) has a closed timelike curve. But it does not seem certain that (M,g) is causally simple. However, it is easily seen that (M,g) is reflecting (cf. Fig. 1.2 of Hawking and Sachs[4]). Hence it follows from Theorem 3.4 that (M,g) is totally vicious.

IV. EXAMPLE OF A COMPACT SPACE-TIME THAT IS NOT TOTALLY VICIOUS

It follows from Proposition 2.2 that a totally vicious space-time has a closed timelike curve. On the other hand, it is well-known that a compact space-time has a closed timelike curve.[2,3] But a compact space-time is not necessarily totally vicious. We shall show that the compact space-time constructed by Galloway[5] is not totally vicious.

Example 4.1: Let (N,g) be a space-time such that

$$N = \mathbb{R}^2,$$

$$g = \cos^2 x(-dt^2 + dx^2) + 2 \sin x \, dt \, dx \quad [(x,t) \in \mathbb{R}^2].$$

The metric g is invariant under the transformations $(x,t) \to (x + 2\pi, t)$ and $(x,t) \to (x, t+1)$. Let G be the group of isometries generated by these transformations. The quotient manifold $M = N/G$ is a compact space-time that is diffeomorphic to the two-dimensional torus. In this space-time (M,g), the lines $x = \pm \pi/2$ are closed null geodesics. For $p = (0,0) \in M$,

$$I^+(p) = \{(x,t) \in \mathbb{R}^2\colon -\pi/2 < x < \pi/2\}/G$$

(cf. Fig. 1 of Galloway[5]). Hence (M,g) is not totally vicious. On the other hand, $I^-(p) = M$.

It follows from Theorem 3.4 that this space-time (M,g) is not reflecting. In fact, for $p = (0,0)$ and $q = (\pi/2, 0)$, $I^+(p) \supset I^+(q)$, but $I^-(p) \not\subset I^-(q)$.

[1] F. J. Tipler, J. Math. Phys. **18**, 1568 (1977).
[2] J. K. Beem and P. E. Ehrich, *Global Lorentzian Geometry* (Dekker, New York, 1981).
[3] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U. P., Cambridge, 1973).
[4] S. W. Hawking and R. K. Sachs, Commun. Math. Phys. **35**, 287 (1974).
[5] G. J. Galloway, Proc. Am. Math. Soc. **98**, 119 (1986).

An embedding problem

István Ozsváth
Programs in Mathematical Sciences, The University of Texas at Dallas, P.O. Box 830688, Richardson, Texas 75083-0688

(Received 8 October 1987; accepted for publication 9 December 1987)

This paper contains two results. First it is shown that the three-dimensional Riemannian space, which is invariant under the transformations of the rotation group, cannot be embedded in a four-dimensional Euclidean space (except, of course, for the three-dimensional sphere). Second, the one parametric family of three-spaces with the above symmetry, which can be embedded in a four-dimensional unit sphere, is found and the embedding is constructed.

I. INTRODUCTION

In a previous paper[1] we developed a formalism in order to embed in a six-dimensional Euclidean space the three-dimensional Riemannian space invariant under the transformations of the rotation group. Here we discuss two special questions emerging from that paper.

More precisely, we consider the metric

$$(ds)^2 = p^2(\omega^1)^2 + q^2(\omega^2)^2 + r^2(\omega^3)^2, \tag{1.1}$$

where

$$p>0, \quad q>0, \quad r>0 \tag{1.2}$$

are arbitrary parameters and the one-forms

$$\omega^1, \quad \omega^2, \quad \omega^3 \tag{1.3}$$

satisfy the relations

$$\begin{aligned} d\omega^1 &= -\omega^2\wedge\omega^3, \quad d\omega^2 = -\omega^3\wedge\omega^1, \\ d\omega^3 &= -\omega^1\wedge\omega^2. \end{aligned} \tag{1.4}$$

We make the following two statements.

Theorem 1: The metric (1.1) can be embedded in the four-dimensional Euclidean space E^4 if and only if

$$p=q=r, \tag{1.5}$$

that is, if and only if (1.1) describes a three-sphere. [For $p=q=r=\frac{1}{2}$ (1.1) is the metric of S^3, the three-dimensional sphere with unit radius.]

Theorem 2: The metric (1.1) can be embedded in a four-dimensional sphere if and only if the parameters (1.2) satisfy the conditions

$$r=p+q \quad \text{and} \quad \kappa^2(p^2+pq+q^2)=3, \tag{1.6}$$

where κ^2 is the curvature of the ambient sphere. We prove Theorems 1 and 2 in Secs. II–IV and construct explicitly the embedding for (1.1) under condition (1.6).

Before that, however, we introduce for later use the vector fields

$$X_1, \quad X_2, \quad X_3 \tag{1.7}$$

defined by

$$\omega^a(X_b) = \delta^a_b, \quad a,b=1,2,3, \tag{1.8}$$

which then satisfy the commutation relations

$$[X_2,X_3]=X_1, \quad [X_3,X_1]=X_2, \quad [X_1,X_2]=X_3. \tag{1.9}$$

Furthermore, we choose the Eulerian angles

$$0\leqslant x^1 = x\leqslant\pi, \quad 0\leqslant x^2=y, \quad x^3=z\leqslant 2\pi \tag{1.10}$$

as our local coordinates. As explained in Ref. 1 the one-forms (1.3) and the vector fields (1.7) are then given by

$$\begin{aligned} \omega^1 &= \cos z\, dx + \sin x \sin z\, dy, \\ \omega^2 &= -\sin z\, dx + \sin x \cos z\, dy, \\ \omega^3 &= \cos x\, dy + dz, \end{aligned} \tag{1.11}$$

and

$$\begin{aligned} X_1 &= \cos z \frac{\partial}{\partial x} + \frac{\sin z}{\sin x}\frac{\partial}{\partial y} - \operatorname{ctg} x \sin z \frac{\partial}{\partial z}, \\ X_2 &= -\sin z \frac{\partial}{\partial x} + \frac{\cos z}{\sin x}\frac{\partial}{\partial y} - \operatorname{ctg} x \cos z \frac{\partial}{\partial z}, \\ X_3 &= \frac{\partial}{\partial z}, \end{aligned} \tag{1.12}$$

respectively.

II. PROOF OF THEOREM 1

As in Ref. 1 we use the one-forms (1.3) and the vector fields (1.7) as frames to span the tensor algebra of tensor fields, describing them by means of their frame components, and use the Koszul connection

$$\nabla_{X_a} X_b = \Gamma_{ab}{}^f X_f \tag{2.1}$$

to carry out covariant differentiation. As explained in Ref. 1 the components of the Koszul connection are given by

$$\begin{aligned} \Gamma_{231} &= \tfrac{1}{2}(p^2-q^2+r^2), \\ \Gamma_{321} &= -\tfrac{1}{2}(p^2+q^2-r^2), \\ \Gamma_{312} &= \tfrac{1}{2}(p^2+q^2-r^2), \\ \Gamma_{132} &= -\tfrac{1}{2}(-p^2+q^2+r^2), \\ \Gamma_{123} &= \tfrac{1}{2}(-p^2+q^2+r^2), \\ \Gamma_{213} &= -\tfrac{1}{2}(p^2-q^2+r^2), \end{aligned} \tag{2.2}$$

that is

 0022-2488/88/040825-10$02.50

$$
\begin{aligned}
\Gamma^1_{23} &= (1/2p^2)(p^2-q^2+r^2),\\
\Gamma^1_{32} &= -(1/2p^2)(p^2+q^2-r^2),\\
\Gamma^2_{31} &= (1/2q^2)(p^2+q^2-r^2),\\
\Gamma^2_{13} &= -(1/2q^2)(-p^2+q^2+r^2),\\
\Gamma^3_{12} &= (1/2r^2)(-p^2+q^2+r^2),\\
\Gamma^3_{21} &= -(1/2r^2)(p^2-q^2+r^2),
\end{aligned}
\tag{2.3}
$$

where the lowering and the raising of the frame indices are carried out with the help of the frame components of the metric

$$g_{ab} = \mathrm{diag}(p^2 \quad q^2 \quad r^2) \tag{2.4}$$

and

$$g^{ab} = \mathrm{diag}\begin{pmatrix}\frac{1}{p^2} & \frac{1}{q^2} & \frac{1}{r^2}\end{pmatrix}, \tag{2.5}$$

respectively.

In order to prove Theorem 1, we show that the Gauss equations

$$R_{abcd} = b_{ac}b_{bd} - b_{ad}b_{bc} \tag{2.6}$$

and the Codazzi–Mainardi equations

$$X_c b_{ab} - X_b b_{ac} + b_{af}C^f{}_{bc} - b_{fb}\Gamma_{ca}{}^f + b_{fc}\Gamma_{ba}{}^f = 0 \tag{2.7}$$

have a solution if and only if (1.5) holds, that is, if and only if $p=q=r$. Here R_{abcd} are the frame components of the Riemann tensor of (1.1) and b_{ab} are the components of the second fundamental form.

We denote a vector field's action on a function as Xf, for example,

$$X_1 f = \cos z\frac{\partial f}{\partial x} + \frac{\sin z}{\sin x}\frac{\partial f}{\partial y} - \cot x \sin z\frac{\partial f}{\partial z}.$$

A straightforward calculation shows that the nonvanishing frame components of the Riemann tensor, that of the Ricci tensor and the Ricci scalar, are given by

$$
\begin{aligned}
R_{2323} &= (1/4p^2)(2p^2(-p^2+q^2+r^2)\\
&\quad -(p^2-q^2+r^2)(p^2+q^2-r^2)),\\
R_{3131} &= (1/4q^2)(2q^2(p^2-q^2+r^2)\\
&\quad -(p^2+q^2-r^2)(-p^2+q^2+r^2)),\\
R_{1212} &= (1/4r^2)(2r^2(p^2+q^2-r^2)\\
&\quad -(-p^2+q^2+r^2)(p^2-q^2+r^2));
\end{aligned}
\tag{2.8}
$$

$$
\begin{aligned}
R_{11} &= -(1/2q^2r^2)(p^2-q^2+r^2)(p^2+q^2-r^2),\\
R_{22} &= -(1/2r^2p^2)(p^2+q^2-r^2)(-p^2+q^2+r^2),\\
R_{33} &= -(1/2p^2q^2)(-p^2+q^2+r^2)(p^2-q^2+r^2);
\end{aligned}
\tag{2.9}
$$

and

$$
\begin{aligned}
R &= \frac{1}{2p^2q^2r^2}\{p^4+q^4+r^4-2q^2r^2-2r^2p^2-2p^2q^2\}\\
&= \frac{1}{2p^2q^2r^2}\{2(p^4+q^4+r^4)-(p^2+q^2+r^2)^2\}\\
&= \frac{1}{2p^2q^2r^2}\{(p^2+q^2-r^2)^2-4p^2q^2\}\\
&= \frac{1}{2p^2q^2r^2}\{(-p^2+q^2+r^2)^2-4q^2r^2\}\\
&= \frac{1}{2p^2q^2r^2}\{(p^2-q^2+r^2)^2-4r^2p^2\}\\
&= -\frac{1}{2p^2q^2r^2}(p+q+r)(-p+q+r)\\
&\quad \times(p-q+r)(p+q-r);
\end{aligned}
\tag{2.10}
$$

respectively.

We introduce the notation

$$(b_{ab}) = \begin{pmatrix} A & F & E\\ F & B & D\\ E & D & C\end{pmatrix}. \tag{2.11}$$

The Gauss and Codazzi–Mainardi equations have the form

$$\mathrm{adj}(b_{ab}) = \mathrm{diag}(R_{2323} \quad R_{3131} \quad R_{1212})$$

or

$$
\begin{pmatrix} BC-D^2 & DE-CF & DF-BE\\ DE-CF & AC-E^2 & EF-AD\\ DF-BE & EF-AD & AB-F^2\end{pmatrix}
= \begin{pmatrix} R_{2323} & 0 & 0\\ 0 & R_{3131} & 0\\ 0 & 0 & R_{1212}\end{pmatrix}
\tag{2.12}
$$

and

$$
\begin{aligned}
&X_3F - X_2E + A - (1/2q^2)(p^2+q^2-r^2)B\\
&\quad -(1/2r^2)(p^2-q^2+r^2)C = 0,\\
&X_3B - X_2D + (1/2p^2)(3p^2+q^2-r^2)F = 0,\\
&X_3D - X_2C + (1/2p^2)(3p^2-q^2+r^2)E = 0;\\
&X_1E - X_3A + (1/2q^2)(p^2+3q^2-r^2)F = 0,\\
&X_1D - X_3F - (1/2p^2)(p^2+q^2-r^2)A + B\\
&\quad -(1/2r^2)(-p^2+q^2+r^2)C = 0,\\
&X_1C - X_3E + (1/2q^2)(-p^2+3q^2+r^2)D = 0;\\
&X_2A - X_1F + (1/2r^2)(p^2-q^2+3r^2)E = 0,\\
&X_2F - X_1B + (1/2r^2)(-p^2+q^2+3r^2)D = 0,\\
&X_2E - X_1D - (1/2p^2)(p^2-q^2+r^2)A\\
&\quad -(1/2q^2)(-p^2+q^2+r^2)B + C = 0;
\end{aligned}
\tag{2.13}
$$

respectively.

We now solve the Gauss equations. Computing the adjoints of the matrices on both sides of Eq. (2.12) we obtain

$$
\begin{pmatrix} \Delta A & \Delta F & \Delta E\\ \Delta F & \Delta B & \Delta D\\ \Delta E & \Delta D & \Delta C\end{pmatrix}
= \begin{pmatrix} R_{3131}R_{1212} & 0 & 0\\ 0 & R_{1212}R_{2323} & 0\\ 0 & 0 & R_{2323}R_{3131}\end{pmatrix},
\tag{2.14}
$$

where Δ is the determinant of (2.11):

$$\Delta = A(BC - D^2) = B(AC - E^2) = C(AB - F^2), \quad (2.15)$$

since

$$DE - CF = DF - BE = EF - AD = 0$$

as a consequence of (2.12).

The components of the Riemann tensor are given by (2.8).

Combining (2.12), (2.14), and (2.15) we see that

$$A^2R_{2323} = R_{3131}R_{1212}, \quad B^2R_{3131} = R_{1212}R_{2323}, \quad C^2R_{1212} = R_{2323}R_{3131}, \quad (2.16)$$

and

$$\Delta D = 0, \quad \Delta E = 0, \quad \Delta F = 0. \quad (2.17)$$

Equations (2.16) impose restrictions on the range of the parameters (1.2).

We distinguish two cases:
case 1,

$$R_{2323}R_{3131}R_{1212} \neq 0; \quad (2.18)$$

case 2,

$$R_{2323}R_{3131}R_{1212} = 0. \quad (2.19)$$

We shall see that in case 1 we are actually embedding a three-sphere in E^4 and case 2 is not possible. This then proves Theorem 1.

Case 1: Inequality (2.18) implies that

$$\Delta \neq 0. \quad (2.20)$$

Therefore A, B, C are given by (2.16) and

$$D = 0, \quad E = 0, \quad F = 0 \quad (2.21)$$

as follows from (2.17).

As a consequence of the above, the second fundamental form is also invariant under the left translations of the rotation group and the Codazzi–Mainardi equations (2.13) reduce to algebraic equations; as a matter of fact Eqs. (2.13) reduce to the following system of linear equations for A, B, and C:

$$\begin{aligned} &A - (1/2q^2)(p^2 + q^2 - r^2)B \\ &\quad - (1/2r^2)(p^2 - q^2 + r^2)C = 0, \\ &-(1/2p^2)(p^2 + q^2 - r^2)A + B \\ &\quad - (1/2r^2)(-p^2 + q^2 + r^2)C = 0, \\ &-(1/2p^2)(p^2 - q^2 + r^2)A \\ &\quad - (1/2q^2)(-p^2 + q^2 + r^2)B + C = 0. \end{aligned} \quad (2.22)$$

The determinant of this system,

$$1 - (1/4p^2q^2)(p^2 + q^2 - r^2)^2, \quad (2.23)$$

is proportional to the Ricci scalar (2.10).

We now distinguish two subcases.

Subcase a: For spaces with nonvanishing Ricci scalar we have

$$A = \lambda p^2, \quad B = \lambda q^2, \quad C = \lambda r^2 \quad (2.24)$$

as the solution of (2.22). Equations (2.8), (2.12), and (2.24) lead to $p^2 = q^2 = r^2$ and (1.1) then describes a three-sphere of radius $2p$.

Subcase b: For spaces with vanishing Ricci scalar we have

$$r = p + q, \quad (2.25)$$

say, as a consequence of (2.10). Using (2.8) and (2.12) we have

$$BC = 2q(p+q), \quad CA = 2p(p+q), \quad AB = -2pq \quad (2.26)$$

as Gauss equations. Equations (2.26), however, imply $C^2 = -2(p+q)^2$, which is not possible.

Case 2: Equation (2.19) implies that at least two components of the Riemann tensor would have to vanish. It is easy to see that all of the three components cannot vanish; consequently the Ricci scalar is different from zero also. For the sake of definiteness we assume that

$$R_{2323} = 0 \quad \text{and} \quad R_{3131} = 0 \quad (2.27)$$

are the two vanishing components, leading to

$$\begin{aligned} &2p^2(-p^2 + q^2 + r^2) \\ &\quad - (p^2 + q^2 - r^2)(p^2 - q^2 + r^2) = 0, \\ &-(p^2 + q^2 - r^2)(-p^2 + q^2 + r^2) \\ &\quad + 2q^2(p^2 - q^2 + r^2) = 0. \end{aligned} \quad (2.28)$$

Consider (2.28) as a homogeneous linear system for the "unknowns"

$$-p^2 + q^2 + r^2 \quad \text{and} \quad p^2 - q^2 + r^2,$$

having only the trivial solution

$$-p^2 + q^2 + r^2 = 0, \quad p^2 - q^2 + r^2 = 0, \quad (2.29)$$

since the determinant, being proportional to the Ricci scalar, must be different from zero. Equation (2.29), however, implies $r^2 = 0$, which is not possible. This concludes the proof of Theorem 1.

III. EMBEDDING IN A FOUR-SPHERE

Obtaining the relevant equations from Eisenhart[2] and in the spirit of Sec. II we have

$$R_{abcd} = \Omega_{ac}\Omega_{bd} - \Omega_{ad}\Omega_{bc} + \kappa^2(g_{ac}g_{bd} - g_{ad}g_{bc}) \quad (3.1)$$

and

$$X_c\Omega_{ab} - X_b\Omega_{ac} + \Omega_{af}C^f{}_{bc} - \Omega_{fb}\Gamma_{ca}{}^f + \Omega_{fc}\Gamma_{ba}{}^f = 0 \quad (3.2)$$

as Gauss and Codazzi–Mainardi equations, respectively. Introducing the notation

$$\overline{R}_{abcd} = R_{abcd} - \kappa^2(g_{ac}g_{bd} - g_{ad}g_{bc}) \quad (3.3)$$

we have the nonvanishing components of $\overline{R}_{abcd}$ as a consequence of (2.4) and (2.8):

$$\bar{R}_{2323} = (1/4p^2)\{2p^2(-p^2+q^2+r^2) - (p^2-q^2+r^2)(p^2+q^2-r^2)\} - \kappa^2 q^2 r^2,$$
$$\bar{R}_{3131} = (1/4q^2)\{2q^2(p^2-q^2+r^2) - (p^2+q^2-r^2)(-p^2+q^2+r^2)\} - \kappa^2 r^2 p^2, \quad (3.4)$$
$$\bar{R}_{1212} = (1/4r^2)\{2r^2(p^2+q^2-r^2) - (-p^2+q^2+r^2)(p^2-q^2+r^2)\} - \kappa^2 p^2 q^2,$$

where κ^2 is the curvature of the ambient sphere. Using the notation

$$(\Omega_{ab}) = \begin{pmatrix} A & F & E \\ F & B & D \\ E & D & C \end{pmatrix} \quad (3.5)$$

the Gauss and Codazzi–Mainardi equations have the form

$$\mathrm{adj}(\Omega_{ab}) = \mathrm{diag}(\bar{R}_{2323} \quad \bar{R}_{3131} \quad \bar{R}_{1212})$$

or

$$\begin{pmatrix} BC-D^2 & DE-CF & DF-BE \\ DE-CF & AC-E^2 & EF-AD \\ DF-BE & EF-AD & AB-F^2 \end{pmatrix} = \begin{pmatrix} \bar{R}_{2323} & 0 & 0 \\ 0 & \bar{R}_{3131} & 0 \\ 0 & 0 & \bar{R}_{1212} \end{pmatrix} \quad (3.6)$$

and

$$\begin{aligned}
&X_3F - X_2E + A - (1/2q^2)(p^2+q^2-r^2)B - (1/2r^2)(p^2-q^2+r^2)C = 0,\\
&X_3B - X_2D + (1/2p^2)(3p^2+q^2-r^2)F = 0,\\
&X_3D - X_2C + (1/2p^2)(3p^2-q^2+r^2)E = 0;\\
&X_1E - X_3A + (1/2q^2)(p^2+3q^2-r^2)F = 0,\\
&X_1D - X_3F - (1/2p^2)(p^2+q^2-r^2)A + B - (1/2r^2)(-p^2+q^2+r^2)C = 0,\\
&X_1C - X_3E + (1/2q^2)(-p^2+3q^2+r^2)D = 0;\\
&X_2A - X_1F + (1/2r^2)(p^2-q^2+3r^2)E = 0,\\
&X_2F - X_1B + (1/2r^2)(-p^2+q^2+3r^2)D = 0,\\
&X_2E - X_1D - (1/2p^2)(p^2-q^2+r^2)A - (1/2q^2)(-p^2+q^2+r^2)B + C = 0;
\end{aligned} \quad (3.7)$$

respectively.

The calculation of this section is very similar, but slightly more complicated than that of Sec. II and furnishes the proof of Theorem 2.

We now solve the Gauss equations. Computing the adjoints of the matrices on both sides of Eq. (3.6) we obtain

$$\begin{pmatrix} \Delta A & \Delta F & \Delta E \\ \Delta F & \Delta B & \Delta D \\ \Delta E & \Delta D & \Delta C \end{pmatrix} = \begin{pmatrix} \bar{R}_{3131}\bar{R}_{1212} & 0 & 0 \\ 0 & \bar{R}_{1212}\bar{R}_{2323} & 0 \\ 0 & 0 & \bar{R}_{2323}\bar{R}_{3131} \end{pmatrix}, \quad (3.8)$$

where Δ is the determinant of (3.5),

$$\Delta = A(BC-D^2) = B(AC-E^2) = C(AB-F^2), \quad (3.9)$$

and the components of the $\bar{R}_{abcd}$ are given by (3.4). Combining (3.6), (3.8), and (3.9) we see that

$$A^2\bar{R}_{2323} = \bar{R}_{3131}\bar{R}_{1212}, \quad B^2\bar{R}_{3131} = \bar{R}_{1212}\bar{R}_{2323}, \quad C^2\bar{R}_{1212} = \bar{R}_{2323}\bar{R}_{3131}, \quad (3.10)$$

and

$$\Delta D = 0, \quad \Delta E = 0, \quad \Delta F = 0. \quad (3.11)$$

Equations (3.10) impose restrictions on the range of the parameters (1.2).

We distinguish two cases:
case 1,

$$\bar{R}_{2323}\bar{R}_{3131}\bar{R}_{1212} \neq 0; \quad (3.12)$$

case 2,

$$\bar{R}_{2323}\bar{R}_{3131}\bar{R}_{1212} = 0. \quad (3.13)$$

We shall see that case 1 contains, aside from S^3, the one parametric family mentioned above and case 2 gives a single member of that family to the special values of the parameters

$$p=1, \quad q=1, \quad r=2, \quad (3.14)$$

which is the simplest and, in a certain sense, most interesting member of the family (1.1).

Case 1: Inequality (3.12) implies that $\Delta \neq 0$ and consequently, A, B, C are constants and

$$D=0, \quad E=0, \quad F=0 \quad (3.15)$$

hold, showing that the second fundamental form is invariant under the left translations of the rotation group.

The Codazzi–Mainardi equations simplify to

$$\Omega_{af}C^f{}_{bc} - \Omega_{fb}\Gamma_{ca}{}^f + \Omega_{fc}\Gamma_{ba}{}^f = 0$$

or

$$\begin{aligned}
&A - (1/2q^2)(p^2+q^2-r^2)B - (1/2r^2)(p^2-q^2+r^2)C = 0,\\
&-(1/2p^2)(p^2+q^2-r^2)A + B - (1/2r^2)(-p^2+q^2+r^2)C = 0,\\
&-(1/2p^2)(p^2-q^2+r^2)A - (1/2q^2)(-p^2+q^2+r^2)B + C = 0,
\end{aligned} \quad (3.16)$$

giving an additional linear system for A, B, C to satisfy.

We now distinguish two subcases:

$$\text{(a)} \quad 4p^2q^2 - (p^2+q^2-r^2)^2 \neq 0, \quad (3.17)$$

$$\text{(b)} \quad 4p^2q^2 - (p^2+q^2-r^2)^2 = 0, \quad (3.18)$$

corresponding to the nonvanishing or vanishing of the Ricci scalar, respectively.

Subcase a: Equations (3.16) and (3.17) imply

$$p^2 = q^2 = r^2, \quad (3.19)$$

that is, the three-sphere of radius $2p$.

Subcase b: We set

$$r = p + q \quad (3.20)$$

and obtain

$$BC = 2q(p+q)(1-(\kappa^2/2)q(p+q)),$$
$$CA = 2p(p+q)(1-(\kappa^2/2)p(p+q)), \tag{3.21}$$
$$AB = -2pq(1+(\kappa^2/2)pq),$$

and

$$C/(p+q) = A/p + B/q \tag{3.22}$$

as the Gauss and Codazzi–Mainardi equations, respectively.

Using (3.22) to eliminate C from (3.21) we have

$$A = 2p(1-\kappa^2p^2/4)^{1/2},$$
$$B = -2q(1-\kappa^2q^2/4)^{1/2}, \tag{3.23}$$
$$C = 2(p+q)[(1-\kappa^2p^2/4)^{1/2} - (1-\kappa^2q^2/4)^{1/2}],$$

with the restriction

$$2(1-\kappa^2p^2/4)^{1/2}(1-\kappa^2q^2/4)^{1/2} = 1+\kappa^2pq/2 \tag{3.24}$$

or

$$\kappa^2(p^2+pq+q^2) = 3. \tag{3.25}$$

This is the desired solution in case 1.

We construct the embedding of the corresponding family in Sec. IV, but at the moment we turn our attention to the following case.

Case 2: Equations (3.10) and (3.13) imply that at least two of the components of $\overline{R}_{abcd}$ have to vanish. For the sake of definiteness we set

$$\overline{R}_{2323} = 0, \quad \overline{R}_{3131} = 0. \tag{3.26}$$

Equations (3.26) and (3.4) imply

$$2p^2(-p^2+q^2+r^2) - (p^2+q^2-r^2)(p^2-q^2+r^2) = 4\kappa^2p^2q^2r^2,$$
$$-(p^2+q^2-r^2)(-p^2+q^2+r^2) + 2q^2(p^2-q^2+r^2) = 4\kappa^2p^2q^2r^2, \tag{3.27}$$

which we intrepret as the system of linear equations for the "unknowns"

$$-p^2+q^2+r^2 \text{ and } p^2-q^2+r^2. \tag{3.28}$$

We now distinguish two subcases:

$$\text{(a)}\ 4p^2q^2-(p^2+q^2-r^2)^2 \neq 0, \tag{3.29}$$
$$\text{(b)}\ 4p^2q^2-(p^2+q^2-r^2)^2 = 0. \tag{3.30}$$

Subcase a: Equations (3.27) and (3.29) imply, using (2.10), that

$$-p^2+q^2+r^2 = -(2\kappa^2/R)(p^2+3q^2-r^2),$$
$$p^2-q^2+r^2 = -(2\kappa^2/R)(3p^2+q^2-r^2). \tag{3.31}$$

The difference of Eqs. (3.31) gives

$$(1+2\kappa^2/R)(p^2-q^2) = 0. \tag{3.32}$$

If

$$1+2\kappa^2/R \neq 0, \tag{3.33}$$

then

$$p^2 = q^2 \tag{3.34}$$

and from (3.31) it follows that

$$r^2 = (2\kappa^2/R)(r^2-4p^2). \tag{3.35}$$

On the other hand, from (2.10) we see that

$$R = (1/2p^4)(r^2-4p^2) \tag{3.36}$$

and so

$$r^2 = 4\kappa^2p^4. \tag{3.37}$$

From (3.4) and (3.37),

$$\overline{R}_{1212} = p^2 - r^2. \tag{3.38}$$

Therefore we have to assume that

$$\overline{R}_{1212} \neq 0$$

in order to avoid the case of the sphere; however, then (3.10) implies $C=0$ and (3.6) in turn implies $D=0$, $E=0$.

Therefore the Gauss equations read as

$$C=0, \quad D=0, \quad E=0, \quad AB-F^2 = p^2-r^2. \tag{3.39}$$

The Codazzi–Mainardi equations are

$$B=-A, \quad X_3A + (1/2p^2)(r^2-4p^2)F = 0,$$
$$X_3F - (1/2p^2)(r^2-4p^2)A = 0, \tag{3.40}$$
$$X_1A + X_2F = 0, \quad X_2A - X_1F = 0.$$

Equations (3.39) imply

$$A^2 + F^2 = r^2 - p^2 \tag{3.41}$$

and using (1.12) we have the following:

$$A_z = -Rp^2F, \quad F_z = Rp^2A,$$
$$A_x + (1/\sin x)F_y - \cot x F_z = 0, \tag{3.42}$$
$$F_x - (1/\sin x)A_y + \cot x A_z = 0.$$

Equations (3.41) and (3.42) can only be consistent if $R=0$.

The first two equations in (3.42) imply

$$A = f\cos(p^2Rz) + g\sin(p^2Rz),$$
$$F = f\sin(p^2Rz) - g\sin(p^2Rz), \tag{3.43}$$

where f and g are functions of x and y only. From (3.41) it follows that

$$f^2+g^2 = r^2-p^2. \tag{3.44}$$

The second two equations in (3.42) imply

$$f_x - g_y/(\sin x) - p^2R\cot x f = 0,$$
$$g_x + f_y/(\sin x) - p^2R\cot x g = 0. \tag{3.45}$$

Equation (3.44) implies

$$ff_x + gg_x = 0, \quad ff_y + gg_y = 0. \tag{3.46}$$

Equation (3.45) implies

$$-gf_x + fg_x = 0. \tag{3.47}$$

The first equation of (3.46) and (3.47) imply that f and g are functions of y only.

Another consequence of (3.45) is

$$-fg_y + gf_y = p^2R(f^2+g^2)\cos x, \tag{3.48}$$

which is only possible if $R=0$, as already mentioned.

If we assume that

$$1+2\kappa^2/R = 0, \tag{3.49}$$

that is,

$$\kappa^2 = -R/2, \tag{3.50}$$

then Eqs. (3.31) lead to

$$r^2 = p^2 + q^2, \tag{3.51}$$

implying that

$$R = -2/(p^2+q^2), \quad \kappa^2 = 1/(p^2+q^2). \tag{3.52}$$

Equations (3.4) show that

$$\overline{R}_{1212} = -2p^2q^2/(p^2+q^2) \neq 0. \tag{3.53}$$

Equations (3.10) and (3.6) now imply

$$\begin{aligned} &C=0, \quad D=0, \quad E=0,\\ &AB - F^2 = -2p^2q^2/(p^2+q^2), \end{aligned} \tag{3.54}$$

representing the Gauss equations.

In order to find A, B, and F we write the Codazzi–Mainardi equations.

Equation (3.7) implies

$$\begin{aligned} &B = -A, \quad X_3A = F, \quad X_3F = -A,\\ &X_1A + X_2F = 0, \quad X_2A - X_1F = 0. \end{aligned} \tag{3.55}$$

Using (1.12) we find

$$A^2 + F^2 = 2p^2q^2/(p^2+q^2), \tag{3.56}$$

$$\begin{aligned} &A_z = F, \quad F_z = -A,\\ &A_x + (1/\sin x)F_y - \cot x F_z = 0,\\ &F_x - (1/\sin x)A_y + \cot x A_z = 0, \end{aligned} \tag{3.57}$$

which lead to

$$A = f\cos z + g\sin z, \quad F = -f\sin z + g\cos z, \tag{3.58}$$

where f and g are functions of x and y only satisfying

$$f^2 + g^2 = 2p^2q^2/(p^2+q^2) \tag{3.59}$$

and

$$\begin{aligned} &f_x + (1/\sin x)g_y + \cot x f = 0,\\ &g_x - (1/\sin x)f_y + \cot x g = 0. \end{aligned} \tag{3.60}$$

Equation (3.59) implies

$$ff_x + gg_x = 0, \quad ff_y + gg_y = 0. \tag{3.61}$$

A consequence of (3.60) is

$$-gf_x + fg_x = 0,$$

implying that f and g are functions of y only. Another consequence of (3.60) is

$$fg_y - gf_y = -(f^2+g^2)\cot x,$$

which is not possible.

This shows that subcase a is not possible.

Subcase b: Equation (3.26) and $R=0$ implies

$$\begin{aligned} &r = p+q,\\ &\overline{R}_{2323} = 2q(p+q)(1-(\kappa^2/2)q(p+q)) = 0,\\ &\overline{R}_{3131} = 2p(p+q)(1-(\kappa^2/2)p(p+q)) = 0,\\ &\overline{R}_{1212} = -2pq(1+(\kappa^2/2)pq). \end{aligned}$$

In rapid succession we obtain

$$p=q, \quad r=2p, \quad \kappa^2 = 1/p^2, \quad \overline{R}_{1212} = -3p^2 \neq 0,$$

and

$$C=0, \quad D=0, \quad E=0, \quad AB-F^2 = -3p^2$$

as Gauss equations, and

$$\begin{aligned} &B=-A, \quad x_3A = 0, \quad X_3F = 0,\\ &X_1A + X_2F = 0, \quad X_2A - X_1F = 0 \end{aligned}$$

as Codazzi–Mainardi equations, showing that A and F are constants. Since $p=q$ we can set $F=0$ without restriction of generality. We then have

$$\begin{aligned} &A = \sqrt{3}p, \quad B = -\sqrt{3}p,\\ &C=0, \quad D=0, \quad E=0, \quad F=0 \end{aligned} \tag{3.62}$$

contained in (3.23)–(3.25) as the limiting case, where $q \to p$.

This concludes our discussion of the Gauss and Codazzi–Mainardi equations verifying the assertions made above.

In Sec. IV we construct the embedding explicitly.

IV. EMBEDDING

Our basic idea is the following: Embedding in a four-sphere is actually embedding in a five-dimensional Euclidean space E^5; as a matter of fact it is a special case of that embedding. We therefore develop the formalism of the embedding of our three-space in E^5. One has to find five functions

$$\xi^\alpha = Z^\alpha(x,y,z), \quad \alpha = 1,2,3,4,5, \tag{4.1}$$

satisfying the six partial differential equations

$$\delta_{\alpha\beta} Z^\alpha{}_a Z^\beta{}_b = g_{ab}, \tag{4.2}$$

where

$$Z^\alpha{}_a = X_a Z^\alpha. \tag{4.3}$$

The relevant part of the classical differential geometry[2,3] instructs us, however, to do something else: We have to introduce the vector fields

$$\eta^\alpha{}_A, \quad A = 4,5 \tag{4.4}$$

such that

$$\delta_{\alpha\beta} Z^\alpha{}_a \eta^\beta{}_A = 0, \quad A=4,5, \quad a = 1,2,3, \tag{4.5}$$

and

$$\delta_{\alpha\beta}\eta^\alpha{}_A\eta^\beta{}_B = \delta_{AB}, \quad A,B = 4,5, \tag{4.6}$$

and integrate the linear system

$$X_a Z^\alpha{}_b - \Gamma_{ab}{}^f Z^\alpha{}_f = b_{4ab}\eta^\alpha{}_4 + b_{5ab}\eta^\alpha{}_5, \tag{4.7}$$

$$\begin{aligned} &X_a\eta^\alpha{}_4 = -b_{4a}{}^f Z^\alpha{}_f - \mu_a\eta^\alpha{}_5,\\ &X_a\eta^\alpha{}_5 = -b_{5a}{}^f Z^\alpha{}_f + \mu_a\eta^\alpha{}_4, \end{aligned} \tag{4.8}$$

where

$$b_{Aa}{}^f = b_{Aah}g^{hf},$$

with the integrability conditions

$$R_{abcd} = b_{4ac}b_{4bd} - b_{4ad}b_{4bc} + b_{5ac}b_{5bd} - b_{5ad}b_{5bc}, \tag{4.9}$$

$$\begin{aligned} &X_c b_{4ab} - X_b b_{4ac} + b_{4af}C^f{}_{bc} - b_{4fb}\Gamma_{ca}{}^f + b_{4fc}\Gamma_{ba}{}^f\\ &\quad = \mu_b b_{5ac} - \mu_c b_{5ab},\\ &X_c b_{5ab} - X_b b_{5ac} + b_{5af}C^f{}_{bc} - b_{5fb}\Gamma_{ca}{}^f + b_{5fc}\Gamma_{ba}{}^f\\ &\quad = \mu_c b_{4ab} - \mu_b b_{4ac}, \end{aligned} \tag{4.10}$$

$$X_b\mu_a - X_a\mu_b + \mu_f C^f{}_{ab} + (b_{4fa}b_5{}^f{}_b - b_{5fb}b_4{}^f{}_a) = 0, \tag{4.11}$$

which are the Gauss, Codazzi–Mainardi, and Ricci equations, respectively.

Looking at Eqs. (3.1) and (3.2), making the identification

$$b_{4ab} = \Omega_{ab}, \quad b_{5ab} = \kappa g_{ab}, \tag{4.12}$$

and setting

$$\mu_a \equiv 0, \tag{4.13}$$

we indeed see that we are dealing with a special case of embedding in E^5. The Ricci equations are identically satisfied since our b's and g are diagonal.

Equations (4.7) and (4.8) take the form

$$X_a Z^\alpha{}_b - \Gamma_{ab}{}^f Z^\alpha{}_f = b_{4ab}\eta^\alpha{}_4 + b_{5ab}\eta^\alpha{}_5, \tag{4.14}$$

$$X_a \eta^\alpha{}_4 = -b_{4a}{}^f Z^\alpha{}_f, \tag{4.15}$$

$$X_a \eta^\alpha{}_5 = -b_{5a}{}^f Z^\alpha{}_f, \tag{4.16}$$

with

$$\Gamma_{23}{}^1 = (p+q)/p, \quad \Gamma_{32}{}^1 = q/p, \quad \Gamma_{31}{}^2 = -p/q, \quad \Gamma_{13}{}^2 = -(p+q)/q, \quad \Gamma_{12}{}^3 = q/(p+q), \quad \Gamma_{21}{}^3 = -p/(p+q), \tag{4.17}$$

$$b_{4ab} = \mathrm{diag}\left(2p\left(1-\frac{\kappa^2p^2}{4}\right)^{1/2} \quad -2q\left(1-\frac{\kappa^2q^2}{4}\right)^{1/2} \quad 2(p+q)\left\{\left(1-\frac{\kappa^2p^2}{4}\right)^{1/2}-\left(1-\frac{\kappa^2q^2}{4}\right)^{1/2}\right\}\right), \tag{4.18}$$

$$b_{5ab} = \mathrm{diag}(\kappa p^2 \quad \kappa q^2 \quad \kappa(p+q)^2), \tag{4.19}$$

$$b_{4a}{}^b = \mathrm{diag}\left(\frac{2}{p}\left(1-\frac{\kappa^2p^2}{4}\right)^{1/2} \quad -\frac{2}{q}\left(1-\frac{\kappa^2p^2}{4}\right)^{1/2} \quad \frac{2}{p+q}\left\{\left(1-\frac{\kappa^2p^2}{4}\right)^{1/2}-\left(1-\frac{\kappa^2q^2}{4}\right)^{1/2}\right\}\right), \tag{4.20}$$

$$b_{5a}{}^b = \mathrm{diag}(\kappa \quad \kappa \quad \kappa), \tag{4.21}$$

and the condition

$$\kappa^2(p^2+pq+q^2) = 3. \tag{4.22}$$

In order to simplify future calculation we set

$$\kappa^2 = 1 \tag{4.23}$$

and use

$$p^2+pq+q^2 = 3 \tag{4.24}$$

to simplify our expressions. Straightforward calculations show that

$$b_{4ab} = \mathrm{diag}\left(\frac{p(p+2q)}{\sqrt{3}} \quad -\frac{q(q+2p)}{\sqrt{3}} \quad -\frac{1}{\sqrt{3}}(p^2-q^2)\right), \tag{4.25}$$

$$b_{5ab} = \mathrm{diag}(p^2 \quad q^2 \quad (p+q)^2), \tag{4.26}$$

$$b_{4a}{}^b = \mathrm{diag}\left(\frac{p+2q}{p\sqrt{3}} \quad -\frac{q+2p}{q\sqrt{3}} \quad -\frac{1}{\sqrt{3}}\frac{p-q}{p+q}\right), \tag{4.27}$$

$$b_{5a}{}^b = \mathrm{diag}(1 \quad 1 \quad 1), \tag{4.28}$$

keeping in mind (4.24)!

Using the new expressions (4.23)–(4.28) we write (4.14)–(4.16) in all details.

$$X_1 Z^\alpha{}_1 = [p(p+2q)/\sqrt{3}]\eta^\alpha{}_4 + p^2\eta^\alpha{}_5, \tag{4.29a}$$

$$X_1 Z^\alpha{}_2 = [q/(p+q)]Z^\alpha{}_3, \tag{4.29b}$$

$$X_1 Z^\alpha{}_3 = -[(p+q)/q]Z^\alpha{}_2, \tag{4.29c}$$

$$X_2 Z^\alpha{}_2 = -[q(q+2p)/\sqrt{3}]\eta^\alpha{}_4 + q^2\eta^\alpha{}_5, \tag{4.29d}$$

$$X_2 Z^\alpha{}_3 = [(p+q)/p]Z^\alpha{}_1, \tag{4.29e}$$

$$X_3 Z^\alpha{}_3 = -(1/\sqrt{3})(p^2-q^2)\eta^\alpha{}_4 + (p+q)^2\eta^\alpha{}_5, \tag{4.29f}$$

$$X_1 \eta^\alpha{}_4 = -[(p+2q)/p\sqrt{3}]Z^\alpha{}_1, \tag{4.30a}$$

$$X_2 \eta^\alpha{}_4 = [(q+2p)/q\sqrt{3}]Z^\alpha{}_2, \tag{4.30b}$$

$$X_3 \eta^\alpha{}_4 = (1/\sqrt{3})[(p-q)/(p+q)]Z^\alpha{}_3, \tag{4.30c}$$

and

$$X_1 \eta^\alpha{}_5 = -Z^\alpha{}_1, \tag{4.31a}$$

$$X_2 \eta^\alpha{}_5 = -Z^\alpha{}_2, \tag{4.31b}$$

$$X_3 \eta^\alpha{}_5 = -Z^\alpha{}_3. \tag{4.31c}$$

Since the vector field $X_3 = \partial/\partial z$ has this simple representation in our coordinate systems, we would like to obtain the z dependence of our functions first.

Equation (4.29f) has the form

$$Z_{zz} = (p+q)(-[(p-q)/\sqrt{3}]\eta_4 + (p+q)\eta_5). \tag{4.32}$$

(We drop the index α for simplicity of notation and denote partial differentiation of a function by attaching the subscripts x, y, or z, respectively.)

The combination of (4.30c) and (4.31c) leads to

$$(p+q)(-[(p-q)/\sqrt{3}]\eta_4 + (p+q)\eta_5)_z = -4Z_z, \tag{4.33}$$

which then implies

$$Z = S\cos 2z + T\sin 2z + U, \tag{4.34}$$

where S, T, U are arbitrary functions of x and y only.

We substitute (4.34) into (4.29c) and (4.29e) and after straightforward calculation we obtain the following system:

$$S_y - 2\cos xT + \sin xT_x = 0, \tag{4.35a}$$

$$T_y + 2\cos xS - \sin xS_x = 0, \tag{4.35b}$$

$$U_x = [(p-q)/(p+q)]S_x, \tag{4.35c}$$

$$U_y = [(p-q)/(p+q)]\sin xT_x. \tag{4.35d}$$

The content of (4.29b) is the following system:

$$(\sin x)^2S_{xx} + S_{yy} - \tfrac{3}{2}\sin 2xS_x + (2(\sin x)^2 + 4(\cos x)^2)S = 0, \tag{4.36a}$$

$$(\sin x)^2 T_{xx} + T_{yy} - \tfrac{3}{2}\sin 2x T_x + (2(\sin x)^2 + 4(\cos x)^2)T = 0, \tag{4.36b}$$

$$T_{xx} + T = 0. \tag{4.36c}$$

Equations (4.36) are the integrability conditions of (4.35).

Observe that the sum of (4.29a), (4.29d), and (4.29f) does not contain η_4, but

$$(X_1X_1 + X_2X_2 + X_3X_3)Z = 6\eta_5. \tag{4.37}$$

Straightforward calculation shows that

$$\eta_5 = \tfrac{1}{6}\{Z_{xx} + \cot x Z_x + [1/(\sin x)^2] \times [Z_{yy} - 2\cos x Z_{zy} + Z_{zz}]\}. \tag{4.38}$$

Substituting (4.34) into (4.38) using (4.35) and (4.36) we obtain

$$\eta_5 = -S\cos 2z - T\sin 2z + \tfrac{1}{3}[(p-q)/(p+q)](S_{xx} + S) \tag{4.39}$$

and from (4.32) we have

$$\eta_4 = (1/\sqrt{3})[(p-q)/(p+q)](S\cos 2z + T\sin 2z) + (1/\sqrt{3})(S_{xx} + S). \tag{4.40}$$

We now integrate Eqs. (4.35) and (4.36). Equation (4.36c) implies that

$$T = f\cos x + g\sin x, \tag{4.41}$$

where f and g are functions of y only. Substituting into (4.36b) we find

$$\frac{d^2f}{dy^2} + 4f = 0, \quad \frac{d^2g}{dy^2} + g = 0, \tag{4.42}$$

that is,

$$f = a\cos 2y + b\sin 2y, \quad g = c\cos y + d\sin y, \tag{4.43}$$

where

$$a,\ b,\ c,\ d \tag{4.44}$$

are arbitrary constants (vectors in E^5) and

$$T = (a\cos 2y + b\sin 2y)\cos x + (c\cos y + d\sin y)\sin x. \tag{4.45}$$

Substituting (4.45) into (4.35a) and (4.35b) and integrating we find

$$S = \tfrac{1}{4}(a\sin 2y - b\cos 2y - e)\cos 2x + \tfrac{1}{2}(c\sin y - d\cos y)\sin 2x + \tfrac{3}{4}(a\sin 2y - b\cos 2y + \tfrac{1}{3}e). \tag{4.46}$$

where e is a constant (vector in E^5). Equations (4.35c) and (4.35d) give

$$U = [(p-q)/(p+q)] \times (X - a\sin 2y + b\cos 2y - e/3). \tag{4.47}$$

The constant of integration is chosen so that

$$\eta_5 = -Z \tag{4.48}$$

should be satisfied! Then Z is given by

$$Z = S\cos 2z + T\sin 2z + [(p-q)/(p+q)] \times (S - a\sin 2y + b\cos 2y - e/3). \tag{4.49}$$

We now have to specify our constants of integration in order to have η_4 and η_5 as mutually orthogonal unit vectors.

If we now introduce the new vectors

$$a = \frac{p+q}{2}A, \quad b = \frac{p+q}{2}B, \quad c = \frac{p+q}{2}C, \quad d = \frac{p+q}{2}D, \quad e = \frac{p+q}{2}\sqrt{3}E, \tag{4.50}$$

where A, B, C, D, E are mutually orthogonal unit vectors in E^5, we then obtain

$$S = A\frac{p+q}{8}(3+\cos 2x)\sin 2y - B\frac{p+q}{8}(3+\cos 2x)\cos 2y + C\frac{p+q}{4}\sin 2x\sin y - D\frac{p+q}{4}\sin 2x\cos y + E\frac{p+q}{8}\sqrt{3}(1-\cos 2x), \tag{4.51}$$

$$T = A\frac{p+q}{2}\cos x\cos 2y + B\frac{p+q}{2}\cos x\sin 2y + C\frac{p+q}{2}\sin x\cos y + D\frac{p+q}{2}\sin x\sin y, \tag{4.52}$$

$$U = -A\frac{p-q}{8}(1-\cos 2x)\sin 2y + B\frac{p-q}{8}(1-\cos 2x)\cos 2y + C\frac{p-q}{4}\sin 2x\sin y - D\frac{p-q}{4}\sin 2x\cos y - E\frac{p-q}{8}\frac{1}{\sqrt{3}}(1+3\cos 2x), \tag{4.53}$$

$$Z=A\left\{\frac{p+q}{8}(3+\cos 2x)\sin 2y\cos 2z+\frac{p+q}{2}\cos x\cos 2y\sin 2z-\frac{p-q}{8}(1-\cos 2x)\sin 2y\right\}$$
$$+B\left\{-\frac{p+q}{8}(3+\cos 2x)\cos 2y\cos 2z+\frac{p+q}{2}\cos x\sin 2y\sin 2z+\frac{p-q}{8}(1-\cos 2x)\cos 2y\right\}$$
$$+C\left\{\frac{p+q}{4}\sin 2x\sin y\cos 2z+\frac{p+q}{2}\sin x\cos y\sin 2z+\frac{p-q}{4}\sin 2x\sin y\right\}$$
$$+D\left\{-\frac{p+q}{4}\sin 2x\cos y\cos 2z+\frac{p+q}{2}\sin x\sin y\sin 2z-\frac{p-q}{4}\sin 2x\cos y\right\}$$
$$+E\left\{\frac{p+q}{8}\sqrt{3}(1-\cos 2x)\cos 2z-\frac{p-q}{8}\frac{1}{\sqrt{3}}(1+3\cos 2x)\right\}, \tag{4.54}$$

and η_4 and η_5 are then mutually orthogonal unit vectors.

Choosing A, B, C, D, E in the direction of the coordinate axes; denoting the Cartesian coordinates in E^5 by $\xi^1, \xi^2, \xi^3, \xi^4$, and ξ^5; and introducing the notation

$$\xi=\xi^1+i\xi^2,\quad \eta=\xi^3+i\xi^4,\quad \zeta=\xi^5; \tag{4.55}$$

we can write our results in a pleasing form:

$$\begin{aligned}\xi&=\frac{p+q}{2}\left\{\cos x\sin 2z-\frac{i}{4}\left((3+\cos 2x)\cos 2z\right.\right.\\&\qquad\left.\left.-\frac{p-q}{p+q}(1-\cos 2x)\right)\right\}e^{2iy},\\ \eta&=\frac{p+q}{2}\left\{\sin x\sin 2z\right.\\&\qquad\left.-\frac{i}{2}\sin 2x\left(\cos 2z+\frac{p-q}{p+q}\right)\right\}e^{iy},\\ \zeta&=\frac{p+q}{8}\sqrt{3}(1-\cos 2x)\cos 2z\\&\qquad-\frac{p-q}{8}\frac{1}{\sqrt{3}}(1+3\cos 2x),\end{aligned} \tag{4.56}$$

where p and q are restricted by $p^2+pq+q^2=3$!

A straightforward calculation shows that

$$\xi\bar{\xi}+\eta\bar{\eta}+\zeta^2=1, \tag{4.57}$$

which verifies that we are indeed embedding in the unit sphere S^4.

Another straightforward calculation shows that

$$\begin{aligned}(ds)^2&=d\xi\,d\bar{\xi}+d\eta\,d\bar{\eta}+(d\zeta)^2\\&=\{p^2(\cos z)^2+q^2(\sin z)^2\}(dx)^2\\&\quad+(p^2-q^2)\sin x\sin 2z\,dx\,dy\\&\quad+\{(p^2(\sin z)^2+q^2(\cos z)^2)(\sin x)^2\\&\quad+(p+q)^2(\cos x)^2\}(dy)^2\\&\quad+2(p+q)^2\cos x\,dy\,dz\\&\quad+(p+q)^2(dz)^2,\end{aligned} \tag{4.58}$$

which is indeed our metric. Please remember

$$p^2+pq+q^2=3 \tag{4.59}$$

must hold!

The family (4.58) contains a very interesting member at $p=q$, which implies $p=q=1$ due to (4.59). The line element and the "parametric equations" (4.56) simplify as

$$\begin{aligned}(ds)^2&=(dx)^2+((\sin x)^2+4(\cos x)^2)(dy)^2\\&\quad+8\cos x\,dy\,dz+4(dz)^2\end{aligned} \tag{4.60}$$

and

$$\begin{aligned}\xi&=\{\cos x\sin 2z-(i/4)(3+\cos 2x)\cos 2z\}e^{2iy},\\ \eta&=\{\sin x\sin 2z-(i/2)\sin 2x\cos 2z\}e^{iy},\\ \zeta&=(\sqrt{3}/4)(1-\cos 2x)\cos 2z;\end{aligned} \tag{4.61}$$

respectively.

Equation (4.60) is also a member of the one parametric family:

$$\begin{aligned}(ds)^2&=(dx)^2+((\sin x)^2+r^2(\cos x)^2)(dy)^2\\&\quad+2r^2\cos x\,dy\,dz+r^2(dz)^2.\end{aligned} \tag{4.62}$$

This family has an additional symmetry generated by the vector field $\partial/\partial z$ in our coordinate systems. Due to this symmetry (4.62) has other interesting features also. The space sections of the Taub solution[4] have this symmetry.

If we drop (4.59) we can embed (4.58) in E^5 by

$$\begin{aligned}\xi&=\frac{p+q}{2}\left\{\cos x\sin 2z-\frac{i}{4}\left((3+\cos 2x)\cos 2z\right.\right.\\&\qquad\left.\left.-\frac{p-q}{p+q}(1-\cos 2x)\right)\right\}e^{2iy},\\ \eta&=\frac{p+q}{2}\left\{\sin x\sin 2z-\frac{i}{2}\sin 2x\left(\cos 2z\right.\right.\\&\qquad\left.\left.+\frac{p-q}{p+q}\right)\right\}e^{iy},\end{aligned} \tag{4.63}$$

$$\zeta = \frac{p+q}{8}\sqrt{3}(1-\cos 2z)\cos 2z + \frac{p-q}{8}\sqrt{3}(1-\cos 2x),$$

without any restrictions on p and q. [See Ref. 1, with (4.63) the major result there.]

The comparison with (4.56) is quite amusing.

ACKNOWLEDGMENT

We are grateful to Mrs. Dorothy Carmack for typing the manuscript.

[1] I. Ozsváth and L. Sapiro, J. Math. Phys. **28**, 2066 (1987).
[2] L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, NJ, 1926).
[3] M. Spivak, *A Comprehensive Introduction to Differential Geometry* (Publish or Perish, Wilmington, DE, 1979).
[4] A. Taub, Ann. Math. **53**, 472 (1951).

On the harmonic radius and the capacity of an inverse ellipsoid

George Dassios
Department of Mathematics, University of Patras, Patras, Greece

(Received 19 August 1987; accepted for publication 25 November 1987)

It is shown that the capacity of a body, obtained by Kelvin inversion, is equal to the inverse of the harmonic radius of the image domain with respect to the center of inversion. Using the monotonicity of the harmonic radius and an appropriate isoperimetric inequality, lower and upper estimates for the capacity of an inverse ellipsoid are obtained.

I. THEORY

Let V be any bounded, closed, and connected domain of $\mathbb{R}^3$ with smooth boundary S. Assume that the origin of a coordinate system is picked up at some interior point of V, and that S' is the Kelvin image[1] of S with respect to a sphere of unit radius, i.e., S' is the image of S under the unique conformal transformation[2] of $\mathbb{R}^3$:

$$\mathbf{r} \mapsto \mathbf{r}' = (1/r^2)\mathbf{r}. \tag{1}$$

The Kelvin transformation (1) maps V into the domain V' exterior to S'.

The capacity C of S is defined by[3]

$$C = \frac{1}{4\pi} \oint_S \frac{\partial \Psi}{\partial n}\, ds, \tag{2}$$

where Ψ is a harmonic function in V^c vanishing on S, and such that $\Psi = 1 + O(r^{-1})$ as $r \to +\infty$. The capacity C' of S' is defined similarly via a harmonic function ϕ' defined on the exterior of S'. The capacities of S and S' can also be obtained by inspection from the asymptotic expansion of the corresponding capacity fields at infinity. In fact, the capacity C' of S' appears in the monopole term of ϕ' as

$$\phi' = 1 - C'/r + O(1/r^2), \quad r \to +\infty. \tag{3}$$

On the other hand, utilizing the fact that the Kelvin transformation (1) preserves harmonicity, we can invert the exterior boundary value problem that determines ϕ' to the interior problem

$$\begin{aligned} &\Delta\phi = 0, && \text{in } V,\\ &\phi = 0, && \text{on } S,\\ &\phi = 1/r + O(1), && r \to 0+, \end{aligned} \tag{4}$$

in which case the expansion (3) at infinity is inverted to the following expansion at the origin:

$$\phi = 1/r - C' + O(r), \quad r \to 0+. \tag{5}$$

Consequently, the capacity of S' is given by

$$C' = \lim_{r \to 0+} (1/r - \phi). \tag{6}$$

Writing (3) in the form

$$\phi(\mathbf{r}) = 1/r - C' + h(\mathbf{r}), \tag{7}$$

with $h(\mathbf{o}) = 0$, and comparing with the expression (see Ref. 4, p. 58)

$$G(\mathbf{r},\mathbf{r}') = \frac{1}{4\pi}\left[\frac{1}{|\mathbf{r}-\mathbf{r}'|} - \frac{1}{R_{\mathbf{r}'}}\right] + H(\mathbf{r},\mathbf{r}') \tag{8}$$

for the Green's function, where $H(\mathbf{r},\mathbf{r}) = 0$ and $R_{\mathbf{r}'}$ is the harmonic radius of V with respect to $\mathbf{r}'$, we conclude that the capacity C' of the Kelvin image S' is equal to the inverse of the harmonic radius of V with respect to the center of inversion.

Furthermore, using an isoperimetric inequality, due to Schiffer [see Ref. 5, formula (4.15)], we obtain that for any Kelvin pair S, S', the product of the corresponding capacities C, C' is greater than or equal to unity, i.e.,

$$C \cdot C' \geqslant 1. \tag{9}$$

The inequality (9) becomes an equality whenever S (and therefore S') becomes a sphere. In this case, if a is the radius of S, then the radius of the sphere S' is equal to a^{-1}, which implies that

$$C \cdot C' = a \cdot a^{-1} = 1. \tag{10}$$

Hence the minimum value of the product of the capacities of a body and its Kelvin image is attained whenever the body becomes a sphere. It can be easily shown that if the sphere of inversion does not have unit radius, then the product of the capacities of any Kelvin pair is greater than or equal to the square of the radius of inversion.

II. APPLICATION

As an application, we consider a triaxial ellipsoid S,

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \frac{x_3^2}{a_3^2} = 1, \tag{11}$$

with $a_1 > a_2 > a_3 > 0$, and its Kelvin image S',

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \frac{x_3^2}{a_3^2} = (x_1^2 + x_2^2 + x_3^2)^2. \tag{12}$$

The nonlinearity of transformation (1) has as a consequence the transformation of the quadratic surface (11) to the biquadratic surface (12). In view of the monotonicity of the harmonic radius and the maximum principle, we obtain that the harmonic radius $R_{\mathbf{r}'}(V)$ of a domain V with respect to a fixed interior point $\mathbf{r}'$ can be estimated to be a number between the minimum and the maximum distance from $\mathbf{r}'$ to the boundary of V (Ref. 4, p. 59). Consequently, the capacity C' of the inverse ellipsoid (12) satisfies the inequalities

$$1/a_1 \leqslant C' \leqslant 1/a_3, \tag{13}$$

 0022-2488/88/040835-02$02.50

while the capacity C of the ellipsoid (11) is given by[6]

$$C=2\left[\int_0^{+\infty}\left(\frac{dx}{\sqrt{x+a_1^2}\sqrt{x+a_2^2}\sqrt{x+a_3^2}}\right)\right]^{-1}. \qquad (14)$$

The lower bound in (13) can be improved by using the isoperimetric inequality (9) as follows.

Since $a_1>a_2>a_3>0$, it follows that

$$\frac{1}{a_1}=\frac{1}{2}\int_0^{+\infty}\frac{dx}{\sqrt{(x+a_1^2)^3}}$$
$$<\frac{1}{2}\int_0^{+\infty}\frac{dx}{\sqrt{x+a_1^2}\sqrt{x+a_2^2}\sqrt{x+a_3^2}}=\frac{1}{C}, \qquad (15)$$

and (13), in view of (9) and (15), implies

$$C/a_1<1\leqslant CC'\leqslant C/a_3, \qquad (16)$$

or

$$1/C\leqslant C'\leqslant 1/a_3. \qquad (17)$$

The estimates (17) for the capacity of the inverse ellipsoid are sharp in the sense that both inequalities become equalities whenever $a_1=a_2=a_3$.

ACKNOWLEDGMENT

The author wants to express his appreciation to Professor L. E. Payne from Cornell University for fruitful and stimulating discussion during the preparation of the present work.

[1]M. W. Thomson (Lord Kelvin), "Extrait d'une lettre de M. William Thomson," J. Math. Pures Appl. **10**, 364 (1845).

[2]O. D. Kellogg, *Foundations of Potential Theory* (Frederick Ungar, New York, 1929).

[3]L. E. Payne, "Isoperimetric inequalities and their applications," SIAM Rev. **9**, 453 (1967).

[4]C. Bandle, *Isoperimetric Inequalities and Applications* (Pitman, Boston, 1980).

[5]M. M. Schiffer, "Partial differential equations of the elliptic type," in *Lecture Series of the Symposium on Partial Differential Equations, Berkeley, California, 1955*, edited by N. Aronszajn and C. B. Morrey, Jr. (University of Kansas, Lawrence, KS, 1957), pp. 97–149.

[6]G. Dassios and L. Payne, "Estimates for low-frequency elastic scattering by a rigid body," J. Elasticity (in press).

Circle maps and reciprocal winding numbers

B. Söderberg
Department of Theoretical Physics, University of Lund, Sölvegatan 14 A, S-223 62 Lund, Sweden

(Received 10 September 1987; accepted for publication 25 November 1987)

A construction that relates circle maps of mutually reciprocal winding number, belonging to the same criticality class, is presented. It is explicitly invariant under smooth conjugations of either map, and displays a series of remarkable properties, in spite of its simplicity.

I. INTRODUCTION

The subject of universality[1] in circle maps[2,3] has been investigated in several papers, and a number of universal properties have so far been discovered, a large part of which concern irrational winding numbers with a periodic continued fraction expansion. Much more remains to be done in this area, though, especially concerning circle maps with more general irrational winding numbers.

In this paper I will present a rigorous construction that might add to the understanding of this field. The construction relates in a very general way circle maps of mutually reciprocal winding numbers within the same universality class. Thus together with the trivial relation between circle maps differing in winding number by an integer, the complete unimodular group is represented. This is precisely the group of self-similarity transformations in the parameter space of a typical one-parameter set of circle maps, relating all rational winding numbers.

The relation is mediated by a function which, for a large class of irrational winding numbers, is conjectured to be at least once continuously differentiable; numerical evidence for this is given. The construction applies to the subcritical regime, as well as to, e.g., cubic critical maps, and the formalism is independent of smooth coordinate changes.

It is interesting to note a certain relation to the renormalization group approach of Ref. 2, where a transformation is used that can be interpreted as a special case of our construction.

The paper is organized as follows. Section II gives a review of some basic concepts relevant to the subsequent discussion. In Sec. III the main pieces of our construction are defined, and analyzed from a general point of view. Section IV gives an analysis relevant for rational winding numbers, while in Sec. V the case of irrational winding numbers, which is the interesting case, is treated. In Sec. VI, some numerical results are presented, and in Sec. VII, finally, our conclusions are summarized.

II. BASICS ON CIRCLE MAPS

A circle map can be represented as a function f from $\mathbb{R}$ to $\mathbb{R}$ that commutes with the unit translation,

$$f(x+1) = f(x) + 1 . \tag{1}$$

Denoting the unit translation by T, this can be written as

$$fT = Tf . \tag{2}$$

In this paper, we will consider circle maps that are monotonous and, unless otherwise stated, at least once continuously differentiable. Specifically, we will be interested in *subcritical circle maps*, i.e., diffeomorphisms, and *cubic critical maps*, defined as differentiable homeomorphisms with a cubic inflection point, by convention chosen to be at zero argument.

A characteristic property of such a circle map is that it has a unique *winding number* ω, defined in the following way:

$$\omega = \lim_{n\to\infty} [f^n(x) - x]/n . \tag{3}$$

As an example of a circle map, we will use the *sine map*,

$$f(x) = x - (k/2\pi)\sin(2\pi x) + \Omega , \tag{4}$$

where Ω is a parameter controlling the winding number, while k is a nonlinearity parameter. For $0 \leqslant k < 1$, the map is subcritical, while for $k = 1$ it is a cubic critical map.

A rational winding number $\omega = P/Q$ implies the existence of a cycle of period Q, i.e., there is an x_0, such that

$$f^Q(x_0) = x_0 + P \tag{5}$$

or, formally,

$$T^{-P} f^Q(x_0) = x_0 . \tag{6}$$

In fact, there must be two Q cycles, one attractive, the other one repulsive. Generically, for a one-parameter family of circle maps $f(x;\Omega)$, such as the sine map with a fixed $k>0$, the winding number ω is rational for Ω in a set of nonzero measure. Typically, for a given rational ω there corresponds a whole interval in Ω, while for an irrational ω the corresponding Ω is unique. Plotting ω as a function of Ω results in a complicated kind of step function, the *devil's staircase*.

For each of the Q cycles, there is another characteristic number: the *stability index* S, which is a measure of the attraction power of the cycle. It is defined as

$$\begin{aligned} S &= \frac{\partial}{\partial x} T^{-P} f^{Q}(x_0) \\ &= f'(x_0)\cdot f'(f(x_0))\cdot \cdots \cdot f'(f^{Q-1}(x_0)) , \end{aligned} \tag{7}$$

with x_0 a cycle point, as before. Thus S is the product of the values of f' in the N different cycle points. For the kind of circle maps we are considering, S is obviously a non-negative number. For an attractive Q cycle, S is required to be less than 1, while for a repulsive one, S is larger than 1.

Thus for a rational ω, there are two additional characteristic numbers, S and S', corresponding to the stable and the unstable cycle, respectively.

Typically, when varying a parameter that controls the winding number, like Ω, over the entire interval correspond-

 0022-2488/88/040837-06$02.50

ing to a given rational ω, the value of S falls from 1 to a minimum (which is 0 for the critical case) and rises back to 1 again, while S' rises from 1 to a maximum, and then falls to 1 again, in such a way that they together trace out a smooth closed curve, as functions of the parameter.

The three numbers ω, S, and S' are all invariant under a *conjugation*, i.e.,

$$f \to hfh^{-1} \tag{8}$$

with h an arbitrary increasing diffeomorphism of the circle.

III. CONSTRUCTION OF A RECIPROCITY RELATION

The idea behind the construction is the following: for a circle map f with a rational winding number $\omega = P/Q$, according to Eq. (6), the combination

$$T^{-P}f^{Q}$$

has a fix point, where T as before stands for the unit translation. On the other hand, for a map g with the inverse winding number $\omega' = Q/P$, the corresponding combination is $T^{-Q}g^{P}$, which is the same as

$$g^{P}T^{-Q}$$

because of the commutativity, Eq. (2). The obvious similarity between the two expressions suggests there might exist a conjugacy relating them. Thus we make the ansatz

$$VfV^{-1} = T^{-1}, \tag{9}$$

and

$$VT^{-1}V^{-1} = g. \tag{10}$$

Note 1: For this to work, the conjugating function V should *not* be a circle map, since these commute with T. (Thus the ansatz has no meaning in the original space, the circle. It is not until we represent circle maps on $\mathbb{R}$ that it makes sense.)

Note 2: It is not completely self-evident that the above construction is possible, even in principle. A key observation is that if f is a circle map it commutes with T^{-1}. Thus also T^{-1} and g have to commute, being conjugate to f and T^{-1}, respectively. Hence only if g is also a circle map is there no contradiction.

Note 3: Although we started out assuming a rational ω, there is no obvious reason to stick with this restriction, and in the following we will consider arbitrary winding numbers.

Note 4: The above ansatz is of course interesting only if a smooth V can be found, i.e., at least once continuously differentiable. (In the cubic critical case, though, V and V^{-1} must be allowed to have inflections at some isolated points.)

Definition: A pair of circle maps (f,g) with mutually inverse winding numbers will be called *admissible*, if f and g admit a relation of the above type.

A natural question to ask is what restrictions does admissibility put on a pair (f,g). To that end, suppose an admissible pair (f,g) is given. Then transform f and g independently by smooth (circle diffeomorphism) conjugacies A and B:

$$\tilde{f} = AfA^{-1}, \tag{11}$$

and

$$\tilde{g} = BgB^{-1}. \tag{12}$$

By inspection, the transform of V,

$$\tilde{V} = BVA^{-1}, \tag{13}$$

will do for the transformed maps.

Thus admissibility is preserved under independent smooth conjugations of (f,g). This does not answer the above question completely, and we will come back to it in the next sections, since the answer is different for rational and irrational winding numbers. The same goes for questions of uniqueness of V, given an admissible pair (f,g).

Instead, we make the general observation that for all maps we can choose the value of V arbitrarily at least in one point, while for critical maps we gain in smoothness for V if we demand it to map one of the critical points of f (there is one in each unit interval) onto a critical point of g. Both statements are consistent with the following *subsidiary condition*, which we adopt from now on:

$$V(1) = 1. \tag{14}$$

Using Eqs. (9) and (10) we immediately obtain

$$V(0) = g(1) = g(0) + 1, \tag{15}$$

$$V(f(1)) = V(f(0) + 1) = 0, \tag{16}$$

$$V(f(0)) = g(0). \tag{17}$$

Assuming positive winding numbers, we further observe that V, f, and g are unambiguously defined at all values, if V is specified on the interval from 0 to $f(1)$, with values ranging from $g(1)$ down to 0. This defines a kind of "unit cell" in an abstract two-dimensional lattice spanned by unit translations in f space and g space. Here f is obtained from relating the interval $0 < x < 1$ $[g(1) > V(x) > 1]$ to the interval $f(0) < x < f(1)$ $[g(0) > V(x) > 0]$, while g is defined in the same way from V^{-1}.

In what follows we will consider the restriction of V to this fundamental interval, to be referred to as the *unit cell*.

In the critical case, we note that if V and V^{-1} are both differentiable in a neighborhood of 1, V must have a cubic inflection at 0, and an inverse cubic point at $f(1)$ (so that also V^{-1} has a cubic inflection at 0).

We end this section by deriving an equation that V must unconditionally satisfy. Recall the condition for f to be a circle map,

$$fT = Tf. \tag{18}$$

Using Eq. (9) to substitute for f, we obtain

$$V = TVTV^{-1}T^{-1}VT^{-1}. \tag{19}$$

We note in passing that the same equation is satisfied by V^{-1}. For the restriction of V to the unit cell, this equation influences only the four corner points discussed above, and corresponds to matching conditions on V in these points.

IV. RATIONAL WINDING NUMBER

In this section we will briefly consider the rational case. Thus suppose f has a winding number $\omega = P/Q$. Obviously, a cycle point of f must be mapped onto a cycle point of g, since if x_0 is a cycle point of f we have

$$T^{-P}f^{Q}(x_0). \tag{20}$$

Acting with V on this, and using Eqs. (9) and (10) give us, with $y_0 = V(x_0)$,

$$T^{-Q}g^{P}(y_0) = y_0 . \tag{21}$$

Next, we consider a point close to x_0. From the definition of the stability S, Eq. (7), we readily obtain

$$T^{-P}f^{Q}(x_0+\epsilon) = x_0 + S\epsilon . \tag{22}$$

Again acting with V, and using the requirement that V be differentiable, gives us

$$T^{-Q}g^{P}(y_0+\epsilon V'(x_0)) = y_0 + S\epsilon V'(x_0) . \tag{23}$$

We conclude that the stability of a cycle is preserved under the mapping. Thus the attractive (repulsive) cycle of f is mapped onto the attractive (repulsive) cycle of g, and the stabilities S and S' are preserved.

This provides us with a necessary condition for the admissibility of a pair (f,g) in the rational case: They should have the same S and the same S'.

The above considerations indicate that the situation very much resembles the one for the well-known problem of conjugate circle maps, and we will not penetrate it further.

We finish off this section by sketching the answer to the question of uniqueness of V, given f and g. Pick one of the attractive cycle points x_0. On each side it will have a repulsive cycle point. Acting with the combination

$$R = T^{-P}f^{Q}$$

on any point in the interval between x_0 and one of the neighboring repulsive points will result in a new point in the same interval, but closer to x_0. The result is, and we give it without further proof, that V can be chosen arbitrarily in the interval between such a point to the right and its image under R, and independently, in the interval between a similar point to the left and its image under R, up to restrictions on the interval boundaries.

V. IRRATIONAL WINDING NUMBER

The situation for irrational winding numbers is very much different. First of all, the orbit of a point is dense in the circle. Given f and g, a dense subset of the unit cell of V is generated from the starting point (1,1) alone, according to the following algorithm.

Algorithm for generating V:

(i) Start with the point (1,1).

(ii) Generate a new point from the old one (x,y) by acting with the navigators

$$(f,T^{-1}) \text{ if } x\leqslant 1 ,$$

or

$$(T^{-1},g) \text{ if } x\geqslant 1 .$$

(When $x=1$, use either.)

(iii) Go to (ii).

We conclude that, given (f,g) and the subsidiary condition $V(1)=1$ (and continuity of V), *V is unique* for irrational winding number.

The next question to answer is for which pairs (f,g) is V a diffeomorphism on the interval $0<x<f(1)$?

To answer this, we first note that given f, we can always construct an admissible partner g by choosing a smooth V arbitrarily on the interval $0<x<f(0)$, subject to the matching condition $V(f(0)) = V(0)-1$, certifying continuity at $f(0)$.

In order also to have a continuous first derivative, V must in addition fulfill a differential matching condition, easiest obtained by differentiating Eq. (9) at 0. For a subcritical map, we obtain

$$V'(0) = f'(0)\cdot V'(f(0)), \tag{24}$$

while for a cubic critical map, V must have a cubic point at 0, and we have instead

$$V'''(0) = f'''(0)\cdot V'(f(0)) . \tag{25}$$

These conditions make sure that the first derivative of V is continuous at $f(0)$, and it is straightforward to match an arbitrary number of derivatives in a similar manner, provided f is sufficiently differentiable.

As an aside, this is where the similarity, but also the difference, to the renormalization transformation of Ref. 2 lies; their construction slightly resembles ours, provided we neglect differential matching, and instead demand that V be a straight line on $0<x<f(0)$.

Having chosen V on the interval $0<x<f(0)$ in the above way, we can generate the values of V in the intervals $f(0)<x<f^2(0)$, $f^2(0)<x<f^3(0)$, etc., until we reach $f(1)$, by using

$$V(f^n(x)) = V(x) - n , \tag{26}$$

which follows from Eq. (9) by iteration. The step from 1 to $f(1)$ will in the critical case force V to have an inverse cubic point at $f(1)$.

Having thus defined V in the whole unit cell, we are ready to extract g from

$$g(V(x)) = V(x-1) . \tag{27}$$

For $x=1$, g will in the critical case pick up a cubic point at $1=V(1)$, due to the cubic point of V at 0. Similarly, for $x=f(1)$, due to the inverse cubic point of V at $f(1)$, g will pick up a cubic point at $0=V(f(1))$.

From the above considerations, it should be clear that g can be constructed to be as smooth as f is.

Thus we have proved two things for critical and subcritical maps with irrational winding numbers.

(i) For every f there exists a g such that the pair (f,g) is admissible.

(ii) The two maps f and g are either both subcritical, or both critical.

We can combine this result with a conjecture from Ref. 2, applying to winding numbers ω in a certain subset A (introduced by Herman[4]) of the irrationals in the unit interval that has Lebesgue measure 1, and stating that for two generic critical maps having ω in A as a winding number, there exists a once continuously differentiable diffeomorphism, relating the two maps with a conjugation. If this conjecture is true, so is the following one.

Conjecture: For ω in Herman's set, and for two arbitrary critical maps with winding numbers ω and ω^{-1}, respectively, the two maps can be related through the construction presented in this paper, with a V that is at least once differentiable in the interior of the unit cell.

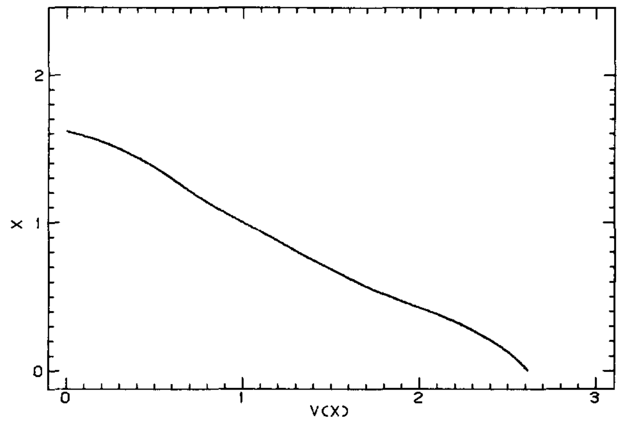


FIG. 1. $V(x)$ for subcritical golden mean (g.m.) sine map, $k_f = k_g = 0.5$.

VI. NUMERICAL RESULTS

I have carried out numerical calculations of V, using sine maps of various criticality, at two different irrational winding numbers, golden mean and $\pi/4$, of which at least the former belongs to Herman's set. Using the algorithm of Sec. V, I have in every case generated 800 points, using double precision arithmetics on the NORD-570 computer of the Physics Institutions in Lund. I have chosen different combinations of critical and subcritical maps for both winding numbers, and the data are presented in graphical form.

For the subcritical case of g.m. (golden mean) winding number, the unit cell of V is shown, for two choices of nonlinearity parameter k, in Figs. 1 and 2.

For the far more interesting critical case, the g.m. V is shown in Fig. 3(a). Note the isolated cubic inflection points at the end points of the unit cell. The importance of the subsidiary condition $V(1) = 1$ is demonstrated in Fig. 3(b), where this condition is relaxed.

The fundamental difference between critical and subcritical maps is illustrated in Figs. 4 and 5, where two examples of f and g of different type are shown, for g.m. winding number. Note the appearance of numerous inflection points over the whole interval.

For winding number $\pi/2$, the analog of Figs. 1, 3(a), 4,

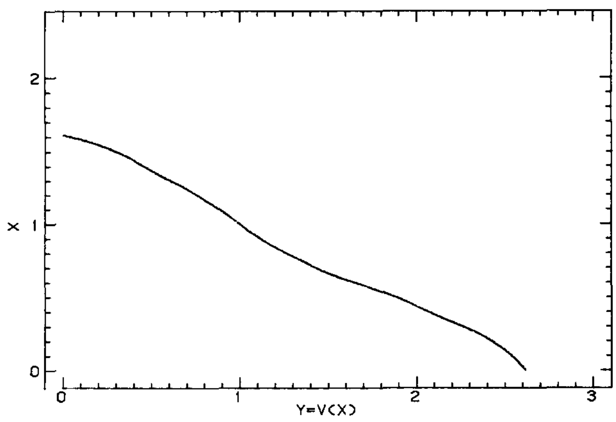


FIG. 2. Same as Fig. 1, but with $k_f = 0.3$ and $k_g = 0.7$.

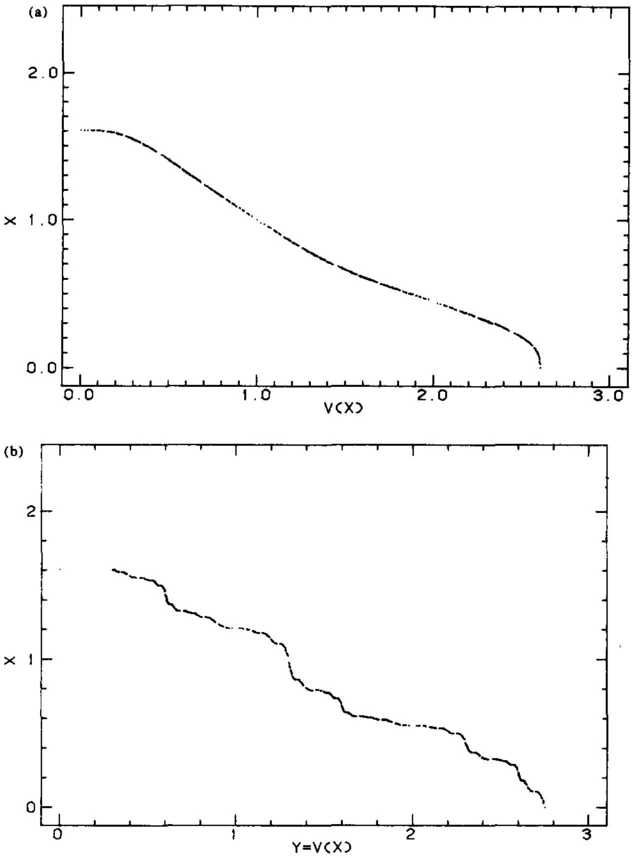


FIG. 3. (a) $V(x)$ for critical g.m. sine map. (b) Same as (a), but without the subsidiary condition $V(1) = 1$. Here, $V(1) = 1.3$.

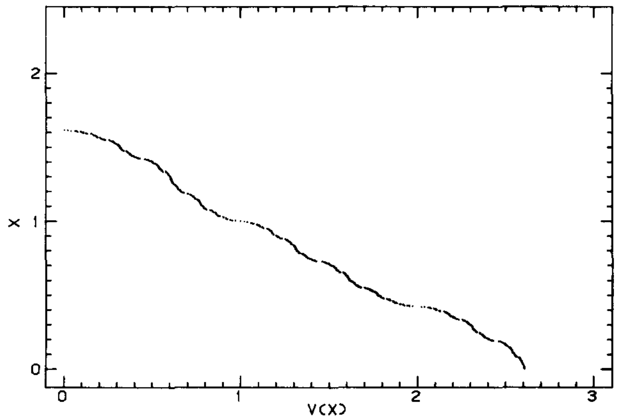


FIG. 4. $V(x)$ for mixed (subcritical f/critical g) g.m. sine map. $k_f = 0.5$.

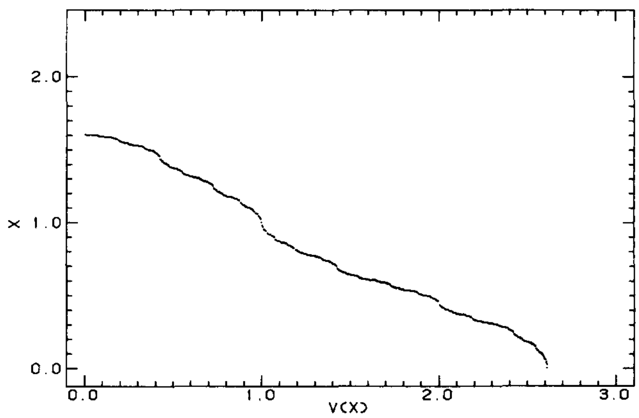


FIG. 5. $V(x)$ for mixed (critical f/subcritical g) g.m. sine map. $k_g = 0.5$.

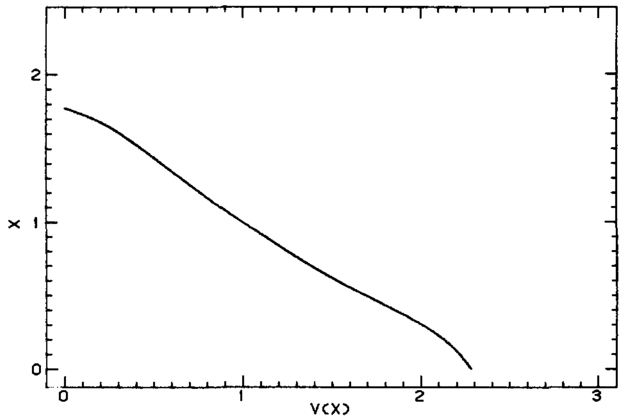


FIG. 6. $V(x)$ for $\omega = \pi/4$ subcritical sine map. $k_f = k_g = 0.5$.

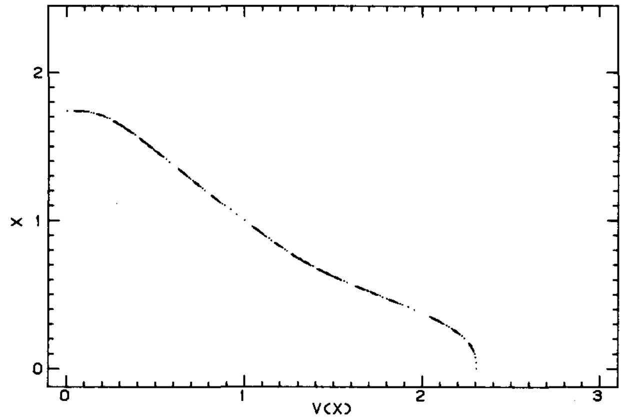


FIG. 7. $V(x)$ for $\omega = \pi/4$ critical sine map.

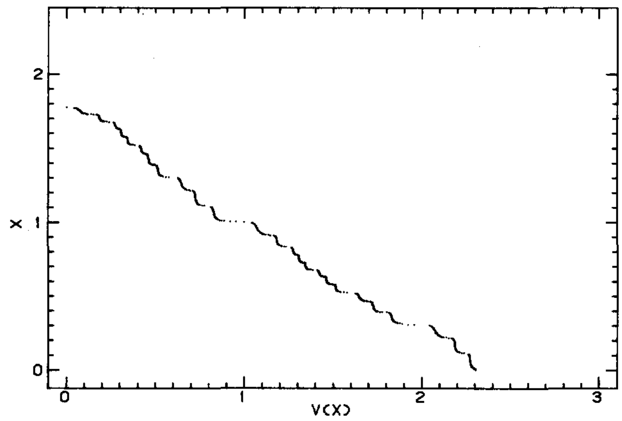


FIG. 8. $V(x)$ for $\omega = \pi/4$ mixed (subcrit. f/crit. g) sine map. $k_f = 0.5$.

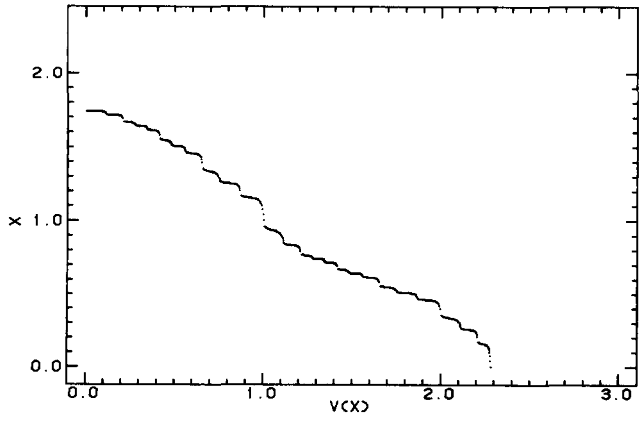


FIG. 9. $V(x)$ for $\omega = \pi/4$ mixed (crit. f/subcrit. g) sine map. $k_g = 0.5$.

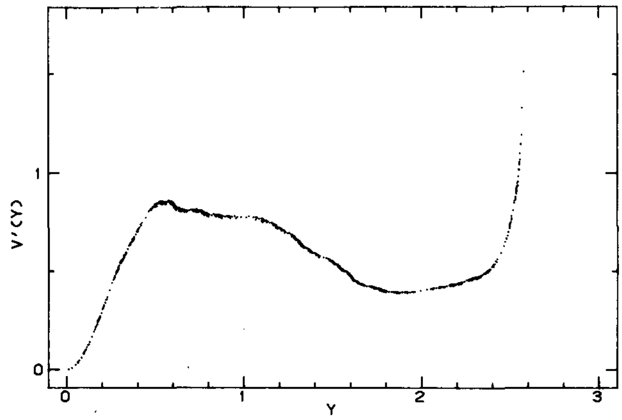


FIG. 10. dV/dx (num.) for critical g.m. sine map.

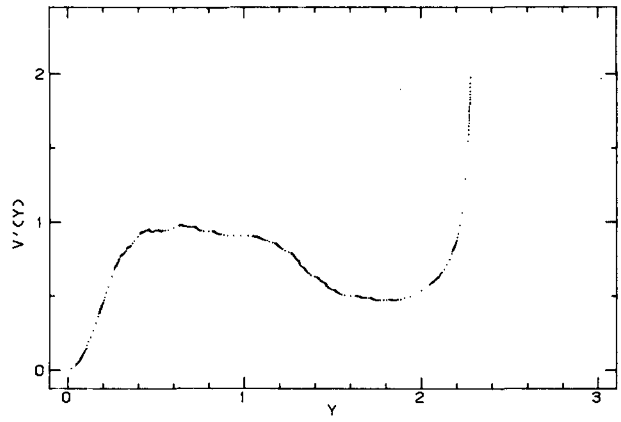


FIG. 11. dV/dx (num.) for critical $\pi/4$ sine map.

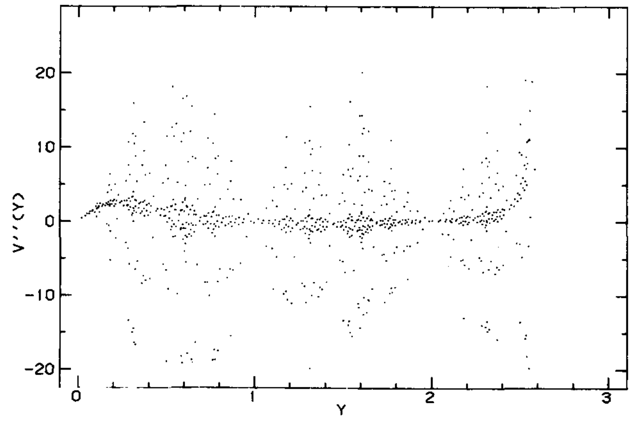


FIG. 12. d^2V/dx^2 (num.) for critical g.m. sine map.

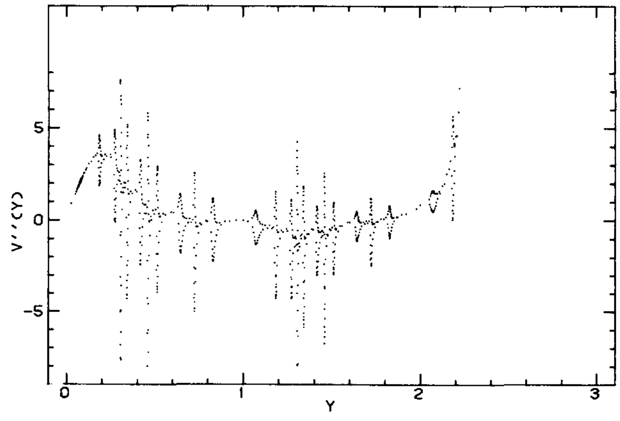


FIG. 13. d^2V/dx^2 (num.) for critical $\pi/4$ sine map.

and 5 are shown in Figs. 6–9, with qualitatively similar results.

For the critical case, and for both winding numbers considered above, I have furthermore computed the two first derivatives of V numerically, simply by using the slope between neighboring data points. The results are shown in Figs. 10 and 11 for the first derivative, and in Figs. 12 and 13 for the second. There is a wiggly structure in Fig. 10; this is probably due to numerical truncation errors, effectively biasing the winding number. Without this bias, the second derivatives would probably be smoother.

Thus the results for both winding numbers indicate an at least once differentiable V when f and g belong to the same class, and an obviously nondifferentiable V for f and g in different classes, or, in the critical case, when the subsidiary condition $V(1) = 1$ is not fulfilled.

VII. CONCLUSIONS

I have presented a construction that relates circle maps in a common criticality class, having mutually reciprocal winding numbers. The construction, in spite of its simplicity, has a number of remarkable properties. One of these is that it seems to have all the properties of a conjugation—except the property of being one. It applies to subcritical as well as cubic maps, and is explicitly invariant under conjugations of either of the maps involved.

This construction thus provides a novel tool that might prove useful for the continuing efforts to acquire a full understanding of the universality in circle maps.

ACKNOWLEDGMENT

Support from the Swedish Natural Science Council, NFR, is greatfully acknowledged.

[1]M. J. Feigenbaum, J. Stat. Phys. **19**, 25 (1975); **21**, 669 (1979).
[2]S. Ostlund, D. Rand, J. Sethna, and E. Siggia, Physica D **8**, 303 (1983).
[3]P. Cvitanović, B. Shraiman, and B. Söderberg, Phys. Scr. **32**, 271 (1985).
[4]M. R. Herman, Publ. Math. I.H.E.S. **49**, 5 (1979).

On a Painlevé test of a coupled system of Boussinesq and Schrödinger equations

Pranab K. Chanda and A. Roy Chowdhury
High Energy Physics Division, Department of Physics, Jadavpur University, Calcutta-700 032, India

(Received 28 March 1986; accepted for publication 4 November 1987)

The complete integrability of a new system of nonlinear equations using the technique of Painlevé analysis has been investigated. The system essentially represents a coupling of Boussinesq and Schrödinger equations through nonlinear terms. While the arbitrariness of the expansion coefficients are proved (for a particular branch) in the formalism of Weiss *et al.* [J. Math. Phys. **24**, 522 (1983)], with the reduced ansatz of Kruskal, the consistency of the truncation is proved by a combination of the methodology due to Weiss [J. Math. Phys. **25**, 13, 2226 (1984)] and Hirota [*Lecture Notes in Physics*, Vol. 515 (Springer, Berlin, 1976)]. On the other hand, the Bäcklund transformation for the equations are obtained via the truncation procedure, without the use of Kruskal's simplification.

I. INTRODUCTION

At present two important approaches exist for the analysis of the nonlinear partial differential equation in two dimensions. One is the method of inverse scattering transform, starting from a Lax pair[1] associated with starting the nonlinear equation for the actual solution of the problem, and the other is a critical analysis of integrability formulated along the line of Painlevé by Ablowitz *et al.*,[2] Weiss *et al.*,[3] Ward,[4] and Jimbo *et al.*[5] It has already been observed that some systems exhibit soliton solutions yet they may not be completely integrable or they may not conform to the criterion of Painlevé. At this point it can be commented that even though a partial differential equation (PDE) passes the test, it may not be completely integrable.[6] Actually, examples already exist illustrating the fact that equations conforming to the Painlevé criterion are not integrable. It can only be said that in many cases the test works. Researchers are trying to understand the properties of more complicated nonlinear systems involving various diverse natures of influence. It has already been observed that KdV, mKdV, sine–Gordon, and Boussinesq[7] equations all are completely integrable. On the other hand, people have observed that the famous equations of Langmuir solitons are not completely integrable but only an approximate version of these can be put into the Lax form,[8] though Ma[9] has obtained by Hirota's approach the N-soliton solution of the original set. Later Goldstein and Infeld[10] made a Painlevé analysis to show (they have used a simplified version of Weiss's approach as suggested by Kruskal) that in spite of the existence of N-soliton solutions these equations (the unapproximated version) of Langmuir solitons do not pass the Painlevé test of Weiss *et al.*

In the wake of such diverse types of results we thought that it is of immense interest to ascertain how integrable systems behave if they are coupled nonlinearly. Of course only those combinations that occur in physical reality will be of interest. Such a system is closely resembled by the equations governing the Whistler mode propagation in a plasma.[11] The equations are essentially a Boussinesq equation coupled to a Schrödinger equation through a nonlinear term. Thus far, to the authors' knowledge, only one soliton solution of such equations has been obtained with no bearing on the question of complete integrability. So here in this communication we have made an analysis of the above-mentioned equation following the approach of Weiss *et al.* Actually, we have sometimes used a variant of this above-mentioned formalism. The arbitrariness of the expansion coefficients has been proved with a reduced ansatz due to Kruskal,[12] by setting $\phi = x - f(t)$. Two branches arise. One of the two branches passes the Painlevé test whereas the other does not. Finally the Bäcklund transformation is obtained by the use of the full machinery of Painlevé analysis and Hirota's[13] technique of bilinearization, without any assumption about $\phi(x,t)$.

II. FORMULATION

The equations under consideration read

$$\left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\sigma}{3}\frac{\partial^4}{\partial x^4}\right) u - \sigma \frac{\partial^2}{\partial x^2}(u^2) = \frac{\partial^2}{\partial x^2}(|\psi|^2),$$
$$\left(i\frac{\partial}{\partial t} + \frac{\partial^2}{\partial x^2} + \lambda + u\right)\psi = 0. \tag{1}$$

For convenience we write this set as

$$\left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\sigma}{3}\frac{\partial^4}{\partial x^4}\right) u - \sigma \frac{\partial^2}{\partial x^2}(u^2) = \frac{\partial^2}{\partial x^2}(\psi\chi),$$
$$\left(i\frac{\partial}{\partial t} + \frac{\partial^2}{\partial x^2} + \lambda + u\right)\psi = 0, \tag{2}$$
$$\left(-i\frac{\partial}{\partial t} + \frac{\partial^2}{\partial x^2} + \lambda + u\right)\chi = 0.$$

The singularity analysis starts by adopting the ansatz

$$u = \sum_{j=0}^{\infty} u_j(x,t)\phi^{\alpha+j}(x,t),$$
$$\psi = \sum_{j=0}^{\infty} \psi_j(x,t)\phi^{\beta+j}(x,t), \tag{3}$$
$$\chi = \sum_{j=0}^{\infty} \chi_j(x,t)\phi^{\gamma+j}(x,t).$$

 0022-2488/88/040843-08$02.50

The singularity manifold is given as $\phi(x,t)=0$. For the leading order analysis we set

$$u \sim u_0(x,t)\phi^{\alpha}(x,t), \quad \psi \sim \psi_0(x,t)\phi^{\beta}(x,t),$$
$$\chi \sim \chi_0(x,t)\phi^{\gamma}(x,t). \tag{4}$$

Then we can have two alternative branches.

(a) $\alpha=-2$, $\beta=\gamma=-1$, $u_0=-2\phi_x^2$, (ψ_0,χ_0) arbitrary. Matching terms are

$$(\sigma/3)u_{xxxx}+\sigma(u^2)_{xx}=0, \tag{5a}$$
$$\psi_{xx}+u\psi=0, \tag{5b}$$
$$\chi_{xx}+u\chi=0. \tag{5c}$$

(b) $\alpha=-2$, $\beta=\gamma=-2$, $u_0=-6\phi_x^2$, $\psi_0\chi_0=-24\sigma\phi_x^4$. Dominant terms are

$$(\sigma/3)u_{xxxx}+\sigma(u^2)_{xx}=-(\psi\chi)_{xx}, \tag{6a}$$
$$\psi_{xx}+u\psi=0, \tag{6b}$$
$$\chi_{xx}+u\chi=0. \tag{6c}$$

At this point it can be mentioned that in the determination of the leading-order coefficients in case (a) the fields ψ, χ become decoupled in leading-order terms but not in case (b). But it can be mentioned that a similar situation is seen to occur in the Hirota–Satsuma case.[14] Here we also observe two branches, but for the case in which (please refer to the Appendix of Ref. 13) $\alpha=-2$, $\beta=-1$, $a=-2\phi_x^2$, $b=$ arbitrary, the dominant terms are

$$-\lambda(uu_x+u_{xxx})=0, \quad v_{xxx}+3uv_x=0,$$

and it is important to note that only this branch passes the Painlevé test and not the other branch that retains the coupling to leading order. Our case is exactly similar to this situation.

III. RESONANCE DETERMINATION

In the following calculation we first consider case (a) and then (b). So we now set

$$u=\sum_{j=0}^{\infty}u_j\phi^{j-2}, \quad \psi=\sum_{j=0}^{\infty}\psi_j\phi^{j-1}, \quad \chi=\sum_{j=0}^{\infty}\chi_j\phi^{j-1} \tag{7}$$

in Eq. (2) and equate the same power of ϕ to get the following recursion relation:

$$\begin{bmatrix} \{-(\sigma/3)(-2+m)(-3+m)(-4+m)(-5+m)+4\sigma(-4+m)(-5+m)\phi_x^4\} & 0 & 0 \\ \psi_0 & [(-1+m)(-2+m)-2]\phi_x^2 & 0 \\ \chi_0 & 0 & \phi_x^2[(-1+m)(-2+m)-2] \end{bmatrix}$$
$$=T\ (\text{the system matrix}) \tag{8}$$

along with

$$T\begin{pmatrix} u_m \\ \psi_m \\ \chi_m \end{pmatrix}=\begin{pmatrix} \text{other terms with} \\ u_j,\psi_j,\chi_j \text{ where} \\ j<m \end{pmatrix}. \tag{9}$$

Now resonances are those values of m for which different arbitrary functions may enter into the expansion u, ψ, χ. These can be obtained by demanding that the characteristic determinant (the determinant of the system matrix) is equal to zero. So if we set

$$\det(T)_{m=r=0}=0,$$

we get

$$r^2(r+1)(r-3)^2(r-4)(r-5)(r-6)=0. \tag{10}$$

So we have resonance at

$$r=0,0,-1,3,3,4,5,6.$$

For the branch (b) we have the system matrix

$$S=\begin{pmatrix} (\sigma/3)[-(-2+m)(-3+m)(-4+m)(-5+m)+36(-4+m)(-5+m)]\,\phi_x^4 & [-(-4+m)(-5+m)\,\phi_x^2\chi_0] & [-(-4+m)(-5+m)\,\phi_x^2\psi_0] \\ \psi_0 & [(-2+m)(-3+m)-6]\,\phi_x^2 & 0 \\ \chi_0 & 0 & [(-2+m)(-3+m)-6]\,\phi_x^2 \end{pmatrix} \tag{11}$$

and the recursion relation

$$S\begin{pmatrix} u_m \\ \psi_m \\ \chi_m \end{pmatrix}=\begin{pmatrix} \text{terms involving } u_j,\psi_j \text{ etc.} \\ \text{with } j<m \end{pmatrix}. \tag{12}$$

Again, setting $\det[S]$ at $m=r=0$ we arrive at the resonance positions

$$r=0,-1,-3,4,5,5,6,8.$$

In the following we will show that only the first branch passes the Painlevé test but the second branch does not. Also it may be pointed out that until this stage of the calculation we have never used the simplified ansatz $\phi=x-f(t)$. As a further remark we may add that the resonance at $r=-3$ does not seem to contribute to the existence of an arbitrary expansion coefficient in (7). Significance of such negative resonances has been studied by Steeb and Louw for ordinary differential equations.[15] Thus we can have seven arbitrary functions corresponding to the resonances at $r = 0, -1, 4, 5, 5, 6, 8$ whereas according to the Cauchy–Kowalevski theorem we should have the number of resonances equal to the number of arbitrary functions. Now we have an indication regarding the failure of the Painlevé test.

In Ref. 15 Steeb and Louw studied an expansion around infinity and made interesting observations. Here one could also assume that

$$u = \sum_{j=0}^{\infty} u_{-j}\phi^{-j-2}, \quad \psi = \sum_{j=0}^{\infty} \psi_{-j}\phi^{-j-1};$$

$$\chi = \sum_{j=0}^{\infty} \chi_{-j}\phi^{-j-1},$$

and use them in (2) in order to get the significance of the resonance $r = -3$ that we get in our case. Before proceeding further one may note an important point regarding rational resonances.[16] It is known that algebraic constants of motion correspond to rational resonances in the case of ordinary differential equations.[16(a),16(b)] This can be extended to partial differential equations.[16(c)] This theorem can be applied to Eq. (5a),

$$\tfrac{1}{3}u_{xxxx} + (u^2)_{xx} = 0.$$

This equation is scale invariant under $x \to \epsilon^{-1}x$, $u \to \epsilon^2 u$.

Since

$$\frac{\partial}{\partial x}\left[\frac{1}{3}u_{xxx} + (u^2)_x\right] = 0,$$

we find [with $I = \frac{1}{3}u_{xxx} + (u^2)_x$] that

$$I(\epsilon^2 u, \epsilon^3 u_x, \ldots) = \epsilon^5 I(u, u_x, \ldots).$$

Thus $r = 5$ must be a resonance. Furthermore, for Eq. (6a) this method can be applied to find a resonance. Equation (6a) is scale invariant under $x \to \epsilon^{-1}x$, $u \to \epsilon^2 u$, $\psi \to \epsilon^2\psi$, $\chi \to \epsilon^2\chi$.

As before, $\partial I'/\partial x = 0$, where $I' = \frac{1}{3}u_{xxx} + (u^2)_x - (\psi\chi)_x$. We find

$$I'(\epsilon^2 u, \epsilon^3 u_x, \ldots, \epsilon^2\psi, \epsilon^2\chi, \ldots) = \epsilon^5 I(u, u_x, \ldots, \psi, \chi),$$

so that $r = 5$ is a resonance. In our above calculation one can see that we really obtained these resonances.

IV. ARBITRARINESS OF THE COEFFICIENTS

We now write out in full the recursion relations (9) and (12) and try to observe what the situations are with the coefficients at the resonance position. We may add that here we take the Kruskal prescription $\phi = x - f(t)$.

For branch (a) we get

$$T\begin{pmatrix} u_m \\ \psi_m \\ \chi_m \end{pmatrix} = \begin{pmatrix} A \\ B \\ C \end{pmatrix} \tag{13}$$

where

$$\begin{aligned} A = {} & -u_{(m-4)tt} + 2u_{(m-3)t}(m-5)f_t \\ & - u_{m-2}(m-4)(m-5)f_t^2 \\ & + u_{m-3}\cdot(m-5)f_{tt} + u_{(m-2)}(m-4)(m-5) \\ & + \sigma\sum_{s=1}^{m-1} u_{m-s}u_s(m-4)(m-5) \\ & + \sum_{s=0}^{m-2} \psi_{m-s-2}(m-4)(m-5), \end{aligned} \tag{14}$$

$$\begin{aligned} B = {} & -i\psi_{(m-2)t} + i\psi_{(m-1)}(m-2)f_t \\ & - \lambda\psi_{m-2} - \sum_{s=1}^{m-1} u_{m-s}\psi_s, \end{aligned}$$

$$\begin{aligned} C = {} & i\chi_{(m-2)} - i\chi_{(m-1)}(m-2)f_t - \lambda\chi_{m-2} \\ & - \sum_{s=1}^{m-1} u_{m-s}\chi_s. \end{aligned}$$

On the other hand, for branch (b) we get

$$S\begin{pmatrix} u_m \\ \psi_m \\ \chi_m \end{pmatrix} = \begin{pmatrix} D \\ E \\ F \end{pmatrix}, \tag{15}$$

$$\begin{aligned} D = {} & -u_{(m-4)tt} + 2(m-5)f_t u_{(m-3)t} \\ & - u_{m-2}(m-4)(m-5)f_t^2 \\ & + u_{m-3}(m-5)f_{tt} + u_{(m-2)}(m-4)(m-5) \\ & + \sigma\sum_{s=1}^{m-1} u_{m-s}u_s(m-4)(m-5) \\ & + \sum_{s=1}^{m-1} \psi_{m-s}\chi_s(m-4)(m-5), \end{aligned} \tag{16}$$

$$\begin{aligned} E = {} & -i\psi_{(m-2)t} + i\psi_{(m-1)}(m-3)f_t \\ & - \lambda\psi_{(m-2)} - \sum_{s=1}^{m-1} u_{m-s}\psi_s, \end{aligned}$$

$$\begin{aligned} F = {} & i\chi_{(m-2)t} - i\chi_{m-1}(m-3)f_t \\ & - \lambda\chi_{(m-2)} - \sum_{s=1}^{m-1} u_{m-s}\chi_s. \end{aligned}$$

Branch (a): $m = -1$ corresponds to arbitrary $f(t)$. Here we get the following: double resonance $m = 0$ corresponds to arbitrary λ_0, χ_0; $u_0 = -2$.

$\underline{m = 1:}$ $u_1 = 0$, $\psi_1 = (i/2)f_t\psi_0$,

$$\chi_1 = -(i/2)\chi_0 f_t, \tag{17}$$

$\underline{m = 2:}$ $u_2 = (1/4\sigma)[2f_t^2 + (\psi_0\chi_0 - 2)]$,

$$\begin{aligned} \psi_2 &= (\psi_0/8\sigma)\,[\,2f_t^2 + (\psi_0\chi_0 - 2) + 4\lambda\sigma\,] + (i/2)\,\psi_{0t}, \\ \chi_2 &= (\chi_0/8\sigma)\,[\,2f_t^2 + (\psi_0\chi_0 - 2) + 4\lambda\sigma\,] - (i/2)\,\chi_{0t}, \end{aligned} \tag{18}$$

$\underline{m = 3:}$ $u_3 = (1/2\sigma)f_{tt}$,

$$\begin{aligned} & u_3\psi_0 + i\,\psi_{1t} - i\psi_2\, f_t + \lambda\psi_1 + u_2\psi_1 = 0, \\ & u_3\chi_0 - i\,\chi_{1t} + i\chi_2\, f_t + \lambda\chi_1 + u_2\chi_1 = 0. \end{aligned} \tag{19}$$

It is easy to observe that use of the first equation of (19) with (17) and (18) leads to $\sigma = 1$, where all of them are identically satisfied. Thus the resonance at this position dictates that the P property is not shown in this branch for $\sigma \neq 1$ and that for $\sigma = 1$ we do have ψ_3 and χ_3 arbitrary.

$\underline{m = 4:}$ This leads to only two equations instead of three as the coefficients in the first row of (13) cancel and we get

$$\psi_4 = -\tfrac{1}{4}[i\psi_{2t} - 2i\psi_3\, f_t + \lambda\psi_2 + u_4\psi_0$$

$$+u_3\psi_1+u_2\psi_2\,],$$
$$\chi_4=-\tfrac{1}{4}[\,-i\chi_{2t}+2i\chi_3\,f_t+\lambda\chi_2+u_4\chi_0+u_3\chi_1+u_2\chi_2\,],$$
$$u_4=\text{arbitrary}. \tag{20}$$

$m=5$: Here also the coefficients of the first row of (13) cancel and we are left with

$$\psi_5=-\tfrac{1}{10}(i\psi_{3t}-3i\psi_4\,f_t+\lambda\psi_3+u_5\psi_0+u_4\psi_1+u_3\psi_2+u_2\psi_3),$$
$$\chi_5=-\tfrac{1}{10}(\,-i\chi_{3t}+3i\chi_4\,f_t+\lambda\chi_3+u_5\chi_0+u_4\chi_1+u_3\chi_2+u_2\chi_3),$$
$$u_5=\text{arbitrary}. \tag{21}$$

$m=6$: The coefficients collected from the first row of (13) do not contain u_6,ψ_6,χ_6 and yield a relation which on use of previous results leads to $\sigma=1$ again. Those obtained from (13) are

$$\psi_6=-\tfrac{1}{18}(i\psi_{4t}-4i\psi_5\,f_t+\lambda\psi_4+u_6\psi_0+u_5\psi_1+u_4\psi_2+u_3\psi_3+u_2\psi_4),$$
$$\chi_6=-\tfrac{1}{18}(\,-i\,\chi_{4t}+4i\chi_5\,f_t+\lambda\chi_4+u_6\chi_0+u_5\chi_1+u_4\chi_2+u_3\chi_3+u_2\chi_4),$$
$$u_6=\text{arbitrary}. \tag{22}$$

From our above analysis we conclude that for $\sigma=1, f(t)$, $u_4, u_5, u_6, \psi_0, \chi_0, \psi_3, \chi_3$ remain arbitrary (equal in number to the number of resonances) satisfying the Cauchy–Kowlavski theorem. Hence at least for branch (a) we can conclude that the equation passes the Painlevé test in the sense of Weiss *et al.* only for $\sigma=1$.

Branch (b): For the second branch we can also follow a similar method of computation as above and our results are as follows. Resonance at $m=-1$ corresponds to arbitrary $f(t)$. For $m=0$ we get

$$u_0=-6,\quad \chi_0\psi_0=-24\sigma. \tag{23}$$

This resonance corresponds to the arbitrariness of ψ_0. Now from the recursion relation we obtain the following.

$m=1$:

$$u_1=0,\quad \psi_1=\tfrac{1}{2}i\psi_0\,f_t,\quad \chi_1=-\tfrac{1}{2}i\chi_0\,f_t. \tag{24}$$

$m=2$: This yields

$$u_2=r/60\sigma+(1/20\sigma)\,[(6-2\sigma)f_t^2-8\lambda\sigma-2],$$
$$\psi_2=(i/6)\,\psi_{0t}+\psi_0\times\left\{\frac{r}{360\sigma}+\frac{[(6-12\sigma)f_t^2+12\lambda\sigma-6\,]}{120\sigma}\right\},$$
$$\chi_2=-(i/6)\,\chi_{0t}+\chi_0\times\left\{\frac{r}{360\sigma}+\frac{[(6-12\sigma)f_t^2+12\lambda\sigma-6\,]}{120\sigma}\right\}, \tag{25}$$

with $r=i\chi_0\psi_{0t}$.

$m=3$: This yields

$$u_3=[(2\sigma+3)/10\sigma\,]f_{tt},$$
$$\psi_3=-\frac{(\sigma-1)}{20\sigma}\psi_0\,f_{tt}-\frac{f_t}{12}\psi_{0t}+\frac{if_t}{12}\left\{\frac{r}{60\sigma}+\frac{1}{20\sigma}[(6-2\sigma)f_t^2+12\lambda\sigma-6]\right\}\psi_0,$$
$$\chi_3=-\frac{(\sigma-1)}{20\sigma}\chi_0\,f_{tt}-\frac{f_t}{12}\chi_{0t}-\frac{if_t}{12}\left\{\frac{r}{60\sigma}+\frac{1}{20\sigma}[(6-2\sigma)f_t^2+12\lambda\sigma-6]\right\}\chi_0. \tag{26}$$

$m=4$: Here the coefficients of the first row of (15) cancel and we get only two equations,

$$\psi_4=\tfrac{1}{4}(i\,\psi_{2t}-i\psi_3\,f_t+\lambda\psi_2+u_3\psi_1+u_2\psi_2+\psi_0u_4),$$
$$\chi_4=\tfrac{1}{4}(\,-i\,\chi_{2t}+i\chi_3\,f_t+\lambda\chi_2+u_3\chi_1+u_2\chi_2+\chi_0u_4), \tag{27}$$

where u_4 is arbitrary.

$m=5$: Again the coefficients collected from the first row of (15) cancel and we are left with

$$\psi_0u_5=-(i\,\psi_{3t}-2i\psi_4\,f_t+\lambda\psi_3+u_4\psi_1+u_3\psi_2+u_2\psi_3),$$
$$\chi_0u_5=-(\,-i\,\chi_{3t}+2i\chi_4\,f_t+\lambda\chi_3+u_4\chi_1+u_3\chi_2+u_2\chi_3). \tag{28}$$

So we conclude that $u_5=0$, ψ_5,χ_5 arbitrary.

To check the consistency between these two equations we eliminate u_5 and then use the previous results to get

$$i(\sigma-1)[6f_{ttt}+5\lambda f_t^3]=0. \tag{29a}$$

For $\sigma\neq 1$, f becomes fixed. On the other hand, for $\sigma=1$, f can be arbitrary, so that for $\sigma=1$ we really have $u_5=0$, ψ_5,χ_5 arbitrary as mentioned above. So if we consider $\sigma=1$, we can proceed further and see what happens at $m=6,7,8$.

It is interesting to note that Eq. (29a) for $\sigma\neq 1$ passes the Painlevé test for ODE's.[2] The analysis is as follows.

The equation is

$$6f_{ttt}+5\lambda f_t^3=0. \tag{29b}$$

The leading-order analysis with $f\sim f_0(t-t_0)^\alpha$ leads to $\alpha=0$. The difficulty is avoided through a transformation $f_t=p$ when Eq. (29b) reduces to

$$6p_{tt}+5\lambda p^3=0. \tag{29c}$$

Now in the leading-order analysis with $p\sim p_0(t-t_0)^\alpha$ one can get only one branch given by

$$\alpha=-1,\quad 5\lambda\,p_0^2+12=0. \tag{29d}$$

We look for a Laurent expansion (in the neighborhood of t_0) of the form

$$p=\sum_{j=0}^{\alpha}p_j(t-t_0)^{\,j-1}. \tag{29e}$$

In order to find the resonance positions we substitute into the equation composed of leading terms ($6p_{tt}+5\lambda p^3=0$) the following form of p:

$$p=p_0(t-t_0)^{-1}+p_r(t-t_0)^{\,r-1}. \tag{29f}$$

The resonances are the roots of the coefficient of $p_r(t-t_0)^{r-3}=0$, and are given by $r=-1,4$, where $r=-1$ corresponds to the arbitrariness of t_0. After substituting (29e) into (29c) and equating the coefficients of different powers of $(t-t_0)$ to zero it may be checked that

$$p_1=0, \quad p_2=0, \quad p_3=0, \quad p_4=\text{arbitrary}.$$

So the expansion (29e) satisfies the Kovalvskaya criterion, and thus (29c) and hence (29b) may be said to pass the Painlevé test.

For $\underline{m=6:}$ ($\sigma=1$, $u_5=0$),

$u_6=$ arbitrary,

$$\psi_6=-(\psi_0 u_6/6)-\tfrac{1}{6}(i\psi_{4t}-3i\psi_5 f_t+\lambda\psi_4+u_4\psi_2 + u_3\psi_3+u_2\psi_4), \tag{30}$$

$$\chi_6=-(\chi_0 u_6/6)-\tfrac{1}{6}(-i\chi_{4t}+3i\chi_5 f_t+\lambda\chi_4+u_4\chi_2 + u_3\chi_3+u_2\chi_4),$$

and the compatibility condition written below

$$\begin{aligned}
&-if_t(\chi_0\psi_5-\psi_0\chi_5)+\tfrac{1}{3}(\chi_0\psi_{4t}-\psi_0\chi_{4t})\\
&\quad+(\lambda/3)(\chi_0\psi_4+\psi_0\chi_4)+(u_4/3)(\chi_0\psi_2+\psi_0\chi_2)\\
&\quad+(u_3/3)(\chi_0\psi_3+\psi_0\chi_3)+(u_2/3)(\chi_0\psi_4+\psi_0\chi_4)\\
&\quad=-u_{2tt}+2u_{3t}f_t+(-2f_t^2+2+4\sigma u_2)u_4+u_3 f_{tt}\\
&\qquad+2\sigma u_3^2+2(\psi_5\chi_1+\psi_1\chi_5)+2(\psi_4\chi_2+\psi_2\chi_4)\\
&\qquad+2\psi_3\chi_3,
\end{aligned} \tag{31}$$

which is satisfied by previous results.

For $\underline{m=7:}$ ($\sigma=1$, $u_5=0$),

$$\begin{aligned}
u_7=\tfrac{1}{32}\{&-[u_{3tt}-4u_{4t}f_t-2u_4 f_{tt}-12u_3u_4]\\
&+6(\psi_7\chi_0+\psi_0\chi_7)+6(\psi_6\chi_1+\psi_1\chi_6)\\
&+6(\psi_5\chi_2+\psi_2\chi_5)+6(\psi_4\chi_3+\psi_3\chi_4)\},
\end{aligned}$$

$$\psi_7=-\tfrac{1}{14}(-i\psi_{5t}-4i\psi_6 f_t+\lambda\psi_5+u_6\psi_1+u_4\psi_3 + u_3\psi_4+u_2\psi_5+\psi_0u_7),$$

$$\chi_7=-\tfrac{1}{14}(-i\chi_{5t}+4i\chi_6 f_t+\lambda\chi_5+u_6\chi_1 + u_4\chi_3+u_3\chi_4+u_2\chi_5+\chi_0u_7). \tag{32}$$

For $\underline{m=8:}$ ($\sigma=1$, $u_5=0$),

$u_8=$ arbitrary,

$$\psi_8=-\tfrac{1}{24}[i\psi_{6t}-5i\psi_7 f_t+\lambda\psi_6+u_7\psi_1 + (u_6\psi_2+u_2\psi_6)+u_3\psi_5+u_4\psi_4+u_8\psi_0],$$

$$\chi_8=-\tfrac{1}{24}[-i\chi_{6t}+5i\chi_7 f_t+\lambda\chi_6+u_7\chi_1 + (u_6\chi_2+u_2\chi_6)+u_3\chi_5+u_4\chi_4+u_8\chi_0], \tag{33}$$

and the compatibility condition

$$\begin{aligned}
&u_{4tt}+12u_6 f_t^2-12u_6-12(2u_6u_2+u_4^2)\\
&\quad=-\tfrac{1}{2}\{i(\chi_0\psi_{6t}-\psi_0\chi_{6t})-5i(\chi_0\psi_7-\psi_0\chi_7)f_t\\
&\qquad+(\lambda+u_2)(\chi_0\psi_6+\psi_0\chi_6)+u_6(\chi_0\psi_2+\psi_0\chi_2)\\
&\qquad+u_4(\chi_0\psi_4+\psi_0\chi_4)+u_3(\chi_0\psi_5+\psi_0\chi_5)\}\\
&\qquad+12(\psi_7\chi_1+\psi_1\chi_7)+12(\psi_5\chi_3+\psi_3\chi_5)+12\psi_4\chi_4,
\end{aligned} \tag{34}$$

which is identically satisfied by our previous results.

However, this branch has a resonance at $r=-3$, which does not contribute to the introduction of arbitrary functions. So here we have eight resonances and seven arbitrary functions: λ_0 (corresponding to $r=0$); $f(t)$ (corresponding to $r=-1$); u_4 (corresponding to $r=4$); ψ_5, χ_5 (corresponding to $r=5$); u_6 (corresponding to $r=6$), u_8 (corresponding to $r=8$), so that the number of arbitrary functions is one less than the number of resonances and one cannot conclude that this branch exhibits the Painlevé property.

V. TRUNCATION OF THE EXPANSION

One of the most important aspects of the Painlevé analysis due to Weiss *et al.* is that sometimes it may be possible to truncate the expansion over the singular manifold at a finite number of terms. But in that case one has an overdetermined set of equations whose consistency is not at all obvious. Let us start from Eqs. (3) and set $u_3=u_4=\cdots=0$, $\psi_i=\chi_i=0$ for $j\geqslant 2$ leading to the following (in all these calculations we set $\sigma=1$):

$$\underline{j=1:}\quad u_1=2\psi_{xx}, \tag{35}$$

$$\underline{j=2:}\quad -12\phi_x^2u_2+6\phi_t^2-6\phi_x^2-8\phi_x\phi_{xxx} + 6\phi_{xx}^2+3\psi_0\chi_0=0, \tag{36}$$

$$\begin{aligned}
\underline{j=3:}\quad &24\phi_x\phi_t\phi_{xt}+6\phi_{xx}\phi_t^2+6\phi_x^2\phi_{tt}-36\phi_x^2\phi_{xx}\\
&-[18\phi_x^2\phi_{xxxx}-6\phi_{xx}^3+72\phi_x^2\phi_{xx}u_2\\
&+24\phi_x^3u_{2x}]+9(\psi_0\chi_0)_x\phi_x=0,
\end{aligned} \tag{37}$$

$$\begin{aligned}
\underline{j=4:}\quad &-12\phi_{xt}^2-12\phi_x\phi_{xtt}-12\phi_t\phi_{xxt}-6\phi_{xx}\phi_{tt}\\
&+18\phi_{xx}^2+24\phi_x\phi_{xxx}-[4\phi_{xxx}^2\\
&-6\phi_{xx}\phi_{xxxx}-12\phi_x\phi_{xxxxx}-12\phi_x^2u_{2xx}\\
&-36\phi_{xx}^2u_2-48\phi_x\phi_{xxx}u_2-72\phi_x u_{2x}\phi_{xx}]\\
&=(3\psi_0\chi_0)_{xx}-6\phi_x(\psi_0\chi_1+\psi_1\chi_0)_x\\
&\quad-3(\psi_0\chi_1+\psi_1\chi_0)\phi_{xx},
\end{aligned} \tag{38}$$

$$\begin{aligned}
\underline{j=5:}\quad &6\phi_{xxtt}-6\phi_{xxxx}-(2\phi_{xxxxxx}+12\phi_{xx}u_{2xx}\\
&+12\phi_{xxxx}u_2+32u_{2x}\phi_{xxx})\\
&=3(\psi_1\chi_0+\chi_1\psi_0)_{xx},
\end{aligned} \tag{39}$$

$\underline{j=6:}$ lastly, for $j=6$ we have

$$u_{2tt}-u_{2xx}-\tfrac{1}{3}u_{2xxxx}-(u_2^2)_{xx}=(\psi_1\chi_1)_{xx}. \tag{40}$$

Similarly, if we again start from the second and third equations of (2) we get the following.

At $\underline{j=1:}$

$$-i\psi_0\phi_t-2\psi_{0x}\phi_x+\psi_0\phi_{xx}-2\phi_x^2\psi_1=0,$$

$$i\chi_0\phi_t-2\chi_{0x}\phi_x+\chi_0\phi_{xx}-2\phi_x^2\chi_1=0. \tag{41}$$

At $\underline{j=2:}$ We get

$$i\psi_{0t}+\psi_{0xx}+\lambda\psi_0+2\phi_{xx}\psi_1+u_2\psi_0=0, \tag{42a}$$

$$-i\chi_{0t}+\chi_{0xx}+\lambda\chi_0+2\phi_{xx}\chi_1+u_2\chi_0=0. \tag{42b}$$

At $\underline{j=3:}$

$$i\psi_{1t}+\psi_{1xx}+\lambda\psi_1+u_2\psi_1=0, \tag{43a}$$

$$-i\chi_{1t}+\chi_{1xx}+\lambda\chi_1+u_2\chi_1=0. \tag{43b}$$

It is also useful to note the following relation that has been repeatedly used in our subsequent calculation:

$$\phi_x^2(\chi_0\psi_1+\psi_0\chi_1)=(\psi_0\chi_0)\phi_{xx}-(\psi_0\chi_0)_x\phi_x. \tag{44}$$

Now after the truncation, we get

$$u=2(\log\phi)_{xx}+u_2,\quad \psi=\psi_0/\phi+\psi_1,$$
$$\chi=\chi_0/\phi+\chi_1. \tag{45}$$

It is also interesting to observe that (u_2,ψ_1,χ_1) satisfy the same set of PDE's as (u,ψ,χ). But the whole process of truncation can only be justified if and only if it can be shown that Eqs. (35)–(39) and (40)–(44) are all mutually consistent. In the following section we set out to prove this compatibility in stages.

VI. COMPATIBILITY OF THE OVERDETERMINED SET

At first we observe that

$$(42a)\times\chi_0+(43a)\times\psi_0$$
$$\Rightarrow -2(\psi_0\chi_0)_x\phi_x+2(\psi_0\chi_0)\,\phi_{xx}$$
$$-2\phi_x^2(\chi_0\psi_1+\psi_0\chi_1)=0, \tag{46}$$

or

$$-(\psi_0\chi_0)_x\phi_x+(\psi_0\chi_0)\,\phi_{xx}=\phi_x^2(\chi_0\psi_1+\psi_0\chi_1).$$

Also (37)–(36)$\times\phi_{xx}$ + $[\partial/\partial x\,(36)]\times\phi_x$ implies

$$\phi_x^2[6\,\phi_{tt}-6\,\phi_{xx}-2\,\phi_{xxxx}-12\,\phi_{xx}u_2$$
$$-3(\chi_0\psi_1+\psi_0\chi_1)\,]=0. \tag{47}$$

For ease of computation we now set $\phi(x,t)=x-f(t)$. From Eq. (36) we get

$$u_2=\tfrac{1}{12}[6f_t^2-6+3(\psi_0\chi_0)]. \tag{48}$$

From (42a) and (43a) we obtain

$$\psi_1=\tfrac{1}{2}[i\psi_0 f_t-2\psi_{0x}],\quad \chi_1=\tfrac{1}{2}[-i\chi_0 f_t-2\chi_{0x}], \tag{49}$$

along with

$$\chi_0\psi_1+\psi_0\chi_1=-(\psi_0\chi_0)_x. \tag{50}$$

So we now substitute these in

$$i\psi_{1t}+\psi_{1xx}+\lambda\psi_1+u_2\psi_1$$

of (43a). To see that this expression vanishes identically if and only if $u_{2x}-\frac{1}{2}f_{tt}=0$, we deduce as follows. We have from (37),

$$6f_{tt}+24\,u_{2x}=9(\psi_0\chi_0)_x. \tag{51}$$

But we also have

$$(\psi_0\chi_0)_x=-(\chi_0\psi_1+\psi_0\chi_1). \tag{52}$$

So from (47),

$$2f_{tt}=-(\chi_0\psi_1+\psi_0\chi_1)$$

leads to $u_{2x}-\frac{1}{2}f_{tt}=0$, which is merely a consequence of Eq. (43a). It is not very difficult to observe that in this way it can be shown that all of these overdetermined sets of equations are compatible with each other.

VII. A GENERALIZED PROOF VIA HIROTA'S APPROACH

A general proof of the truncation, and thereby a proof of the justification of the associated Bäcklund transformation, can be made if we convert our full set of equations to bilinear form. If we set

$$u=(2\log f)_{xx}=D_x^2 f{\cdot}f/f^2,\quad \psi=g/f,\quad \chi=h/f \tag{53}$$

in Eqs. (2) then we can write these as

$$(D_t^2-D_x^2-(\sigma/3)\,D_x^4+\rho)f{\cdot}f=gh, \tag{54}$$

where ρ is a constant of integration, along with

$$(iD_t+D_x^2+\lambda)g{\cdot}f=0,$$
$$(-iD_t+D_x^2+\lambda)h{\cdot}f=0. \tag{55}$$

An important observation that encourages us to combine the techniques of Hirota and Weiss is that the Bäcklund transformation (45) can be written in a neat form

$$f^{(n)}=\phi_{(n-1)}f^{(n-1)},$$
$$g^{(n)}=\psi_0\,f^{(n-1)}+\phi_{(n-1)}g^{(n-1)}, \tag{56}$$
$$h^{(n)}=\chi_0\,f^{(n-1)}+\phi_{(n-1)}h^{(n-1)}.$$

We now assume that if the set (f^n,g^n,h^n) satisfies the Hirota set of equations (54) and (55), then also $(f^{n-1},g^{n-1},h^{n-1})$ if and only if the overdetermined system of equations obtained from the Weiss analysis is satisfied. To proceed with the induction procedure we substitute (56) in (54) and (55) when (f,g,h) are replaced by (f^n,g^n,h^n), for $\sigma=1$:

$$D_t^2(\phi_{n-1}f^{n-1}{\cdot}\phi_{n-1}f^{n-1})$$
$$-D_x^2(\phi_{n-1}f^{n-1}{\cdot}\phi_{n-1}f^{n-1})$$
$$-\tfrac{1}{3}D_x^4(\phi_{n-1}f^{n-1}{\cdot}\phi_{n-1}f^{n-1})$$
$$+\rho(\phi_{n-1}f^{n-1}{\cdot}\phi_{n-1}f^{n-1})$$
$$=(\psi_0\,f^{n-1}+\phi_{n-1}g^{n-1})(\chi_0\,f^{n-1}+\phi_{n-1}h^{n-1}) \tag{57}$$

or

$$\{2[\phi_{(n-1)}\phi_{(n-1)tt}-\phi^2_{(n-1)t}](f^{n-1})^2$$
$$+\phi^2_{n-1}D_t^2(f^{n-1}{\cdot}f^{n-1})\}-\{2[\phi_{(n-1)}\phi_{(n-1)xx}$$
$$-\phi^2_{(n-1)x}](f^{n-1})^2+\phi^2_{n-1}D_x^2[f^{n-1}{\cdot}f^{n-1}]\}$$
$$+\;-\tfrac{1}{3}[2\phi_{(n-1)}\phi_{(n-1)xxxx}-8\,\phi_{(n-1)x}\phi_{(n-1)xxx}$$
$$+6\,\phi^2_{(n-1)xx}](f^{n-1})^2-\tfrac{8}{3}[2\,\phi_{(n-1)}\phi_{(n-1)}$$
$$-2\,\phi^2_{(n-1)x}][2f^{n-1}f^{n-1}_{xx}-2(f^{n-1}_x)^2]$$
$$-\tfrac{1}{3}\phi^2_{n-1}D_x^4(f^{n-1}{\cdot}f^{n-1})$$
$$+\rho(\phi_{n-1}f^{n-1}{\cdot}\phi_{n-1}f^{n-1})$$
$$=\psi_0\chi_0(f^{n-1})^2+(f^{n-1})^2\phi_{n-1}[\chi_0(g^{n-1}/f^{n-1})$$
$$+\psi_0(h^{n-1}/f^{n-1})]+\phi^2_{(n-1)}g^{n-1}h^{n-1} \tag{58}$$

or

$$2[\phi_{(n-1)}\phi_{(n-1)tt}-\phi^2_{(n-1)t}](f^{n-1})^2$$
$$-2[\phi_{(n-1)}\phi_{(n-1)xx}-\phi^2_{(n-1)x}](f^{n-1})^2$$
$$-\tfrac{1}{3}[2\phi_{(n-1)}\phi_{(n-1)xxxx}$$
$$-8\phi_{(n-1)x}\phi_{(n-1)xxx}+6\,\phi^2_{(n-1)xx}](f^{n-1})^2$$
$$-\tfrac{8}{3}[2\,\phi_{(n-1)}\phi_{(n-1)xx}-2\,\phi^2_{(n-1)x}]$$
$$\times[2f^{n-1}f^{n-1}_{xx}-2(f^{n-1}_x)^2]$$

$$+\left[\phi^2_{(n-1)}(D_t^2-D_x^2-\tfrac{1}{3}D_x^4+\rho)(f^{n-1}\cdot f^{n-1})\right.$$
$$\left.-\phi^2_{(n-1)}\cdot g^{n-1}h^{n-1}\right]$$
$$=\psi_0\chi_0(f^{n-1})^2+(f^{n-1})^2\phi_{n-1}[\chi_0(g^{n-1}/f^{n-1})$$
$$+\psi_0(h^{n-1}/f^{n-1})]. \qquad (59)$$

Now the terms in the last bracket on the left-hand side equal zero, and if we now divide by $(f^{n-1})^2$ on both sides then we get

$$2[\phi_{(n-1)}\phi_{(n-1)tt}-\phi^2_{(n-1)t}]$$
$$-2[\phi_{(n-1)}\phi_{(n-1)xx}-\phi^2_{(n-1)x}]$$
$$-\tfrac{1}{3}[2\phi_{(n-1)}\phi_{(n-1)xxxx}-8\phi_{(n-1)x}\phi_{(n-1)xxx}$$
$$+6\phi^2_{(n-1)xx}]-\tfrac{6}{3}[2\phi_{(n-1)}\phi_{(n-1)xx}-2\phi^2_{(n-1)x}]$$
$$\cdot[2f^{n-1}f^{n-1}_{xx}-2(f^{n-1}_x)^2]\cdot[1/(f^{n-1})^2]$$
$$=\psi_0\chi_0+\phi_{n-1}[\chi_0(g^{n-1}/f^{n-1})$$
$$+\psi_0(h^{n-1}/f^{n-1})]. \qquad (60)$$

Now it is easy to observe that

$$\frac{2[f^{n-1}\cdot f^{n-1}_{xx}-(f^{n-1}_x)^2]}{[f^{n-1}]^2}=2(\ln f^{n-1})_{xx}=u_2, \qquad (61)$$

and then by rewriting ϕ for ϕ_{n-1} and ψ_1 for g^{n-1}/f^{n-1} and χ_1 for h^{n-1}/f^{n-1}, we have

$$2(\phi\phi_{tt}-\phi_t^2)-2(\phi\phi_{xx}-\phi_x^2)-\tfrac{1}{3}(2\phi\phi_{xxxx}$$
$$-8\phi_x\phi_{xxx}+6\phi^2_{xx})-2[2\phi\phi_{xx}-2\phi_x^2]u_2$$
$$=\phi(\chi_0\psi_1+\psi_0\chi_1)+\psi_0\chi_0. \qquad (62)$$

So we can write

$$6(\phi\phi_{tt}-\phi_t^2)-6(\phi\phi_{xx}-\phi_x^2)-(2\phi\phi_{xxxx}$$
$$-8\phi_x\phi_{xxx}+6\phi^2_{xx})-6[2\phi\phi_{xx}-2\phi_x^2]u_2$$
$$=3\psi_0\chi_0+3\phi(\chi_0\psi_1+\psi_0\chi_1), \qquad (63)$$

so that

$$\phi[6\phi_{tt}-6\phi_{xx}-2\phi_{xxxx}-12\phi_{xx}u_2$$
$$-3(\chi_0\psi_1+\psi_0\chi_1)]-[6\phi_t^2-6\phi_x^2-8\phi_x\phi_{xxx}$$
$$+6\phi^2_{xx}-12\phi_x^2u_2+3\psi_0\chi_0]=0, \qquad (64)$$

and this equation is identically satisfied if and only if the expression in each of the brackets vanishes individually, which is obtained from Eqs. (36) and (37).

Let us differentiate Eq. (36) with respect to x and multiply by ϕ_x to get

$$12\phi_x^3u_{2x}=-24\phi_x^2\phi_{xx}u_2+12\phi_x\phi_t\phi_{xt}-12\phi_x^2\phi_{xx}$$
$$+4\phi_x\phi_{xx}\phi_{xxx}-8\phi_x^2\phi_{xxxx}+3\phi_x(\psi_0\chi_0)_x. \qquad (65)$$

Also by multiplying (36) by ϕ_{xx} and by rearranging we get

$$6\phi_{xx}\phi_t^2=12\phi_x^2\phi_{xx}u_2+6\phi_x^2\phi_{xx}$$
$$+8\phi_x\phi_{xx}\phi_{xxx}-6\phi^3_{xx}-3\phi_{xx}(\psi_0\chi_0). \qquad (66)$$

Now using (65), (66), and (37) we arrive at

$$6\phi_{tt}-6\phi_{xx}-2\phi_{xxxx}-12\phi_{xx}u_2-3(\chi_0\psi_1+\psi_0\chi_1)=0.$$

It is then easily observed that (64) is identically satisfied when (36) and (37) are used. We similarly get

$$\phi(i\psi_{0t}+\psi_{0xx}+\psi_0u_2+2\phi_{xx}\psi_1+\lambda\psi_0)$$
$$+(-i\psi_0\phi_t+\psi_0\phi_{xx}-2\psi_{0x}\phi_x-2\phi_x^2\psi_1)=0 \qquad (67)$$

along with

$$\phi(-i\chi_0+\chi_{0xx}+\chi_0u_2+2\phi_{xx}\chi_1+\lambda\chi_0)$$
$$+(i\chi_0\phi_t+\chi_0\phi_{xx}-2\chi_{0x}\phi_x-2\phi_x^2\chi_1)=0, \qquad (68)$$

which are identically satisfied when Eqs. (41) and (42) are satisfied.

VIII. BÄCKLUND TRANSFORMATION

Since (u,u_2), $(\psi,\psi_1,\chi,\chi_1)$ form, respectively, two sets of equations for the same equation we can try to iterate the Bäcklund transformation BT to construct the new nontrivial solution by starting from a trivial one. Since the starting solution looks similar in both of the formalisms, we start from the Hirota set of equations (53) and (55), and set

$$f^{(0)}=1+\epsilon f_1+\epsilon^2f_2+\cdots,$$
$$g^{(0)}=Ae^{i(kx-wt)}[1+\epsilon g_1+\epsilon^2g_2+\cdots], \qquad (69)$$
$$h^{(0)}=Be^{-i(kx-wt)}[1+\epsilon h_1+\epsilon^2h_2+\cdots],$$

which immediately leads to

$$f_1=-h_1=-g_1,\quad f_{1tt}-f_{1xx}-\tfrac{1}{3}f_{1xxxx}+2\rho f_1=0,$$
$$f^2_{1x}-f_1f_{1xx}=0,\quad (D_t^2-D_x^2-\tfrac{1}{3}D_x^4)f_1\cdot f_1=0. \qquad (70)$$

To obtain a general solution we set

$$f_1=p(x)q(t)=Pe^{ax+bt}+Qe^{ax-bt}, \qquad (71)$$

$b=(k-2\rho)^{1/2}$, $a^2+a^4/3=k$. But the third equation of (70) sets either P or Q equal to zero. So finally we obtain

$$f_1=Pe^{ax+bt},\quad g_1=-f_1,\quad h_1=-f_1,$$
$$f^{(0)}=1+Pe^{ax+bt},\quad AB=\rho,$$
$$g^{(0)}=Ae^{i(kx-wt)}[1-Pe^{ax+bt}],$$
$$h^{(0)}=Be^{-i(kx-wt)}[1-Pe^{ax+bt}],$$
$$b=\sqrt{k-2\rho},\quad 3a^2+a^4=3k,\quad w-k^2+\lambda=0, \qquad (72)$$
$$u=\frac{2a^2Pe^{ax+bt}}{(1+Pe^{ax+bt})^2},$$
$$\psi=\frac{A[1-Pe^{ax+bt}]e^{i(kx-wt)}}{(1+Pe^{ax+bt})},$$
$$\chi=\frac{B[1-Pe^{ax+bt}]e^{-i(kx-wt)}}{(1+Pe^{ax+bt})},$$

which is nothing but the one-soliton solution. Since in the previous section we have proved the equivalence of the formalism of Weiss *et al.* and that of Hirota for $\sigma=1$, this solution is actually the solution of the BT determined by the truncation. Actually, as observed by Steeb *et al.*[13(b)] and later by Gibbon *et al.*[13(c)] (through a more detailed investigation) the truncation in Weiss formalism and in the formalism of Hirota is identical.

ACKNOWLEDGMENT

It is a pleasure to acknowledge helpful criticism of the referee for the improvement of the paper.

[1]"Backlund transformation," in *Lecture Note in Mathematics*, Vol. 515, edited by R. M. Miura (Springer, Berlin, to be published).
[2]M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715 (1980).
[3]J. M. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1983).
[4]R. Ward, Phys. Lett. A **102**, 279 (1984).
[5]M. Jimbo, M. D. Kruskel, and T. Miwa, Phys. Lett. A **92**, 59 (1982).
[6]P. Clarkson, Phys. Lett. A **109**, 205 (1985).
[7]J. Weiss, J. Math. Phys. **25**, 13, 2226 (1984).
[8]A. Roy Chowdhury and T. Roy, J. Math. Phys. **20**, 1559 (1979).
[9]Y. C. Ma, Stud. Appl. Math. **60**, 73 (1979).
[10]P. Goldstein and E. Infeld, Phys. Lett. A **103**, 8 (1984).
[11]I. L. Bogolubsky and V. O. Makhankov, Dubna preprint E4-9425 (1975).
[12]M. D. Kruskal (private communication); see also Ref. 5.
[13](a) R. Hirota, *Lecture Notes in Physics*, Vol. 515 (Springer, Berlin, 1976); (b) W.-H. Steeb, M. P. Kloke, and B. M. Spieker, J. Phys. A: Math. Gen. **17**, L825 (1984); (c) J. D. Gibbon, P. Rodmore, M. Tabor, and D. Wood, Stud. Appl. Math. **72**, 39 (1985).
[14]See the discussion in the Appendix of Ref. 13.
[15]W.-H. Steeb and J. A. Louw, J. Math. Phys. **27**, 2458 (1986).
[16](a) H. Yosida, Celes. Mech. **31**, 363 (1983); (b) **31**, 381 (1983); (c) W.-H. Steeb and J. A. Louw, Prog. Theor. Phys. **76**, 1177 (1986).

Scattering of waves from a random cylindrical surface

Hisanao Ogura and Hiroshi Nakayama
Department of Electronics, Kyoto University, Yoshida, Kyoto, 606, Japan

(Received 13 May 1986; accepted for publication 18 November 1987)

The present paper deals with the scattering of waves in two-dimensional space by the random surface of a circular object, which is meant to be a preliminary study for treating three-dimensional scattering by a random sphere. The theory is formulated using a stochastic functional method and a group-theoretic consideration related to the rotation of the circle, in a manner analogous to the authors' previous treatment of the scattering by a planar random surface [Radio Sci. **15**, 1049 (1980); J. Math. Phys. **22**, 471 (1981); Radio Sci. **16**, 831, 847 (1981); J. Opt. Soc. Am. **A 2**, 2208 (1985)]. First, the randomly scattered wave for cylindrical wave injection is given in terms of the Wiener–Hermite functional of the random field on the circle, and then the scattered field for plane-wave injection is synthesized by superposing cylindrical waves. The differential cross sections for the coherent and incoherent scattering are obtained, and a statistical version of the optical theorem is shown to hold. Some numerical calculations are made for the Mie scattering by the random circular surface with Dirichlet and Neumann conditions.

I. INTRODUCTION

The problem of wave scattering by a sphere with random surface is often encountered in practical problems, such as light scattering by interplanetary dust particles,[1–3] Raman scattering by microcrystalline particles,[4–5] the radio-wave scattering by distorted rain drops,[6] and radio-wave propagation along the spherically curved irregular surface of the earth. The present work has been motivated actually by the need of theoretical analysis for such scattering phenomena, which also is a theoretically interesting problem as the stochastic scattering by a random spherical object.

On the other hand, the problem of the scattering by a random planar surface, including theories and applications, has been studied so far by a number of scholars using various techniques such as the small-perturbation method,[7] Kirchhoff approximation,[8] renormalization techniques,[9–14] etc.; theoretical difficulty being mainly due to the multiple scattering. In a series of preceding works of the authors,[15–18] the scattering of a plane wave by a random planar surface has been treated successfully by means of the stochastic functional approach, which was previously introduced in the theory of propagation in random media.[19,20] In those works, the random planar surface is regarded as a statistically homogeneous Gaussian random field, and the scatterred wave field is regarded as a stochastic nonlinear functional of the Gaussian random surface. A nonlinear functional of a Gaussian process can be dealt with by means of Wiener–Hermite functionals; the theory originally due to Wiener and others[21–24] has found diverse applications such as the turbulence[25–27] and system theory.[28–30] By a group-theoretic consideration on the stochastic homogeneity of the random surface, the random wave field scattered by it is shown to have a specific form analogous to the Floquet theorem for a periodic surface. The wave field as a stochastic functional is then developed into a series of orthogonal Wiener–Hermite (WH) functionals satisfying the wave equation, and as a result the boundary condition on the random surface is transformed into the hierarchy of equations for the expansion coefficients, so-called Wiener kernels, which can be solved approximately assuming the roughness to be small. The statistical quantities of the random wave field involving the effect of multiple scattering, such as the coherent amplitude (average part) and the angular distribution or the differential cross section for the incoherent scattering (random part), can be easily calculated from the stochastic functional solution so obtained. Notice that the divergence difficulty arising in the perturbation theory[7] can be circumvented in the present method without recourse to the renormalization technique (e.g., the scalar wave with Neumann surface,[31] electromagnetic wave with perfectly conducting surface,[17] and surface plasmon mode of silver film.[32] A treatment based on the WH functional calculus was recently given by Ref. 33.

Although the scattering by a random sphere has been treated in several ways by assuming a suitable model for a random spherical surface,[1–3,34] no work has yet been made of the theoretical formulation as a spherical scattering problem just like the well-known scattering theory for a nonrandom sphere;[35] it is perhaps because of the lack of techniques to handle a random field on the sphere and partly because of the complex manipulation of spherical functions in the perturbation calculus. In the present paper we intend to develop such a formulation based on the stochastic functional approach, which is particularly essential when we deal with the Mie scattering for a random sphere. However, before going into the three-dimensional (3-D) scattering by a random sphere, we start with the simpler problem of 2-D scattering; that is, the scattering by a random circular surface in 2-D space. Though such a problem may not be practical at the moment, it will serve as a preliminary model for a more complicated spherical problem: an experimental work is reported of the light scattering by a metallic wire with random cylindrical surface.[36] The circular problem is theoretically much simpler than the spherical one without the knowledge of the rotation group,[37] and is easily comprehensible because of the similarity of various theoretical formulas between the cases of random planar surface[15,16] and the random circular surface in the present paper. In the 2-D analysis, the random circular surface is assumed to be a Gaussian random field

 0022-2488/88/040851-10$02.50

statistically homogeneous with respect to circular rotations, and the random wave field that undergoes the transformation as an irreducible representation of the rotational motions gives the stochastic analog of the Floquet theorem. Therefore, we first solve the scattering problem for the cylindrical wave injection corresponding to an irreducible representation and then obtain the wave field for plane-wave injection by superposing the stochastic cylindrical solutions; this implies the decomposition into irreducible components. The stochastic Floquet solution, as a functional of the random surface, is again expanded in terms of the WH functionals, and the boundary condition on the circular random surface is solved to yield the expansion coefficients. From the stochastic solution the statistical properties of the scattering are easily calculated, such as the differential cross sections for the coherent and incoherent scatterings, and the stochastic version of the optical theorem which gives the relationship between the total cross section and the coherent amplitude for the forward scattering.

At this point it should be noted that there is some similarity between the two cases for the planar random surface and the circular one: this is basically a result of the fact that the underlying groups, i.e., the translational motions on the plane and the rotational motions of the circle, are both commutative additive groups and that the irreducible representations are given by 1-D exponential functions,[38] which are trivially easy for multiplicative calculations. In the case of a random sphere, however, we have to deal with the noncommutative rotation group associated with the rotational homogeneity of the random field on the sphere, where the multiplication of the irreducible components involves a more intricate calculus.[37,38] It will be shown in a succeeding paper that the representation theory of the rotation group plays an important role in the theoretical formulation of the stochastic scattering by a random sphere, which is to be compared with the 2-D case in the present paper. The theory and the results in the present paper, therefore, will be very helpful for the understanding of the more complex 3-D formulation and for the comparison of the results as well.

II. HOMOGENEOUS RANDOM FIELD ON A CIRCLE

The formulas and the arguments in the following are quite analogous to those in the planar case[15,16] (see also Refs. 19 and 20). Let the homogeneous random field on a circle S_2 $(0\leqslant\theta<2\pi)$ be denoted by

$$r=f(\theta,\omega),\quad \langle f(\theta,\omega)\rangle=0, \tag{1}$$

where $\omega\in\Omega$ indicates the probability parameter denoting the sample point in the sample space Ω and $\langle\ \rangle$ denotes the probabilistic average over Ω. For a strictly homogeneous random field on the circle, we can define the group of measure-preserving transformations T^θ (θ: mod 2π) on Ω such that

$$f(\theta,\omega)=f(0,T^\theta\omega)\equiv f(T^\theta\omega),\quad f(\omega)\equiv f(0,\omega), \tag{2}$$

where the transformations T^α $(0\leqslant\alpha<2\pi)$ have the properties

$$T^{\alpha_1}T^{\alpha_2}=T^{\alpha_1+\alpha_2},\quad [T^\alpha]^{-1}=T^{-\alpha},$$
$$T^0=1\quad (\alpha:\ \mathrm{mod}\ 2\pi), \tag{3}$$

which gives a representation of the additive group on a circle S_2, namely, the group of rotational motions.

Let $\psi(\theta,\omega)$ be a random field on the circle generated by the homogeneous random field (1) and define the shift transformation of the random field $\psi(\theta,\omega)$ by

$$D^\alpha\psi(\theta,\omega)\equiv\psi(\theta+\alpha,T^{-\alpha}\omega),\quad \alpha:\ \mathrm{mod}\ 2\pi, \tag{4}$$

$$D^{\alpha_1}D^{\alpha_2}=D^{\alpha_1+\alpha_2},\quad [D^\alpha]^{-1}=D^{-\alpha},$$
$$D^0=1,\quad \alpha:\ \mathrm{mod}\ 2\pi, \tag{5}$$

where (5) again gives a representation of the rotation group on S_2 similar to (3).

Now a D^α-invariant random field $u(\theta,\omega)$ such that

$$D^\alpha u(\theta,\omega)=u(\theta+\alpha,T^{-\alpha}\omega)=u(\theta,\omega) \tag{6}$$

is a homogeneous random field on S_2; namely,

$$u(\theta,\omega)=u(T^\theta\omega),\quad u(\omega)\equiv u(0,\omega), \tag{7}$$

which is of the same form with (2). The random field $\phi_m(\theta,\omega)$, which gives the 1-D representation (group index $e^{im\alpha}$) of the transformations D^α such that

$$D^\alpha\phi_m(\theta,\omega)=\phi_m(\theta+\alpha,T^{-\alpha}\omega)$$
$$=e^{im\alpha}\phi_m(\theta,\omega),\ \mathrm{mod}\ 2\pi, \tag{8}$$

can be written in the form

$$\phi_m(\theta,\omega)=e^{im\theta}u_m(T^\theta\omega),\quad m=0,\pm1,\pm2,\ldots. \tag{9}$$

Generally speaking a random field on the cirlce $\psi(\theta,\omega)$ can be decomposed into the sum of the random fields of the form (9);

$$\psi(\theta,\omega)=\sum_{m=-\infty}^{\infty}e^{im\theta}u_m(T^\theta\omega), \tag{10}$$

where the homogeneous random field $u_m(T^\omega\omega)$ can be given by

$$u_m(T^\theta\omega)=\frac{1}{2\pi}\int_0^{2\pi}e^{-im\alpha}D_\alpha\psi(\theta,\omega)d\alpha. \tag{11}$$

The homogeneous random field (2) with zero mean has the following spectral representation:

$$f(\theta,\omega)\equiv f(T^\theta\omega)=\sum_{n=-\infty}^{\infty}e^{in\theta}F_nB_n(\omega), \tag{12}$$

where $B_n(\omega)$, $n=$ integer, denotes a set of normalized orthogonal random variables with zero means:

$$\langle B_n(\omega)\rangle=0,$$
$$\langle\overline{B_n(\omega)}B_m(\omega)\rangle=\delta_{nm},\quad n,m=0,\pm1,\pm2,\ldots, \tag{13}$$

the overbar indicating the complex conjugate quantity. It should be noted that $\overline{B}_n=B_{-n}$, and $\overline{F}_n=F_{-n}$ because of real-valued f. The spectral representation of the correlation function readily follows from (12) and (13);

$$R(\theta)=\langle f(\omega)f(T^\theta\omega)\rangle=\sum_{n=-\infty}^{\infty}e^{in\theta}|F_n|^2, \tag{14}$$

so that $|F_n|^2$ corresponds to the so-called power spectrum, which we will call it also in the present case. The orthogonality (13) is a consequence of the homogeneity of the random field, and can be derived also from the random variable $F_nB_n(\omega)$ given by the Fourier transform:

$$F_n B_n(\omega) = \frac{1}{2\pi}\int_0^{2\pi} e^{-in\theta} f(T^\theta \omega)\,d\theta. \tag{15}$$

In what follows we assume $f(x,\omega)$ to be a Gaussian random field so that B_n's are a set of independent complex Gaussian random variables with zero means, that is, a complex sequence of Gaussian white noise. By (15) it is shown that the random variable $B_n(\omega)$ is transformed under D^α or T^α like

$$D^\alpha B_n(\omega) = B_n(T^{-\alpha}\omega) = e^{-in\alpha} B_n(\omega), \tag{16}$$

that is, $B_n(\omega)$ is an eigenvector of D^α with the eigenvalue $e^{-in\alpha}$.

III. SCATTERING BY A CIRCULAR CYLINDER WITH HOMOGENEOUS RANDOM SURFACE

Homogeneous random cylindrical surface: Let the cylindrical coordinates in a 2-D space be denoted by (r,θ) and a homogeneous random field on a circle with radius a be

$$r = a + f(T^\theta \omega), \quad \langle f(T^\theta \omega)\rangle = 0, \tag{17}$$

where $f(T^\theta\omega)$ denotes a homogeneous Gaussian random field on the circle with the spectral representation given by (12)–(14). We regard the variance of the random surface, i.e.,

$$\sigma^2 \equiv R(0) = \langle |f(\omega)|^2\rangle = \sum_{n=-\infty}^{\infty} |F_n|^2, \tag{18}$$

as the parameter describing the surface roughness.

Wave equation and the boundary condition: Let the 2-D Helmholtz equation for the random wave field $\psi(kr,\theta,\omega)$ be given by

$$(\nabla^2 + k^2)\psi(kr,\theta,\omega) = 0, \tag{19}$$

k being the wave number. We consider the two kinds of homogeneous boundary conditions on the circular random surface defined by (17):

$$\psi(kr,\theta,\omega) = 0; \quad r = a + f(\theta,\omega) \quad \text{(Dirichlet)}, \tag{20}$$

$$\frac{\partial \psi(kr,\theta,\omega)}{\partial n} = 0; \quad r = a + f(\theta,\omega) \quad \text{(Neumann)}, \tag{21}$$

where $\partial/\partial n$ denotes the normal derivative on the random surface. The boundary condition on the random surface (20) or (21) can be represented by the approximated boundary condition on the surface $r = a$ as follows:

$$\left[\psi + f\frac{\partial \psi}{\partial r}\right]_{r=a} = 0 \quad \text{(Dirichlet)}, \tag{22}$$

$$\left[\frac{\partial \psi}{\partial r} - \frac{1}{r^2}\frac{\partial f}{\partial \theta}\frac{\partial \psi}{\partial \theta} + f\frac{\partial^2 \psi}{\partial r^2}\right]_{r=a} = 0 \quad \text{(Neumann)}, \tag{23}$$

which we will use as a model of the random boundary condition on the circle.

Cylindrical-wave expansion of a plane wave: The plane wave with the wave vector $\mathbf{k} = (k,\alpha)$ can be expanded into cylindrical waves by the formula

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_{m=-\infty}^{\infty} i^m J_m(kr) e^{im(\theta-\alpha)}, \tag{24}$$

$\mathbf{r} = (r,\theta)$ denoting the position vector, and $J_m(z)$ the Bessel function. Therefore, the wave solution for the plane-wave incidence can be given by superposing the solutions for the injection of cylindrical waves $J_m(kr)e^{im\theta}$, $m =$ integer.

Unperturbed field (primary wave): In the nonrandom case with $\sigma^2 = 0$ (smooth cylinder), as well known, the unperturbed wave field for the cylindrical wave injection can be given by

$$\psi_m^0(kr,\theta) = [J_m(kr) + \alpha_m^0 H_m^{(1)}(kr)]e^{im\theta} \tag{25}$$

$$= \tfrac{1}{2}[H_m^{(2)}(kr) + e^{2i\delta_m} H_m^{(1)}(kr)]e^{im\theta}, \tag{26}$$

$$\alpha_m^0 \equiv i e^{i\delta_m}\sin\delta_m$$

$$= -J_m(ka)/H_m^{(1)}(ka) \quad \text{(Dirichlet)} \tag{27}$$

$$= -\dot{J}_m(ka)/\dot{H}_m^{(1)}(ka) \quad \text{(Neumann)}, \tag{28}$$

where the overdot indicates the differentiation; $\dot{J}_m(z) = dJ_m(z)/dz$, and $H_m^{(1)}(z)$ denotes the Hankel function of the first kind. The first and the second term of (26) gives the incident cylindrical wave and the scattered wave, respectively, and δ_m denotes the phase shift defined by (27) and (28).

It should be noticed that the cylindrical wave $\psi_m^0(kr,\theta)$ (independent of ω) is transformed under the shift (rotation) D^α as

$$D^\alpha \psi_m^0(kr,\theta) = \psi_m^0(kr,\theta+\alpha) = e^{im\alpha}\psi_m^0(kr,\theta), \tag{29}$$

that is, it is an eigenfunction (basis of an irreducible representation) with the eigenvalue (group index) $e^{im\alpha}$ (Ref. 38).

Perturbed wave field (secondary wave): In the case of random surface ($\sigma^2 > 0$), let the total wave field for cylindrical wave injection be

$$\psi_m(kr,\theta,\omega) = \psi_m^0(kr,\theta) + \psi_m^s(kr,\theta,\omega). \tag{30}$$

We note here that the wave equation (19) and the boundary condition (20), or (21), associated with the homogeneous random surface (2) are invariant under the operation D^α because of (6), so that the wave solution (30) and therefore the perturbed field ψ_m^s undergoes the same transformation as (29); which means the transformation (8),

$$D^\alpha \psi_m^s(kr,\theta,\omega) \equiv \psi_m^s(kr,\theta+\alpha,T^{-\alpha}\omega)$$

$$= e^{im\alpha}\psi_m^s(kr,\theta,\omega). \tag{31}$$

Wiener–Hermite expansion of the perturbed field: The perturbed wave field is a nonlinear functional of the Gaussian random surface represented as (12), and therefore, is expressible as a functional of the complex Guassian random sequence $B_n(\omega)$. We represent it as a WH expansion in the form

$$\psi_m^s(kr,\theta,\omega) = \sum_{n=0}^{\infty}\sum_{j_1=-\infty}^{\infty}\cdots\sum_{j_n=-\infty}^{\infty} A_m^n(j_1,\ldots,j_n) H_{m+j_1+\cdots+j_n}^{(1)}(kr) e^{i(m+j_1+\cdots+j_n)\theta}\hat{H}_n(B_{j_1},\ldots,B_{j_n})$$

$$= A_m^0 H_m^{(1)}(kr)e^{im\theta} + \sum_{j=-\infty}^{\infty} A_m^1(j) H_{m+j}^{(1)}(kr) e^{i(m+j)\theta}\hat{H}_1(B_j)$$

$$+ \sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty} A_m^2(j_1,j_2) H_{m+j_1+j_2}^{(1)}(kr) e^{i(m+j_1+j_2)\theta}\hat{H}_2(B_{j_1},B_{j_2}) + \cdots, \tag{32}$$

where $A^n_m(j_1,...,j_n)$, which we call the WH coefficient, denotes the n-variate coefficient symmetric with respect to its n arguments, and $\hat{H}_n(x_{j_1},...,x_{j_n})$ is the n-variate complex Hermite polynomial, some lowest-order polynomials of which are

$$\hat{H}_0 = 1, \quad \hat{H}_1(x) = x, \quad \hat{H}_2(x_{j_1},x_{j_2}) = x_{j_1}x_{j_2} - \delta_{j_1,-j_2},... . \tag{33}$$

[For the WH expansion and multivariate Hermite polynomials, see the Appendices of Refs. 15, 16, and 19, and the references therein. Equation (32) corresponds to a discrete version of the WH expansion. For more details, see Ref. 39.] The first term of the expansion (32) gives the average (coherent) part and the second term, and the following are the random (incoherent) part with zero average which we denote by ψ^{ic} in the following. In view of the transformation rule of $\hat{H}_n(B_{j_1},...,B_{j_n})$ under D^α, that is,

$$D^\alpha \hat{H}_n(B_{j_1},...,B_{j_n}) = e^{-i(j_1+\cdots+j_n)\alpha}\hat{H}_n(B_{j_1},...,B_{j_n}), \tag{34}$$

we note that the nth-order term of the WH expansion (32), i.e.,

$$\phi_m(kr,\theta,\omega) \equiv H^{(1)}_{m+j_1+\cdots+j_n}(kr)e^{(m+j_1+\cdots+j_n)\theta} \times \hat{H}_n(B_{j_1},...,B_{j_n}), \tag{35}$$

which is a cylindrical wave satisfying the wave equation (19), obeys the same transformation rule as (8) or (31):

$$D^\alpha \phi^n_m(kr,\theta,\omega) = e^{im\alpha}\phi^n_m(kr,\theta,\omega). \tag{36}$$

Scattered wave field for cylindrical wave injection: To summarize, the total wave field ψ_m for the cylindrical wave injection can be written

$$\psi_m(kr,\theta,\omega) = \psi^0_m(kr,\theta) + \psi^s_m(kr,\theta,\omega) \tag{37}$$

$$= \psi^c_m(kr,\theta) + \psi^{ic}_m(kr,\theta,\omega), \quad \langle \psi^{ic}_m \rangle = 0, \tag{38}$$

where

$$\psi^c_m(kr,\theta) = [J_m(kr) + \alpha_m H^{(1)}_m(kr)]e^{im\theta}, \tag{39}$$

$$\alpha_m \equiv \alpha^0_m + A^0_m, \tag{40}$$

$$\psi^{ic}_m(kr,\theta,\omega) \equiv \sum_{n=1}^{\infty} \sum_{j_1} \cdots \sum_{j_n} A^n_m(j_1,...,j_n)\phi^n_m(kr,\theta,\omega), \tag{41}$$

ψ^c_m gives the coherent part and ψ^{ic}_m the incoherent part for cylindrical wave injection. We call α_m the coherent scattering coefficient.

Scattered wave field for plane-wave injection: The total wave field for plane-wave incidence with the wave vector $k = (k,\alpha)$ can be obtained by superposing ψ_m in m according to (24), which implies the decomposition (10). Corresponding to the equations (37)–(41), we have

$$\psi(kr,\theta,\omega) = \psi^0(kr,\theta) + \psi^s(kr,\theta,\omega) \tag{42}$$

$$\equiv \psi^c(kr,\theta) + \psi^{ic}(kr,\theta,\omega), \quad \langle \psi^{ic} \rangle = 0, \tag{43}$$

where

$$\psi^c(kr,\theta) \equiv \sum_{m=-\infty}^{\infty} i^m \psi^c_m(kr,\theta)e^{-im\alpha} \tag{44}$$

$$= \sum_{m=-\infty}^{\infty} i^m[J_m(kr) + a_m H^{(1)}_m(kr)]e^{im(\theta-\alpha)}, \tag{45}$$

$$\psi^{ic}(kr,\theta,\omega) = \sum_{m=-\infty}^{\infty} i^m \psi^{ic}_m(kr,\theta,\omega)e^{-im\alpha} \tag{46}$$

$$= \sum_{m=-\infty}^{\infty} i^m e^{-im\alpha}\Big[\sum_j A^1_m(j)H^{(1)}_{m+j}(kr) \times e^{i(m+j)\theta}\hat{H}_1(B_j) + \sum_{j_1}\sum_{j_2} A^2_m(j_1,j_2)H^{(1)}_{m+j_1+j_2}(kr) \times e^{i(m+j_1+j_2)\theta}\hat{H}_2(B_{j_1},B_{j_2}) + \cdots\Big]. \tag{47}$$

IV. STATISTICAL CHARACTERISTICS OF THE SCATTERED WAVE FIELD

Coherent scattering amplitude: Without loss of generality we assume that the direction of incidence of the plane wave is in the z axis, that is, $\alpha = 0$. The asymptotic form at $r \to \infty$ of the coherent scattered wave $\psi^c_s(kr,\theta)$, which is given by the second term of (45) without the incident wave, is given by

$$\psi^c_s(kr,\theta) = \sum_{m=-\infty}^{\infty} i^m \alpha_m H^{(1)}_m(kr)e^{im\theta} \tag{48}$$

$$\sim \sqrt{\frac{2}{\pi kr}}\, e^{i(kr-(1/4)\pi)} \sum_{m=-\infty}^{\infty} \alpha_m e^{im\theta} \equiv \frac{e^{ikr}}{\sqrt{r}}\Phi(\theta), \tag{49}$$

$$\Phi(\theta) = \sqrt{\frac{2}{k\pi i}} \sum_{m=-\infty}^{\infty} \alpha_m e^{im\theta}, \quad \alpha_m \equiv \alpha^0_m + A^0_m, \tag{50}$$

where $\Phi(\theta)$ gives the coherent scattering amplitude.

Total coherent power flow: The total power flow due to ψ^c, which is integrated over a surrounding circle with radius r, can be evaluated at $r \to \infty$;

$$S_c = \lim_{r\to\infty} \frac{r}{k}\int_0^{2\pi} \mathrm{Im}\Big[\overline{\psi^c}\,\frac{\partial \psi^c}{\partial r}\Big]d\theta$$

$$= \frac{1}{r}\sum_{m=-\infty}^{\infty} (-1 + |1 + 2\alpha_m|^2) \tag{51}$$

$$= \frac{4}{k}\sum_{m=-\infty}^{\infty} (\mathrm{Re}\,\alpha_m + |\alpha_m|^2). \tag{52}$$

Total incoherent power: The total power flow due to ψ^{ic}, integrated over a surrounding circle, is shown to be constant with probability 1, so that it equals to its average, which can be similarly evaluated at infinity:

$$\sigma_{ic} \equiv \lim_{r\to\infty} \frac{r}{k}\int_0^{2\pi} \mathrm{Im}\Big\langle \overline{\psi^{ic}}\,\frac{\partial \psi^{ic}}{\partial r}\Big\rangle d\theta \tag{53}$$

$$\equiv \frac{4}{k}\sum_{m=-\infty}^{\infty}\sum_{n=1}^{\infty} n! \sum_{j_1}\cdots\sum_{j_n} |A^n_m(j_1,...,j_n)|^2. \tag{54}$$

Power flow conservation law: If we apply the Gauss theorem to the circular region with radius r surrounding the random circle, we obtain the conservation law of the total power flow $S_c + \sigma_{ic} = 0$ with probability 1 (regardless of the random surface), that is,

$$\frac{4}{k}\sum_{m=-\infty}^{\infty}\Big[\mathrm{Re}\,\alpha_m + |\alpha_m|^2 + \sum_{n=1}^{\infty} n! \sum_{j_1}\cdots\sum_{j_n} |A^n_m(j_1,\ldots,j_n)|^2\Big] = 0. \tag{55}$$

Total scattering cross section: The second term in the bracket of (55) gives the power flow σ_c due to the coherent scattered wave $\psi^c_s(\mathrm{kr},\theta)$. Hence, the sum of σ_c and σ_{ic} gives the total scattering cross section:

$$S = \sigma_c + \sigma_{ic} = \frac{4}{k}\sum_{m=-\infty}^{\infty}\Big[|\alpha_m|^2 + \sum_{n=1}^{\infty} n! \sum_{j_1}\cdots\sum_{j_n} |A^n_m(j_1,\ldots,j_n)|^2\Big]. \tag{56}$$

Optical theorem: In view of (50) and (56), the power-flow conservation law (55) can be rewritten in the following form:

$$S = \sqrt{8\pi/k}\,\mathrm{Im}(\Phi(0)/\sqrt{i}) \tag{57}$$

that is, the total scattering cross section is given in terms of the imaginary part of the coherent foward-scattering amplitude; this is an extended version of the optical theorem to a random surface scattering.

Angular distribution of coherent scattering (differential cross section for coherent scattering): From (49) we obtain

$$\sigma_c(\theta)d\theta = |\Phi(\theta)|^2 d\theta = \frac{2}{\pi k}\Big|\sum_{m=-\infty}^{\infty}\alpha_m e^{im\theta}\Big|^2 d\theta. \tag{58}$$

Angular distribution of incoherent scattering (differential cross section for incoherent scattering): From (35), (47), and (53), we obtain

$$\sigma_{ic}(\theta)d\theta = \frac{2}{\pi k}\sum_{n=1}^{\infty} n! \sum_{j_1}\cdots\sum_{j_1} \times \Big|\sum_{m=-\infty}^{\infty} A^n_m(j_1,\ldots,j_n)e^{im\theta}\Big|^2 d\theta. \tag{59}$$

Equations (58) and (59), when integrated, give σ_c and σ_{ic}, respectively.

Power conservation law for cylindrical wave injection: It is shown that the power conservation law for the plane-wave incidence (55) does hold termwise for each m, corresponding to the conservation law for the cylindrical wave injection. We express this in the form of (51):

$$1 = |1 + 2\alpha_m|^2 + 4\sum_{n=1}^{\infty} n! \sum_{j_1}\cdots\sum_{j_n} |A^n_m(j_1,\ldots,j_n)|^2, \tag{60}$$

where the left-hand side gives the total incident power of the cylindrical wave, whereas the first term in the right-hand side represents the total power of the coherent scattering P_c, and the second term including the summation gives the total power of the incoherent scattering P_{ic}, so that (60) can be expressed as $1 = P_c + P_{ic}$. The power conservation law (60) as well as the optical theorem (57) could be used to check the validity range of approximate solutions.

V. METHOD OF APPROXIMATE SOLUTION FOR THE BOUNDARY CONDITION

Dirichlet condition: The boundary condition (22), that is,

$$\Big[\psi_m + f\frac{\partial\psi_m}{\partial r}\Big]_{r=a} = 0, \tag{61}$$

can be transformed into the hierarchical set of equations for determining the WH coefficients $A^n_m(j_1,\ldots,j_n)$, if we substitute (12) for f and (37)–(41) for ψ_m followed by using the recurrence formula and the orthogonality relations for complex multivariate Hermite polynomials;

$$B_j\hat{H}_n(B_{j_1},B_{j_2},\ldots,B_{j_n}) = \hat{H}_{n+1}(B_j,B_{j_1},\ldots,B_{j_n}) + \sum_{k=1}^{n}\hat{H}_{n-1}(B_{j_1},\ldots,B_{j_{k-1}},B_{j_{k+1}},\ldots,B_{j_n})\delta_{j-j_k}, \tag{62}$$

$$\langle\overline{\hat{H}_n(B_{i_1},\ldots,B_{i_n})}\hat{H}_m(B_{j_1},\ldots,B_{j_m})\rangle = \delta_{nm}\,\delta^{(n)}_{ij}, \tag{63}$$

$$\delta^{(n)}_{ij} \equiv \sum_{\text{all pair}}\prod_{(\nu,\mu)}\delta_{i_\nu j_\mu}, \tag{64}$$

where δ_{mn} denotes the Kronecker delta, and the sum in (64) is to be taken over all distinct products of n Kronecker deltas $\delta_{i_\nu j_\mu}$, i_ν and j_μ being taken from $i \equiv (i_1,\ldots,i_n)$ and $j \equiv (j_1,\ldots,j_n)$, respectively; hence (64) contains $n!$ terms. Omitting the details of the calculation, we show the first three equations for the WH coefficients:

$$H_n(ka)A^0_m + \sum_{j=-\infty}^{\infty} k\,\overline{F_j}\dot{H}_{m+j}(ka)A^1_m(j) = 0 \qquad (n=0), \tag{65}$$

$$H_{m+j_1}(ka)A^1_m(j_1) + kF_{j_1}[\dot{J}_m(ka) + \alpha_m\dot{H}_m(ka)] + 2\sum_{j=-\infty}^{\infty} k\,\overline{F_j}\dot{H}_{m+j_1+j}(ka)A^1_m(j_1,j) = 0 \qquad (n=1), \tag{66}$$

$$\begin{aligned}&H_{m+j_1+j_2}(ka)A^2_m(j_1,j_2) + \tfrac{1}{2}[k\,\overline{F_{j_1}}\dot{H}_{m+j_2}(ka)A^1_m(j_2) + k\,\overline{F_{j_2}}\dot{H}_{m+j_1}(ka)A^1_m(j_1)]\\ &\quad + 3\sum_{j=-\infty}^{\infty} k\,\overline{F_j}\dot{H}_{m+j_1+j_2+j}(ka)A^3_m(j_1,j_2,j) = 0 \qquad (n=2),\end{aligned} \tag{67}$$

where we have used the abbreviation $H_m(kr) \equiv H^{(1)}_m(kr)$. The first equation (65) for $n=0$ corresponds to the average part of (61). We assume the small roughness, i.e., $k\sigma \ll 1$. Since kF_j is of the order of $k\sigma$, A^0_m and A^n_m $(n \geqslant 1)$ are of the order of $(k\sigma)^2$ and $(k\sigma)^n$, respectively. Therefore, neglecting A^3_m in (67), we can easily obtain an approximate solution for A^2_m. Substituting this into (66) and neglecting terms of the order $(k\sigma)^3$ compared to $k\sigma$, we obtain an approximate solution for A^1_m:

$$A^1_m(j_1) \cong -\frac{kF_{j_1}(ka)[\dot{J}_m(ka) + \alpha_m\dot{H}_m(ka)]}{[1 + M(m+j_1)]H_{m+j_1}(ka)}, \tag{68}$$

$$M(m+j_1) \equiv -\frac{\dot{H}_{m+j_1}(ka)}{H_{m+j_1}(ka)}\sum_j |kF_j|^2\frac{\dot{H}_{m+j_1+j}(ka)}{H_{m+j_1+j}(ka)}. \tag{69}$$

If necessary, the approximate solution (68) can be corrected taking account of the discarded terms. Thus $M(m+j)$, given by (69), which corresponds to the mass operator in the renormalization theory, gives the term of the order of $(k\sigma)^2$. Equation (69) for $M(m+j)$ can be slightly corrected if we take A_m^3 into account; the corrected equation is shown below.

The mass operator representing the effect of multiple scattering plays an important role to suppress the divergence in the case of a random planar surface with Neumann condition[31] or in the presence of a surface wave mode.[32] However, in the case of circular surface, we never have $H_{m+j_1}(ka)=0$ in (68) (no resonance for real m and ka), so that when $(k\sigma)^2$ is sufficiently small enough we can neglect $M(m+j)\ll 1$ in (68). Furthermore, in the numerator of (68), the coherent scattering coefficient

$$\alpha_m=\alpha_m^0+A_m^0=-J_m(ka)/H_m(ka)+A_m^0 \tag{70}$$

involves A_m^0 of the order of $(k\sigma)^2$, which gives rise to the term of the order of $(k\sigma)^3$ in (68). Therefore, neglecting $M(m+j)$ and A_m^0 in (68), we get the solution for the single-scattering approximation:

$$A_m^1(j_1)\simeq-\frac{kF_{j_1}[\dot J_m(ka)+\alpha_m^0\dot H_m(ka)]}{H_{m+j_1}(ka)}=\frac{2ikF_{j_1}}{\pi kaH_m(ka)H_{m+j_1}(ka)}, \tag{71}$$

where the right-hand equality is due to the Lommel formula.

Now, substituting (68) and (70) into (65), we obtain

$$A_m^0=\frac{2i}{\pi ka\dot H_m(ka)H_m(ka)}\frac{M(m)}{1+M(m)}, \tag{72}$$

where

$$M(m)=-\frac{\dot H_m(ka)}{H_m(ka)}\sum_j|kF_j|^2\frac{\dot H_{m+j}(ka)}{H_{m+j}(ka)}\times\frac{1}{1+M(m+j)} \tag{73}$$

gives an equation for the mass operator, which is reduced to Eq. (69) when $M(m+j)$ is neglected in the right-hand side. If we further neglect $M(m)$ in the denominator of (72), we get the single-scattering approximation

$$A_m^0\cong 2iM(m)/\pi ka\dot H_m(ka)H_m(ka). \tag{74}$$

Neumann condition: In the same manner as above we can solve the boundary condition (23);

$$\left[\frac{\partial\psi_m}{\partial r}-\frac{1}{r^2}\frac{\partial f}{\partial\theta}\frac{\partial\psi_m}{\partial\theta}+f\frac{\partial^2\psi_m}{\partial r^2}\right]_{r=a}=0. \tag{75}$$

Substituting (12), (30), and (32) into (75), we obtain a set of equations for WH coefficients. The first three equations are

$$k\dot H_m(ka)A_m^0-\sum_{j=-\infty}^{\infty}\overline{F}_j[(1/a^2)j(m+j)H_{m+j}(ka)-k^2\ddot H_{m+j}(ka)]A_m(j)=0\quad(n=0), \tag{76}$$

$$\begin{aligned}&k\dot H_{m+j}(ka)A_m(j_1)+F_{j_1}\left\{\frac{j_1m}{a^2}[J_m(ka)+\alpha_mH_m(ka)]+k^2[\ddot J_m(ka)+\alpha_m\ddot H_m(ka)]\right\}\\&-2\sum_{j=-\infty}^{\infty}\overline{F}_j\left[\frac{j(m+j_1+j)}{a^2}H_{m+j_1+j}(ka)-k^2\ddot H_{m+j_1}+j(ka)\right]A_m^2(j_1,j)=0\quad(n=1),\end{aligned} \tag{77}$$

$$\begin{aligned}&k\dot H_{m+j_1+j_2}(ka)A_m^2(j_1,j_2)+\frac{1}{2}F_{j_1}\left[\frac{j_1(m+j_2)}{a^2}H_{m+j_2}(ka)+k^2\ddot H_{m+j_2}(ka)\right]A_m^1(j_2)\\&+\frac{1}{2}F_{j_2}\left[\frac{j_2(m+j_1)}{a^2}H_{m+j_1}(ka)+k^2\ddot H_{m+j_1}(ka)\right]A_m^1(j_1)\\&-\sum_j\overline{F}_j\left[\frac{j(m+j_1+j_2+j)}{a^2}H_{m+j_1+j_2+j}(ka)-k^2\ddot H_{m+j+j_1+j_2}(ka)\right]A_m^3(j_1,j_2,j)=0\quad(n=2),\end{aligned} \tag{78}$$

where

$$\alpha_m=\alpha_m^0+A_m^0=-\dot J_m(ka)/\dot H_m(ka)+A_m^0. \tag{79}$$

Now we substitute (78) with $A_m^3=0$ into (77) and use the Lommel formulas as to obtain an approximation solution for A_m^1:

$$A_m^1(j_1)=\frac{-2iF_{j_1}(j_1m+m^2-k^2a^2)}{\pi a(ka)^2\dot H_m(ka)\dot H_{m+j_1}(ka)[1+M(m+j_1)]}-\frac{F_{j_1}[j_1mH_m(ka)+k^2a^2\ddot H_m(ka)]}{ka^2H_{m+j_1}(ka)[1+M(m+j_1)]}A_m^0, \tag{80}$$

and similarly, substituting this into (76), we obtain

$$A_m^0=\frac{1}{1+M(m)}\frac{-2i}{\pi(ka)^5[\dot H_m(ka)]^2}\sum_j\frac{|kF_j|^2}{\dot H_{m+j}(ka)}\frac{[jm+m^2-k^2a^2][j(m+j)H_{m+j}(ka)-k^2a^2\ddot H_{m+j}(ka)]}{1+M(m+j)}, \tag{81}$$

where the mass operator equation is given by

$$M(m)=\sum_j\frac{|kF_j|^2[j(m+j)H_{m+j}(ka)-k^2a^2\ddot H_{m+j}(ka)][jmH_m(ka)+k^2a^2\ddot H_m(ka)]}{(ka)^4\dot H_m(ka)\dot H_{m+j}(ka)[1+M(m+j)]}, \tag{82}$$

which can be approximately evaluated, by neglecting $M(m+j)$ in the right-hand side, or by iterating substitutions for better evaluation.

Particularly, in the single-scattering approximation to neglect M in (80) and (81), we obtain

$$A_m^1(j_1) \simeq -\frac{2ikF_{j_1}(j_1 m + m^2 - k^2a^2)}{-(ka)^3\dot{H}_m(ka)\dot{H}_{m+j_1}(ka)}, \quad (83)$$

$$A_m^0 \simeq \frac{-2i}{\pi(ka)^5[\dot{H}_m(ka)]^2} \times \sum_j \frac{|kF_j|^2}{\dot{H}_{m+j}(ka)}[jm + m^2 - k^2a^2] \times [j(m+j)H_{m+j}(ka) - k^2a^2\ddot{H}_{m+j}(ka)]. \quad (84)$$

VI. NUMERICAL EVALUATIONS FOR SCATTERING CHARACTERISTICS

Once the approximate solutions are obtained for zeroth- and first-order WH coefficients A_m^0 and A_m^1 as in (68) and (72) for the Dirichlet condition, and (80) and (81) for the Neumann condition, we can evaluate various scattering characteristics using the formulas (48) to (60) with an appropriate power spectrum for the random surface. As is well known, the cylindrical-wave expansion of the type (45) and (47) is only effective in the Mie scattering range, so that the Mie parameter is chosen to be $ka = 1, 2$ in the following numerical calculation.

Although there is no typical power spectrum known for the random field on a circle, we assume the power spectrum to be of the Gaussian form for the numerical calculation:

$$|F_n|^2 = (\sigma^2/\vartheta)e^{-K^2n^2/2}, \quad n = 0, \pm 1, \pm 2, \ldots, \quad (85)$$

where

$$\vartheta \equiv \vartheta\left(\frac{K^2}{2\pi}\right) = \sum_{n=-\infty}^{\infty} e^{-K^2n^2/2}, \quad (86)$$

$\vartheta(x)$ denoting the theta function. The correlation function is given by

$$R(\theta) = \frac{\sigma^2}{\vartheta}\frac{\sqrt{2\pi}}{K}\sum_{n=-\infty}^{\infty}\exp\left[-\frac{1}{2K^2}(\theta + 2\pi n)^2\right], \quad (87)$$

which can be approximated by a single Gaussian term with $n = 0$ if K is small enough; then the parameter K ($\ll 2\pi$) can be regarded as the correlation distance (rad) on the circle, and we assume $K = 0.2, 0.6$ in the following calculation.

We then first check the validity of the approximate solutions by means of the power flow conservation (60) for the mth cylindrical wave injection:

$$1 = P_c + P_{ic}, \quad (88)$$

where in the present approximation with small roughness, the total coherent power P_c and the total incoherent power P_{ic} can be written as

$$P_c \simeq |1 + 2(\alpha_m^0 + A_m^0)|^2, \quad P_{ic} \simeq 4\sum_j |A_m^1(j)|^2. \quad (89)$$

An example for $m = 0$ is shown in Figs. 1 and 2 for the Dirichlet and Neumann cases, respectively, where $P_c + P_{ic}$ is plotted against the roughness parameter $k\sigma$; the equality

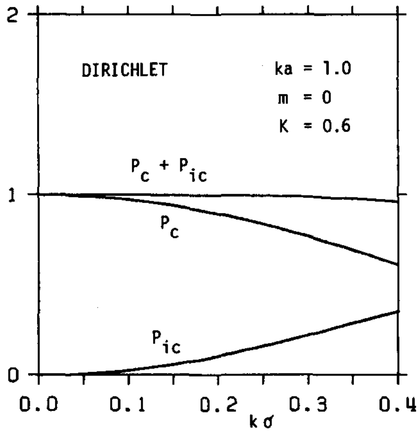


FIG. 1. Power flow conservation for a cylindrical wave incidence (Dirichlet, $m = 0$, $ka = 1$, $K = 0.6$). Here P_c and P_{ic} denote the total coherent power flow and the incoherent power flow, respectively. For a rigorous solution, $P_c + P_{ic} = 1$.

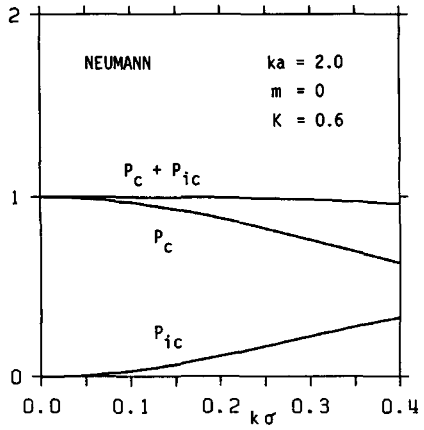


FIG. 2. Power flow conservation for a cylindrical wave incidence (Neumann, $m = 0$, $ka = 1$, $K = 0.6$). Here P_c and P_{ic} denote the total coherent power flow and the incoherent power flow, respectively. For rigorous solution, $P_c + P_{ic} = 1$.

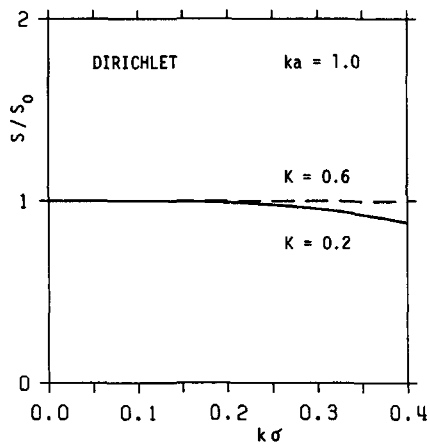


FIG. 3. Optical theorem for a plane-wave incidence (Dirichlet, $ka = 1.0$, $K = 0.2, 0.6$). Here S denotes the total cross section calculated by (56) and S_0 the right-hand side of (57) due to the imaginary part of the forward-scattering amplitude. $S/S_0 = 1$ for rigorous solution.

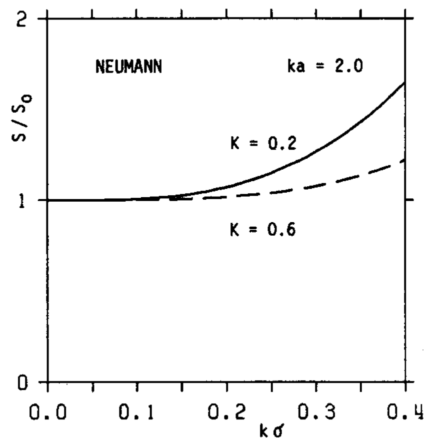


FIG. 4. Optical theorem for plane-wave incidence (Neumann, $ka = 1.0$, $K = 0.2, 0.6$). Here S denotes the total cross section calculated by (56) and S_0 the right-hand side of (57) due to the imaginary part of the forward-scattering amplitude. For rigorous solution, $S/S_0 = 1$.

(88) is nearly satisfied for $k\sigma < 0.2$. It is shown by numerical calculations that $1 - P_c$ and P_{ic} rapidly approach 0 for larger m, so that in the Mie scattering range, the equality (55) consisting of the sum over m does hold for plane-wave incidence. However, to check the power equality in the plane-wave case, we can make use of the optical theorem (57), namely, $S = S_0$, S denoting the total cross section (56) and S_0 the right-hand member of (57) due to the coherent forward-scattering amplitude. Figs. 3 and 4 shows the ratio

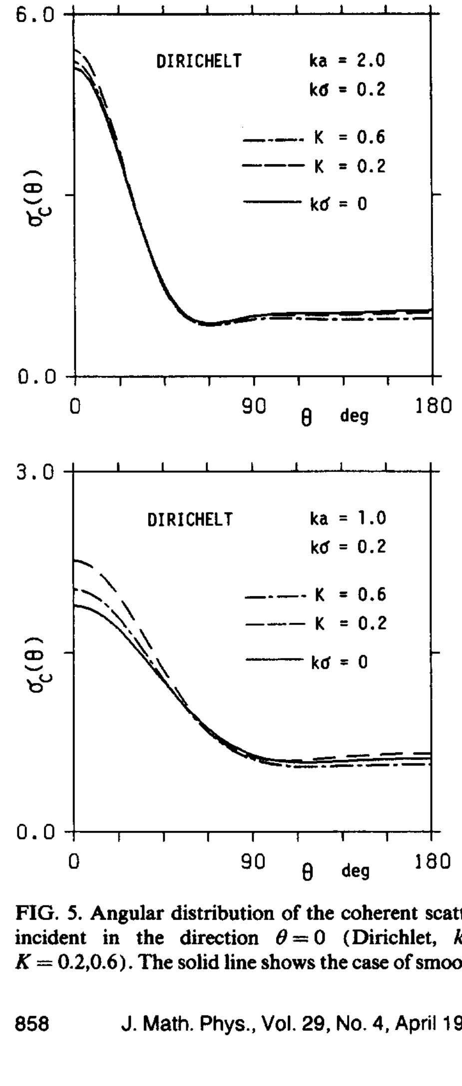


FIG. 5. Angular distribution of the coherent scattering for a plane wave incident in the direction $\theta = 0$ (Dirichlet, $ka = 1.0, 2.0$, $k\sigma = 0.2$, $K = 0.2, 0.6$). The solid line shows the case of smooth surface with $k\sigma = 0$.

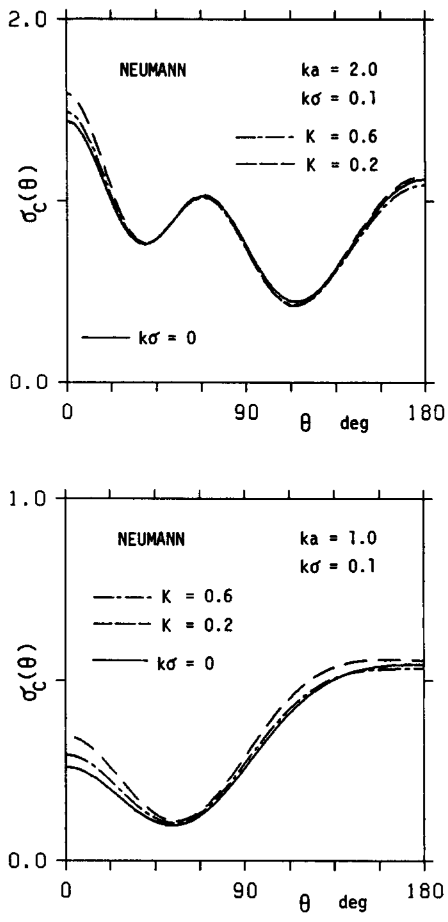


FIG. 6. Angular distribution of the coherent scattering for a plane wave incident in the direction $\theta = 0$ (Neumann, $ka = 1.0, 2.0$, $k\sigma = 0.1$, $K = 0.2, 0.6$). The solid line shows the case of smooth surface with $k\sigma = 0$.

S/S_0 plotted against $k\sigma$, which nearly equals 1 within $k\sigma < 0.2$ for the parameter values shown in the figures.

We are then ready to calculate the angular distributions of coherent scattering (58) for a plane-wave injection, which are shown in Figs. 5 and 6 for the Dirichlet and Neumann case, respectively, with $ka = 1, 2$ and $k\sigma = 0.2$. Correspondingly, the angular distributions of the incoherent scattering calculated by (59) are shown in Figs. 7 and 8, respectively, which show that the incoherent scattering is generally stronger in the backward direction than in the forward in either case.

In the Mie scattering such as $ka = 1$ or 2, the single-scattering approximation actually does not differ appreciably from the results shown if $k\sigma$ is so small. This means that the multiple scattering due to small roughness does not produce an appreciable effect on the Mie scattering because of the absence of real resonance or surface modes on a circular surface. On the other hand, in the case of planar random surface, which should correspond to the limiting case $ka \to \infty$, the multiple scattering has an important effect on the scattering characteristics even if the roughness $k\sigma > 0$ is negligibly small (scalar wave with Neumann surface,[31] electromagnetic wave with perfectly conducting surface,[17] and surface plasmon mode[32]). Therefore, the multiple scattering is expected to become more and more effective as radius of

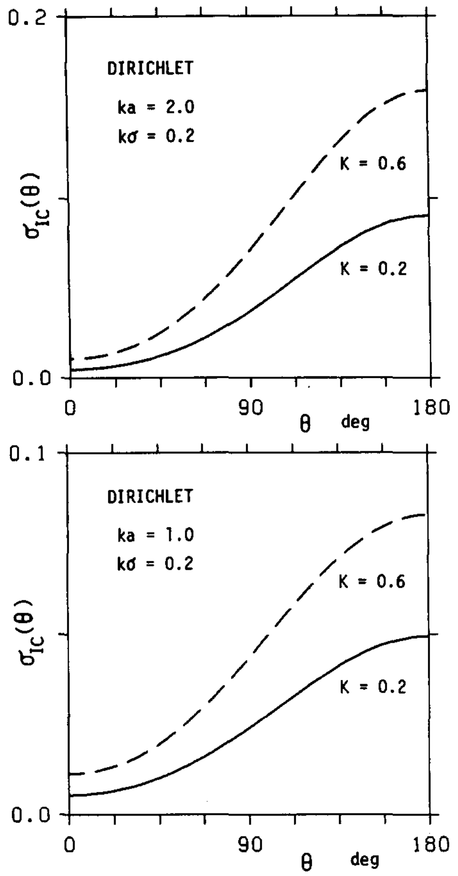


FIG. 7. Angular distribution of the coherent scattering for a plane-wave incidence (Dirichlet, $ka = 1.0, 2.0$, $k\sigma = 0.2$, $K = 0.2, 0.6$).

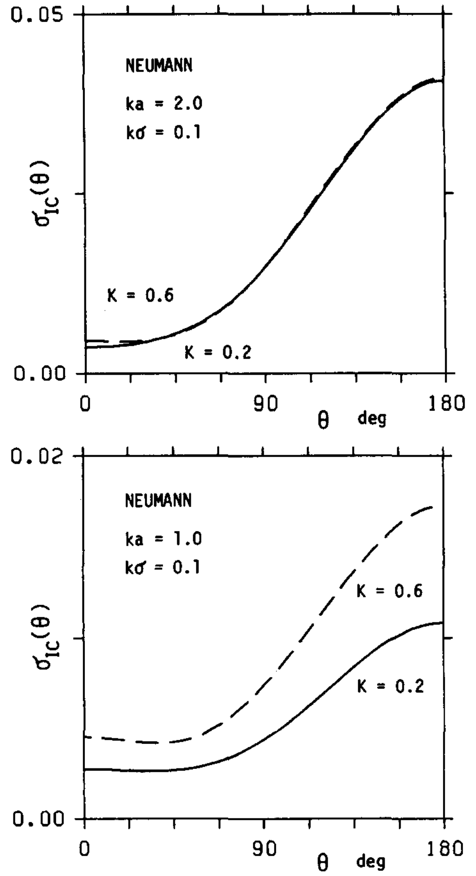


FIG. 8. Angular distribution of the incoherent scattering for the plane-wave incidence (Neumann, $ka = 1.0, 2.0$, $k\sigma = 0.1$, $K = 0.2, 0.6$).

the circle, i.e., ka, is made much larger beyond the Mie scattering range. The analysis of scattering for such large ka will be reported in a succeeding paper.

It is important to mention that for rougher surfaces, i.e., for larger $k\sigma$, we have to take into account the terms due to A^2_m or higher-order WH coefficients, which, however, could be obtained in many intractable forms from higher-order equations similar to (65)–(67) or (76)–(78). Another point is that our approximate boundary condition (22) or (23), which retains only linear terms in f to model a slightly random surface, is not sufficient enough to deal with the scattering by a very rough surface. In the very rough case involving higher-order WH coefficients, therefore, the boundary condition has to be treated differently along the line suggested by Nakayama[40] and also their recent work based on the WH expansion by Meecham and Lin.[33]

ACKNOWLEDGMENTS

The authors are deeply indebted to Professor T. Mukai of Kanazawa Institute of Technology for his suggesting the problem and related references.

[1] *Light Scattering by Irregularly Shaped Particles*, edited by D. W. Schuerman (Plenum, New York, 1980).
[2] S. Mukai, T. Mukai, R. H. Giese, K. Weiss, and R. H. Zerull, Moon Planets **26**, 197 (1982).
[3] R. Schiffer, Astron. Astrophys. **148**, 347 (1985).
[4] C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1983).
[5] S. Hayashi, Jpn. J. Appl. Phys. **23**, 665 (1984).
[6] T. Oguchi, Proc. IEEE **71**, 1029 (1983).
[7] S. O. Rice, Commun. Pure Appl. Math. **4**, 351 (1951).
[8] P. Beckmann and A. Sppichichino, *The Scattering of Electromagnetic Waves from Rough Surfaces* (Pergamon, New York, 1963).
[9] V. D. Freiliker and I. M. Fuks, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **13**, 98 (1970).
[10] G. G. Zipfel and J. A. DeSanto, J. Math Phys. **13**, 1903 (1972).
[11] J. A. DeSanto, J. Math. Phys. **15**, 283 (1974).
[12] G. S. Brown, IEEE Trans. Antennas Propag. **AP-30**, 1135 (1982).
[13] G. S. Brown, IEEE Trans. Antennas Propag. **AP-31**, 5 (1983).
[14] S. Ito, J. Radio Res. Lab. **31**, 1 (1984).
[15] J. Nakayama, H. Ogura, and B. Matsumoto, Radio Sci. **15**, 1049 (1980).
[16] J. Nakayama, H. Ogura, and M. Sakata, J. Math. Phys. **22**, 471 (1981).
[17] J. Nakayama, H. Ogura, and M. Sakata, Radio Sci. **16**, 831, 847 (1981).
[18] H. Ogura and N. Takahashi, J. Opt. Soc. Am. A **2**, 2208 (1985).
[19] H. Ogura, Phys. Rev. A **11**, 942 (1975).
[20] H. Ogura and J. Nakayama, Phys. Rev. A **11**, 957 (1975).
[21] N. Wiener, *Nonlinear Problems in Random Theory* (M.I.T., Cambridge, MA, 1958).
[22] R. H. Cameron and W. T. Martin, Ann. Math. **48**, 385 (1947).
[23] K. Ito, J. Math. Soc. Jpn. **13**, 157 (1951).
[24] K. Ito, Jpn. J. Math. **22**, 63 (1952).
[25] A. Siegel, T. Imamura, and W. C. Meecham, J. Math. Phys. **6**, 707 (1965).
[26] H. D. Hogge and W. C. Meecham, J. Fluid Mech. **85**, 325 (1978).
[27] C. P. Lee, W. C. Meecham, and H. D. Hogge, Phys. Fluids **25**, 1322 (1982).
[28] J. F. Barrett, J. Electron. Control **16**, 107 (1963).
[29] Y. Lee and M. Schetzen, Inst. J. Control **2**, 237 (1965).
[30] M. Schetzen, *The Volterra and Wiener Theory of Nonlinear Systems* (Wiley, New York, 1980).
[31] J. Nakayama, Doctoral thesis, Kyoto University, 1982.
[32] J. Nakayama, K. Mizutani, H. Ogura, and S. Hayashi, J. Appl. Phys. **56**, 1465 (1984).
[33] W. C. Meecham and W. W. Lin, J. Wave Mater. Interaction **1**, 4 (1986).

[34]R. Schiffer and K. O. Thielheim, J. Appl. Phys. **57**, 2437 (1985).
[35]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[36]F. Andorf, Master thesis, Ruhr University, 1978; T. Mukai (private communication).
[37]I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications*, English translation (Pergamon, London, 1963).
[38]N. Ja. Vilenkin, *Special Functions and the Theory of Group Representations*, English translation (American Mathematical Society, Providence, 1968).
[39]H. Ogura, *Theory of Random Processes*, in Japanese (Corona, Tokyo, 1978).
[40]J. Nakayama, Radio Sci. **21**, 707 (1986).

On modeling discontinuous media. One-dimensional approximations

P. C. Sabatier and B. Dolveck-Guilpard
Laboratoire de Physique Mathématique, Université des Sciences et Techniques du Languedoc, Place Eugène-Bataillon, 34060 Montpellier Cédex, France

(Received 16 July 1987; accepted for publication 28 October 1987)

Distorted wave Born series are derived for the impedance equation $[A^{-1}(d/dx)A(d/dx) + k^2]p = 0$, where A is a piecewise differentiable function. The first and second orders are given explicitly. They hold for impedances A, which include discontinuities, before and/or after perturbation, so that the problem could not be reduced to the Schrödinger one. The results are used for discussing homogenizations currently practiced in modeling discontinuous media.

I. INTRODUCTION

Wave propagation in discontinuous media is observed in several domains of physics. Obvious examples include heat conduction, neutron transport theory, electromagnetic waves, acoustic waves, elastic waves, etc. For the sake of simplicity, modeling these problems reduces the medium either to a multilayered one, with homogeneous layers (inside each layer the parameters are constant) or to a continuously varying one (mathematical simplicity usually requires that the parameters are twice differentiable). These approximations are certainly not "innocent" since (a) they lead to quite different analyses in spectral problems[1]; (b) the reflection coefficient from a discontinuous medium does not go to zero at high frequency (it is an almost periodical function); and (c) in two-dimensional or three-dimensional cases, the difference between continuous and discontinuous media is dramatic: surface waves can be generated and propagate along discontinuity surfaces and they considerably modify the results of a sounding. In acoustical microscopy for instance, the resolving length for layered media is very different from that for continuously varying media.

Now, a glance at any drilling situation shows that (unless layers are very thin) homogeneous assumptions are unfounded and that there are many discontinuities.

In the present paper, we begin a study of this modeling in the simplest case, that of one dimension. In Sec. II we construct the mathematical tools to be used: the essential tool being a distorted wave Born series valid altogether for (small) discontinuous and (small) continuous perturbations. The term "distorted wave" means that we start from the solution of an (arbitrary) reference model—which may or may not be continuous. This tool is used in Sec. III to analyze the "homogenization" problem's "intra" layers.

For a general reference on the one-dimensional wave equations and their application in geophysics, the reader is suggested to refer to papers by Newton,[2] Howard,[3] and Sabatier.[4] The physical domain of applicability for these classical models is that of wave frequencies up to, but not including, infrared or optical frequencies, where quantum many-body effects become important. For a first derivation of the central equations of wave propagation through a discontinuous impedance, the reader is referred to the papers by Sabatier[5,6] and Degasperis and Sabatier.[7]

II. DISTORTED WAVE BORN SERIES IN DISCONTINUOUS CASES

In the one-dimensional model studied here, we assume that the "wave equation" is replaced by the "impedance equation":

$$\left(A^{-1}\frac{d}{dx}A\frac{d}{dx}+k^2\right)p(k,x)=0,\quad p, A\frac{dp}{dx}\ \text{continuous,} \tag{2.1}$$

where $A=\alpha^2$ is a positive function of x, everywhere twice differentiable except at the finite number of points $x_0<x_1<\cdots<x_N$, where A and A' show jumps, and such that, for any $x\notin S=\{x_n\}$,

$$\alpha^{-1}\frac{d^2}{dx^2}\alpha=U \tag{2.2}$$

is a "potential" whose properties guarantee the usual scattering theory [say, $U\in L_1^1=\{U|\int_{-\infty}^{+\infty}(1+|t|)U(t)|dt<\infty\}$]. These conditions guarantee the "impedance theory" developed by one of us in previous papers.[5–7] We assume also that the impedance factor α has been chosen in the class of "standard equivalent" ones[7] so that it goes to the constant 1 as x goes to $+\infty$.

At each "singular" point x_n we define the transmission factor t_n, the reflection factor r_n, and the slope factor s_n by the following formulas, where α_n^+ stands for $\alpha(x_n^+)$, $\alpha_n'^+$ for $\alpha'(x_n^+)$, etc.:

$$t_n^{-1}=\tfrac{1}{2}(\alpha_n^+/\alpha_n^- + \alpha_n^-/\alpha_n^+)\,, \tag{2.3a}$$

$$t_n^{-1}r_n=\tfrac{1}{2}(\alpha_n^+/\alpha_n^- - \alpha_n^-/\alpha_n^+)\,, \tag{2.3b}$$

$$t_n^{-1}s_n=\tfrac{1}{2}(\alpha_n'^-/\alpha_n^+ - \alpha_n'^+/\alpha_n^-)\,. \tag{2.3c}$$

For $x<x_0$ or $x>x_N$, or more generally at each "regular" point, setting $p=\alpha^{-1}f$ in (2.1) yields the Schrödinger equation

$$\left[\frac{d^2}{dx^2}+(k^2-U)\right]f=0\,; \tag{2.4}$$

thus the impedance equation could also be considered as a certain chain of Schrödinger equations with the particular conserved quantities p and $A\,dp|dx$ at each discontinuity. Hence we can define, for real k, the Jost solutions of (2.1) by their asymptotic behavior

 0022-2488/88/040861-08$02.50

$$p_+(k,x) \sim \alpha^{-1} \exp(ikx) \quad (x \to +\infty),$$
$$p_-(k,x) \sim \alpha^{-1} \exp(-ikx) \quad (x \to -\infty). \tag{2.5}$$

The basic property of linearly independent solutions enables us to define the reflection coefficient $R(k)$ and the transmission coefficient $T(k)$:

$$T(k)p_-(k,x) = p_+(-k,x) + R(k)p_+(k,x). \tag{2.6}$$

They can be expressed in terms of the Wronskian $p_+p'_- - p_-p'_+$ of p_+ and p_-. In particular,

$$T(k) = 2ik[\alpha(x)]^{-2}[W(p_-(k,x),p_+(k,x))]^{-1}. \tag{2.7}$$

Note that $\alpha^2 W$ does not depend on x. These equations can be used, for instance, in a model of the sounding of a one-dimensional medium by electromagnetic or elastic waves; $R(k)$ is then the result to be used for parameter identification.

Now let the "unperturbed" problem be characterized by a potential $U \in L_1^1$ and a set of singular points S, with the corresponding reflection, transmission, and slope factors, all this together being equivalent to giving the unperturbed impedance factor α. We assume that the Jost solutions $p_\pm$ and the reflection and transmission coefficients R and T have been calculated for this problem. Let the perturbed problem be characterized by $\widetilde{U}, \widetilde{S}$, and singular data, and let the corresponding functions be $\widetilde{p}_\pm$, $\widetilde{R}$, and $\widetilde{T}$. Let $S \cup \widetilde{S} = \mathbf{S}$. The set $\mathbf{S}$ contains subsets where every point of S is associated with a corresponding point of $\widetilde{S}$, which may be identical with itself (we call it a double point, and denote their subset as S_1) or not identical (we call the two points a pair and denote their subset as S_2). Let S_3 be the set of all the other points of $\mathbf{S}$. In the following, we assume that the reflection and slope factors of all points of S_3 are of order ε (this meaning, for example, that they have been obtained by multiplying fixed numbers, of order 1 or less, by a factor ε that need not be written down in our formulas). Also, for a pair of points of S_1, we assume that the reflection and slope factor of the S point and that of the $\widetilde{S}$ point differ by numbers of order ε—and make a similar assumption on the two components of a double point of S_2. We also assume that the distances between two points of S_2 associated in pairs are of order ε. Now our aim is to show a way for obtaining $\widetilde{p}_\pm$, $\widetilde{R}(k)$, and $\widetilde{T}(k)$ up to any orders $O(\varepsilon^p)$, and to write down explicitly the first order (which could also be called a distorted wave Born approximation). It is important to remark that there is not any Schrödinger equation that is equivalent to (2.1) and that enables one to derive the standard generalized Born series. This is because the δ' distributions that appear after blindly using (2.2) cannot be multiplied by (piecewise) continuous functions and none of the usual terms of the generalized Born series will be obtained. As a matter of fact, in the derivations to follow, we shall be cautious never to introduce anything more singular than Dirac measures at the given discrete points of S, and then to multiply them only by functions that are continuous at their point support (of course they may have discontinuities elsewhere).

A. Basic formulas

Now, from (2.5), with $\widetilde{k} = \pm k$, and since $\alpha(\infty) = \widetilde{\alpha}(\infty) = 1$, and $\alpha'(\infty) = \widetilde{\alpha}'(\infty) = 0$, we see that

$$\lim_{x\to\infty} W(p_+(k,x),\widetilde{p}_+(\widetilde{k},x)) = \begin{cases} 0, & \text{if } k = \widetilde{k}, \\ 2i\widetilde{k}, & \text{if } k = -\widetilde{k}. \end{cases} \tag{2.8}$$

This result and (2.6) can be used to prove that

$$\lim_{x\to\infty} W(p_-(k,x),\widetilde{p}_-(k,x)) = 2ik\,[\widetilde{R}(k) - R(k)]/T(k)\widetilde{T}(k), \tag{2.9}$$

$$\lim_{x\to\infty} W[(\widetilde{p}_-(k,x) - p_-(k,x),p_+(k,x)] = 2ik\,[T(k) - \widetilde{T}(k)]/T(k)\widetilde{T}(k). \tag{2.10}$$

B. Basic equation for the Jost solution

We need to calculate $\widetilde{p}_-(k,x)$. Following Ref. 7 (where it was called $\bar{\sigma}$), we first introduce the singular data function $\boldsymbol{\sigma}$, which is constant on each regular interval. It shows at each singular point (i.e., each discontinuity of α or α') the reflection and transmission factors of α, and is normalized by $\boldsymbol{\sigma}(\infty) = 1$. It can be written explicitly as

$$\boldsymbol{\sigma}(x) = \exp\Big[-\sum_{x_j\in S} \rho_j \Theta(x_j - x)\Big], \tag{2.11}$$

where Θ is the Heaviside function and

$$\rho_j = \log[(1 + r_j)/t_j] = \tfrac{1}{2}\log[(1 + r_j)/(1 - r_j)]. \tag{2.12}$$

The functions $\boldsymbol{\sigma}$ and $\widetilde{\boldsymbol{\sigma}}$ enable us to write the equations for

$$\rho = \alpha p_-/\boldsymbol{\sigma}, \quad \widetilde{\rho} = \widetilde{\alpha}\widetilde{p}_-/\widetilde{\boldsymbol{\sigma}}. \tag{2.13}$$

They read[7]

$$\Big[\frac{d}{dx}\boldsymbol{\sigma}^2\frac{d}{dx} + \boldsymbol{\sigma}^2(k^2 - U) + \sum_{x_i\in S} u(x_i)\delta(x - x_i)\Big] \times\rho(k,x) = 0, \tag{2.14}$$

$$\Big[\frac{d}{dx}\widetilde{\boldsymbol{\sigma}}^2\frac{d}{dx} + \widetilde{\boldsymbol{\sigma}}^2(k^2 - \widetilde{U}) + \sum_{x_j\in \widetilde{S}} \widetilde{u}(x_j)\delta(x - x_j)\Big] \times\widetilde{\rho}(k,x) = 0, \tag{2.15}$$

where

$$u(x_i) = 2\boldsymbol{\sigma}(x_i^+)\boldsymbol{\sigma}(x_i^-)t_i^{-1}s_i, \tag{2.16}$$

and a similar formula holds for $\widetilde{u}(x_j)$. In the following it is convenient to number the singularities as an increasing sequence $x_0, x_1, \ldots$ inside $S \cup \widetilde{S} = \mathbf{S}$. In each regular interval between the points of $\mathbf{S}$, both Eqs. (2.14) and (2.15) do not contain the Dirac measures. Setting

$$\tau(k,x) = \widetilde{\rho}(k,x)/\rho(k,x) \tag{2.17}$$

inside (2.15) and using the identity

$$\frac{d}{dx}\Big(\widetilde{\boldsymbol{\sigma}}^2\frac{d}{dx}\rho\tau\Big) = \rho^{-1}\frac{d}{dx}\Big(\rho^2\widetilde{\boldsymbol{\sigma}}^2\frac{d\tau}{dx}\Big) + \tau\frac{d}{dx}\widetilde{\boldsymbol{\sigma}}^2\frac{d\rho}{dx}, \tag{2.18}$$

we obtain on each of these regular intervals

$$\frac{d}{dx}\Big(\rho^2\widetilde{\boldsymbol{\sigma}}^2\frac{d\tau}{dx}\Big) + \rho^2\widetilde{\boldsymbol{\sigma}}^2(U - \widetilde{U})\tau = 0. \tag{2.19}$$

It is tedious but very easy to calculate directly the discontinuities of $\rho^2\widetilde{\boldsymbol{\sigma}}^2(d\tau/dx)$ [$= \widetilde{\boldsymbol{\sigma}}^2 W(\rho,\widetilde{\rho})$] across each point of $\mathbf{S}$, it being understood that if α (resp. $\widetilde{\alpha}$) is continuously differentiable at x_n, then t_n (resp. $\tilde{t}_n$) is equal to 1, whereas r_n, s_n (resp. $\tilde{r}_n, \tilde{s}_n$) vanish. The result is

$$\left[\rho^2\tilde{\sigma}^2\frac{d\tau}{dx}\right]_{x_n^-}^{x_n^+} = -w_n\tau(k,x_n), \tag{2.20}$$

where

$$\begin{aligned} w_n &= \rho^2(k,x_n)\tilde{\sigma}(x_n^+)\tilde{\sigma}(x_n^-)\Bigg[2\frac{\tilde{s}_n}{\tilde{t}_n} - 2\frac{s_n}{t_n}\frac{1-r_n\tilde{r}_n}{t_n\tilde{t}_n} \\ &\quad + \frac{\tilde{r}_n - r_n}{t_n\tilde{t}_n}\left(\frac{\alpha'(x_n^+)}{\alpha(x_n^-)} + \frac{\alpha'(x_n^-)}{\alpha(x_n^+)}\right) \\ &\quad + \frac{(\alpha^2p'_-)(x_n)/p_-(x_n)}{\alpha(x_n^-)\alpha(x_n^+)}\left(\frac{1+\tilde{r}_n}{\tilde{t}_n} - \frac{1+r_n}{t_n}\right) \\ &\quad \times\left(\frac{1-\tilde{r}_n}{\tilde{t}_n} + \frac{1-r_n}{t_n}\right)\Bigg], \end{aligned} \tag{2.21a}$$

$$= \rho^2(k,x_n)\tilde{\sigma}(x_n^+)\tilde{\sigma}(x_n^-)\omega_n(k). \tag{2.21b}$$

Combining (2.19) and (2.20) yields the equation

$$\left[\frac{d}{dx}\rho^2\tilde{\sigma}^2\frac{d}{dx} + \rho^2\tilde{\sigma}^2(U-\tilde{U}) + \sum_{x_n\in S} w_n\delta(x-x_n)\right] \times\tau(k,x) = 0. \tag{2.22}$$

Integrating (2.22) twice yields a Volterra integral equation for $\tau(k,x)$ (the fixed parameter k is omitted):

$$\begin{aligned} \tau(x) &= \gamma - \int_{-\infty}^{x} dt\,\rho^{-2}(t)\tilde{\sigma}^{-2}(t) \\ &\quad \times\int_{-\infty}^{t} ds\,\rho^2(s)\tau(s)w(s), \end{aligned} \tag{2.23a}$$

$$= \gamma - \int_{-\infty}^{x} ds\,\rho^2(s)w(s)\tau(s)\int_s^x dt\,\rho^{-2}(t)\tilde{\sigma}^{-2}(t), \tag{2.23b}$$

where

$$\gamma = \tau(-\infty) = \sigma(-\infty)/\tilde{\sigma}(-\infty), \tag{2.24}$$

$$w(s) = \tilde{\sigma}^2(s)(U(s) - \tilde{U}(s)) + \sum_{x_n\in S}\tilde{\sigma}(x_n^+)\tilde{\sigma}(x_n^-)\omega_n\delta(s-x_n). \tag{2.25}$$

C. Iterative solution of the integral equation

The convergence of the iterative solution of (2.23b) is guaranteed by that of the iterative solution of a majorant equation. It is easy to see that there exists, in general, a function $\bar{w}(s)$ in L_1^1 and constants $C, \bar{w}_1, \bar{w}_2,\ldots,\bar{w}_n$, such that

$$\begin{aligned} &\left|\int_s^x dt\,\rho^{-2}(t)\tilde{\sigma}^{-2}(t)\right| \\ &\quad < C|x-s| < C[1+|x|\theta(x)][1+|s|\theta(-s)] \\ &\quad (s<x), \end{aligned} \tag{2.26}$$

$$|\rho^2 w(s)| < \bar{w}(s) + \sum_{x_n\in S}\bar{w}_n\delta(x_n - s). \tag{2.27}$$

Thus a majorant equation for (2.23b) is[5]

$$\bar{\tau}(x) = |\gamma| + \theta(x)\int_{-\infty}^{x} ds\,\bar{\tau}(s)\beta(s), \tag{2.28}$$

with

$$\alpha(x) = C[1+|x|\theta(x)], \tag{2.29a}$$

$$\beta(x) = [1+|x|\theta(-x)]\left[\bar{w}(x) + \sum_{x_n\in S}\bar{w}_n\delta(x_n - x)\right]. \tag{2.29b}$$

The iterative solution to (2.28) is easy to write explicitly since

$$\bar{\tau}(x) = |\gamma| + \alpha(x)\int_{-\infty}^{x}\exp\left[\int_s^x \alpha(t)\beta(t)dt\right]\beta(s)ds. \tag{2.30}$$

Hence the convergence is uniform for any fixed x, provided that $|V|\in L_1^1$. The series can also be differentiated and $\tau'(x)$ can be calculated readily. If there were no pairs of singular points, it would be obvious that the nth term of the iterated series would be $O(\epsilon^n)$. For $x_n\in S_1$, w_n is also of order ε [trivially from (2.21a)]. Let us consider a pair of points in S_2, say, x_n, x_{n+1}. The values of w_n and w_{n+1} are $O(1)$ only, but, since $|x_n - x_{n+1}| = O(\varepsilon)$, one easily shows that $|\rho(x_n) - \rho(x_{n+1})|$ and $|\alpha(x_n^+) - \alpha(x_{n+1}^-)|$ are $O(\varepsilon)$. Supposing, for instance, $x_n\in S$, $x_{n+1}\in\tilde{S}$, we obtain from (2.21a) and (2.3) the following formulas:

$$\begin{aligned} w_n &\sim \rho^2(x_n)[\tilde{\sigma}(x_n]^2\bigg\{-\frac{2s_n}{t_n} - \frac{r_n}{t_n}\left(\frac{\alpha'(x_n^+)}{\alpha(x_n^-)} + \frac{\alpha'(x_n^-)}{\alpha(x_n^+)}\right) \\ &\quad - \frac{2r_n}{t_n}\frac{(\alpha^2p'_-)(x_n)/p(x_n)}{\alpha(x_n^-)\alpha(x_n^+)}\bigg\}, \end{aligned} \tag{2.31}$$

$$\begin{aligned} w_{n+1} &\sim \rho^2(x_n)[\tilde{\sigma}(x_n)]^2\frac{\tilde{\sigma}(x_{n+1}^+)}{\tilde{\sigma}(x_{n+1}^-)}[1+O(\varepsilon)] \\ &\quad \times\bigg\{2\frac{\tilde{s}_{n+1}}{\tilde{t}_{n+1}} + 2\frac{\tilde{r}_{n+1}}{\tilde{t}_{n+1}}\frac{\alpha'(x_{n+1})}{\alpha(x_{n+1})} \\ &\quad + 2\frac{\tilde{r}_{n+1}}{\tilde{t}_{n+1}}\frac{(\alpha^2p'_-)(x_{n+1})/p_-(x_{n+1})}{\alpha^2(x_{n+1})}\bigg\}; \end{aligned} \tag{2.32}$$

and after some algebra, using (2.3) yields

$$\begin{aligned} w_n &= -2\rho^2(x_n)[\bar{\sigma}(x_n)]^2\bigg\{\frac{s_n(1+r_n)}{t_n^2} + \frac{\alpha'(x_n^+)}{\alpha(x_n^+)} \\ &\quad \times\frac{r_n(1+r_n)}{t_n^2} + \frac{r_n(1+r_n)(\alpha^2p'_-/p_-)(x_n)}{t_n^2\,\alpha^2(x_n^+)}\bigg\}, \end{aligned} \tag{2.33}$$

$$w_{n+1} = -w_n[1+O(\varepsilon)]. \tag{2.34}$$

Now we easily show that if a function $F_1(x)$ has the property

$$|F_1(x)| = O(\varepsilon), \quad |F_1(x) - F_1(y)| = O(\varepsilon)|x-y|, \tag{2.35}$$

then the function F_2 defined by

$$F_2(x) = \int[w_n\delta(s-x_n) + w_{n+1}\delta(s-x_{n+1})] \times[F_1(x) - F_1(s)], \tag{2.36}$$

where w_n satisfies (2.34) and $|x_{n+1} - x_n| = O(\varepsilon)$, has the property

$$|F_2(x)| = O(\varepsilon^2), \quad |F_2(x) - F_2(y)| = O(\varepsilon^2)|x-y|. \tag{2.37}$$

The function $\int_s^x dt\,\rho^{-2}(t)\tilde{\sigma}^{-2}(t)$, which is given in (2.23b), is equal to $F(x)-F(s)$, where F has the property of F_1, and the property (2.34) is not modified if w_n is multiplied by a piecewise differential function. Using these remarks in the iterated terms of the solution of (2.23b), it is not difficult to show that the first term is $O(\varepsilon)$ and the remainder is $O(\varepsilon^2)$.

D. Generalized Born approximation

Noticing that $\tilde{\sigma}^{-2}(t)=\sigma^{-2}(t)\ [1+O(\varepsilon)]$, we can use in the first iterated term the approximate value of $F(x)$:

$$F(x)\simeq\int^x dt\,\rho^{-2}(t)\sigma^{-2}(t)=\int^x dt\,\alpha^{-2}(t)[p^-(t)]^{-2}=\frac{1}{2ik}\frac{p_-(-k,x)}{p_-(k,x)}. \tag{2.38}$$

Using also (2.13) and keeping only the $O(\varepsilon)$ terms, we obtain the Born approximation for $\tilde{p}_-(k,x)$ and $\tilde{p}'_-(k,x)$:

$$\begin{aligned}\tilde{p}_-(k,x)=\gamma(x)p_-(k,x)-\frac{1}{2ik}\int_{-\infty}^x ds\,\alpha^2(s)\sigma^{-2}(s)\\ \times w(s)p_-(k,s)[p_-(k,s)p_-(-k,x)\\ -p_-(k,x)p_-(-k,s)]+O(\varepsilon^2),\end{aligned} \tag{2.39}$$

$$\begin{aligned}\tilde{p}'_-(k,x)=\ &\gamma(x)p'_-(k,x)+\gamma'(x)p_-(k,x)\\ &-\frac{1}{2ik}\int_{-\infty}^x ds\,\alpha^2(s)\sigma^{-2}(s)w(s)p_-(k,s)\\ &\times[p_-(k,s)p'_-(-k,x)\\ &-p'_-(k,x)p_-(-k,s)]+O(\varepsilon^2),\end{aligned} \tag{2.40}$$

where

$$\gamma(x)=\gamma(\tilde{\sigma}/\tilde{\alpha})(\alpha/\sigma). \tag{2.41}$$

Inserting (2.39) and (2.40) into (2.9) and (2.10) shows that $\tilde{T}(k)=T(k)(1+O(\varepsilon))$ and that, up to $O(\varepsilon^2)$,

$$\begin{aligned}2ik\frac{\tilde{R}(k)-R(k)}{[T(k)]^2}\\ =-\int_{-\infty}^{+\infty}ds\,\alpha^2(s)\sigma^{-2}(s)[p_-(k,s)]^2\\ \times\Big[\tilde{\sigma}^2(s)(U(s)-\tilde{U}(s))+\sum_{x_n\in S}\tilde{\sigma}(x_n^+)\tilde{\sigma}(x_n^-)\\ \times\omega_n(k)\delta(s-x_n)\Big],\end{aligned} \tag{2.42a}$$

which is the Born approximation for $\tilde{R}(k)-R(k)$. It can also be written, up to the same order, as

$$\begin{aligned}2ik\frac{\tilde{R}(k)-R(k)}{T^2(k)}\\ =-\int_{-\infty}^{+\infty}ds[p_-(k,s)]^2\Big\{\alpha^2(s)(U(s)-\tilde{U}(s))\\ +\sum_{x_n\in S}\alpha(x_n^+)\alpha(x_n^-)\omega_n(k)\delta(s-x_n)\Big\}.\end{aligned} \tag{2.42b}$$

E. Study of the other Jost solution

A quite similar analysis can be done for $p_+(k,x)$. Setting

$$P=\alpha p_+/\sigma,\quad \tilde{P}=\tilde{\alpha}\tilde{p}_+/\tilde{\sigma},\quad T=\tilde{P}/P, \tag{2.43}$$

we easily derive the integral equation for $T(k,x)$:

$$T(x)=1-\int_x^\infty ds\,P^2(s)T(s)W(s)\int_x^s P^{-2}(t)\tilde{\sigma}^{-2}(t)dt, \tag{2.44}$$

where

$$\begin{aligned}W(s)=\tilde{\sigma}^2(s)(U(s)-\tilde{U}(s))\\ +\sum_{x_n\in S}\tilde{\sigma}(x_n^+)\tilde{\sigma}(x_n^-)\Omega_n\delta(x-x_n),\end{aligned} \tag{2.45}$$

$$\begin{aligned}\Omega_n=\ &2\frac{\tilde{s}_n}{\tilde{t}_n}-2\frac{s_n}{t_n}\frac{1-r_n\tilde{r}_n}{t_n\tilde{t}_n}\\ &+\frac{\tilde{r}_n-r_n}{t_n\tilde{t}_n}\left(\frac{\alpha'(x_n^+)}{\alpha(x_n^-)}+\frac{\alpha'(x_n^-)}{\alpha(x_n^+)}\right)\\ &+\frac{(\alpha^2p'_+)(x_n)/p_+(x_n)}{\alpha(x_n^-)\alpha(x_n^+)}\left(\frac{1+\tilde{r}_n}{\tilde{t}_n}-\frac{1+r_n}{t_n}\right)\\ &\times\left(\frac{1-\tilde{r}_n}{\tilde{t}_n}+\frac{1-r_n}{t_n}\right).\end{aligned} \tag{2.46}$$

The Born approximation for $\tilde{p}_+(k,x)$ is

$$\begin{aligned}\tilde{p}_+(k,x)=\Gamma(x)p_+(k,x)-\frac{1}{2ik}\int_x^\infty ds\,\alpha^2(s)\sigma^{-2}(s)\\ \times W(s)p(k,s)[p_+(k,s)p_+(-k,x)\\ -p_+(-k,s)p_+(k,x)]+O(\varepsilon^2),\end{aligned} \tag{2.47}$$

where

$$\Gamma(x)=(\tilde{\sigma}/\tilde{\alpha})(\alpha/\sigma). \tag{2.48}$$

F. Second order approximation

Let us set for the sake of simplicity,

$$w(x)=V(x)+\sum_{x_n\in S}v_n\delta(x-x_n), \tag{2.49a}$$

$$V(x)=\tilde{\sigma}^2(x)(U(x)-\tilde{U}(x)), \tag{2.49b}$$

$$v_n=\tilde{\sigma}(x_n^+)\tilde{\sigma}(x_n^-)\omega_n, \tag{2.49c}$$

and notice that from (2.38),

$$\begin{aligned}2ik\int_s^x dt\,\rho^{-2}(t)\tilde{\sigma}^{-2}(t)\\ =\frac{p_-(-k,x)\sigma^2(x)}{p_-(k,x)\tilde{\sigma}^2(x)}-\frac{p_-(-k,s)}{p_-(k,s)}\frac{\sigma^2(s)}{\tilde{\sigma}^2(s)}\\ +\sum_{x_n\in S}\frac{p_-(-k,x_n)}{p_-(k,x_n)}\eta_n[\theta(x-x_n)-\theta(s-x_n)]\\ =2ikF(x,s),\end{aligned} \tag{2.50}$$

where

$$\eta_n=-2\frac{\sigma(x_n^+)\sigma(x_n^-)}{\tilde{\sigma}(x_n^+)\tilde{\sigma}(x_n^-)}\frac{r_n-\tilde{r}_n}{t_n\tilde{t}_n}.$$

Then we calculate the second order iterative solution of (2.23b), namely,

$$\tau(x)=\gamma-\gamma\int_{-\infty}^{x} ds\,\rho^2(s)w(s)F(x,s)$$

$$+\gamma\int_{-\infty}^{x} ds\,\rho^2(s)w(s)$$

$$\times F(x,s)\int_{-\infty}^{s} dt\,\rho^2(t)w(t)F(s,t)+O(\varepsilon^3), \tag{2.51}$$

by using (2.49) and the formula (2.50) each time $F(x,s)$ is multiplied by piecewise continuous functions, but not when it is multiplied by Dirac measures.

Our aim is to obtain an approximation of $\tau(x)$ containing the $O(\varepsilon^2)$ terms and neglecting $O(\varepsilon^3)$, so that in the last term of (2.51) we can replace $\tilde{\sigma}$ by σ. The result is

$$\begin{aligned}\tau(x)=\gamma&-\frac{\gamma}{2ik}\int_{-\infty}^{x} ds\,\rho^2(s)V(s)\left[\frac{p_-(-k,x)}{p_-(k,x)}\frac{\sigma^2(x)}{\tilde{\sigma}^2(x)}-\frac{p_-(-k,s)}{p_-(k,s)}\frac{\sigma^2(s)}{\tilde{\sigma}^2(s)}\right]\\&-\frac{\gamma}{2ik}\sum_{x_n\in S}\frac{p_-(-k,x_n)}{p_-(k,x_n)}\eta_n\theta(x-x_n)\int_{-\infty}^{x_n} dt\,\rho^2(t)V(t)-\gamma\sum_{x_n\in S}\rho^2(x_n)v_n\theta(x-x_n)\int_{x_n}^{x} dt\,\rho^{-2}(t)\tilde{\sigma}^{-2}(t)\\&-\frac{\gamma}{4k^2}\int_{-\infty}^{x} ds\,\rho^2(s)w(s)\left[\frac{p_-(-k,x)}{p_-(k,x)}-\frac{p_-(-k,s)}{p_-(k,s)}\right]\int_{-\infty}^{s} dt\,\rho^2(t)w(t)\left[\frac{p_-(-k,s)}{p_-(k,s)}-\frac{p_-(-k,t)}{p_-(k,t)}\right].\end{aligned} \tag{2.52}$$

Using (2.13) and (2.17) we obtain, for $x>x_N$,

$$\begin{aligned}&\tilde{p}_-(k,x)-\Gamma(x)\gamma p_-(k,x)\\&\quad=-\frac{\Gamma(x)\gamma}{2ik}\int_{-\infty}^{x} ds\,\frac{\alpha^2(s)}{\sigma^2(s)}V(s)p_-(k,s)\left[\frac{\sigma^2(x)}{\tilde{\sigma}^2(x)}p_-(-k,x)p_-(k,s)-\frac{\sigma^2(s)}{\tilde{\sigma}^2(s)}p_-(-k,s)p_-(k,x)\right]\\&\qquad-\frac{\Gamma(x)\gamma}{2ik}\sum_{x_n\in S}p_-(k,x)\frac{p_-(-k,x_n)}{p_-(k,x_n)}\eta_n\int_{-\infty}^{x_n} dt\,\rho^2(t)V(t)-\Gamma(x)\gamma\sum_{x_n\in S}p_-(k,x)\rho^2(x_n)v_n\int_{x_n}^{x} dt\,\rho^{-2}(t)\tilde{\sigma}^{-2}(t)\\&\qquad-\frac{\Gamma(x)\gamma}{4k^2}\int_{-\infty}^{x} ds\,\frac{\alpha^2(s)}{\tilde{\sigma}^2(s)}w(s)\left[p_-(-k,x)p_-(k,s)-p_-(-k,s)p_-(k,x)\right]\\&\qquad\times\int_{-\infty}^{s} dt\,\frac{\alpha^2(t)}{\sigma^2(t)}w(t)p_-(k,t)\left[p_-(-k,s)p_-(k,t)-p_-(-k,t)p_-(k,s)\right],\end{aligned} \tag{2.53}$$

and the derived formula for $\tilde{p}'_-(k,x)$. Using $\Gamma(x)\to 1$ and $\Gamma'(x)\to 0$ as $x\to\infty$, it is not difficult to calculate the limit Wronskian that appears in the left-hand side of (2.9) and (2.10). Hence we obtain

$$\begin{aligned}2ik\frac{\tilde{R}(k)-R(k)}{T(k)\tilde{T}(k)}=&-\gamma\int_{-\infty}^{+\infty} ds\,\frac{\alpha^2(s)}{\sigma^2(s)}\tilde{\sigma}^2(s)[U(s)-\tilde{U}(s)]p_-^2(k,s)\\&-\gamma\sum_{x_n\in S}\left(\frac{\alpha}{\sigma}\right)^2(x_n)\tilde{\sigma}(x_n^+)\tilde{\sigma}(x_n^-)\omega_n p_-^2(k,x_n)+\frac{1}{2ik}\int_{-\infty}^{+\infty}\frac{\alpha^2(s)}{\sigma^2(s)}w(s)ds\int_s^{\infty}\frac{\alpha^2(t)}{\sigma^2(t)}w(t)p_-(k,s)\\&\times[p_-(-k,s)p_-(k,t)-p_-(-k,t)p_-(k,s)]p_-(k,t)dt+O(\varepsilon^3),\end{aligned} \tag{2.54}$$

$$\begin{aligned}2ik\left(\frac{1}{\tilde{T}(k)}-\frac{1}{T(k)}\right)=&\,2ik\frac{\gamma-1}{T(k)}-\frac{R^+(-k)}{T(-k)}\int_{-\infty}^{+\infty} ds\,\alpha^2(s)[U(s)-\tilde{U}(s)]p_-^2(k,s)\\&+\frac{1}{T(k)}\int_{-\infty}^{+\infty} ds\,\alpha^2(s)[U(s)-\tilde{U}(s)]p_-(k,s)p_-(-k,s)\\&-\frac{R^+(-k)}{T(-k)}\sum_{x_n\in S}v_n\left(\frac{\alpha}{\sigma}\right)^2(x_n)p_-^2(k,x_n)\\&+\frac{1}{T(k)}\sum_{x_n\in S}v_n\left(\frac{\alpha}{\sigma}\right)^2(x_n)p_-(k,x_n)p_-(-k,x_n)+O(\varepsilon^2).\end{aligned} \tag{2.55}$$

Using (2.6) and (2.49) we can simplify this formula to

$$2ik\left(\frac{1}{\tilde{T}(k)}-\frac{1}{T(k)}\right)=2ik\frac{\gamma-1}{T(k)}-\int_{-\infty}^{+\infty} ds\frac{\alpha^2(s)}{\sigma^2(s)}w(s)p_-(k,s)p_+(k,s)+O(\varepsilon^2). \tag{2.56}$$

Hence we obtain $2ik[\tilde{R}(k)-R(k)]/T^2(k)$ up to order $O(\varepsilon^3)$ if we multiply the right-hand side of (2.54) by

$$2-\gamma+(2ik)^{-1}T(k)\int_{-\infty}^{+\infty} ds\,\sigma^{-2}(s)\alpha^2(s)w(s)p_-(k,s)p_+(k,s)\,.$$

Since $\gamma(2-\gamma)$ is equal to 1 up to $O(\varepsilon^2)$, we finally obtain

$$\begin{aligned}2ik\frac{\widetilde{R}(k)-R(k)}{T^2(k)} &= -\int_{-\infty}^{+\infty} ds\,\frac{\alpha^2(s)}{\sigma^2(s)}\,\tilde{\sigma}^2(s)[U(s)-\widetilde{U}(s)]p_-^2(k,s)-\sum_{x_n\in S}\left(\frac{\alpha}{\sigma}\right)^2(x_n)\tilde{\sigma}(x_n^+)\tilde{\sigma}(x_n^-)\omega_n(k)p_-^2(k,x_n)\\ &\quad-\frac{T(k)}{2ik}\int_{-\infty}^{+\infty} ds\,\frac{\alpha^2(s)}{\sigma^2(s)}w(s)p_-^2(k,s)\int_{-\infty}^{+\infty} dt\,\alpha^2(t)w(t)p_-(k,t)p_+(k,t)\\ &\quad+\frac{1}{2ik}\int_{-\infty}^{+\infty}\frac{\alpha^2(s)}{\sigma^2(s)}\,w(s)\int_s^{\infty} dt\,\frac{\alpha^2(t)}{\sigma^2(t)}\,w(t)g(k,s,t)+O(\varepsilon^3)\,,\end{aligned} \tag{2.57}$$

where

$$g(k,s,t)=p_-(k,s)[p_-(-k,s)p_-(k,t)-p_-(-k,t)p_-(k,s)]p_-(k,t)\,.$$

Needless to say, the reader will observe that the Born approximation (2.42) could be derived from this further approximation. We have preferred to give a separate derivation of the Born approximation for the sake of clarity and simplicity.

III. THE HOMOGENIZATION PROBLEM

In practical descriptions of heterogeneous or discontinuous media one tries to replace them at least by "partially" homogeneous ones, i.e., either media whose parameter variations are everywhere smooth ("smooth media") or piecewise homogeneous media, with parameters that are constant inside the layers ("layered media"). These so-called (generalized) homogenizations, sometimes feasible only at the price of modifying the structure of the wave equation, are justified if the observed property of interest is *practically not modified.* Here this property is the function $R(k)$, and we wish to keep the structure of the wave equation.

Let us first observe that the exact result (which could be calculated from the recurrence formulas of Ref. 5) as well as the Born approximation (2.42) show that the asymptotic behavior of $\widetilde{R}(k)$ always gives evidence for the existence and nature of singular points. At a point x_n of S_3, for instance, if $\tilde{r}_n\neq 0$, the factors $(\alpha^2p'_-)(x_n)p_-(x_n)$ appearing in (2.21a) introduce in the right-hand side of (2.42) for the asymptotic behavior of $R(k)$ in the k plane a term $(2ik)\tilde{r}_n[p_-(k,x_n)]^2/\tilde{t}_n$, and hence an oscillating variation of $R(k)/[T(k)]^2$ at $\pm\infty$, which is the Born approximation for the modification of the exact almost-periodic structure of $R(k)$ at $\pm\infty$. If $\tilde{r}_n=0$, introducing a point x_n into S_3 yields a variation of the asymptotic behavior of $R(k)$, which goes to zero at ∞ like k^{-1}, and hence decreases more slowly than the asymptotic behavior due to *any* regular potential U. Hence if we talk of "exact" results, *no homogenization is ever possible.* Homogenization becomes possible if (a) we allow small errors on the measurements of the reflection coefficients, and (b) we "observe" $R(k)$ through a filter. Needless to say, these assumptions are physically sound in all problems. Now, suppose we start from a known problem characterized by U and singular data, and that a set of perturbations of the singular data is introduced, like in Sec. II, that is sufficiently small to allow Born approximation.

(1) A "continuous" homogenization is obtained if these perturbations may be replaced by a perturbation $\delta U=\widetilde{U}-U$ of the regular potential in such a way that the calculated reflection coefficients obtained by either perturbation cannot be distinguished from each other. The homogenization is useful only if δU is smooth enough. Hence, if we assume that a band filter has the effect of multiplying $\widetilde{R}-R=\delta R$ by a factor $\mathscr{F}(k)$ equal to 1 for $|k|<\Delta^{-1}$, and is rapidly vanishing for $|k|>\Delta^{-1}$, e.g., $\exp[-k^2\Delta^2]$, it is sound to assume that $\widetilde{U}-U=\delta U$ is such that for any $k\in\mathbb{R}$

$$\left|(\mathscr{F}(k)-1)\int_{-\infty}^{+\infty} ds[p_-(k,s)]^2\alpha^2(s)\delta U(s)\right|\lesssim\frac{1}{2}\varepsilon'\,, \tag{3.1}$$

where ε' is the error allowed on measurements. Thus a sufficient condition for "ε'-homogenizing" by means of δU is

$$\left|\int_{-\infty}^{+\infty} ds\,p_-^2(k,s)\alpha^2(s)\delta U(s)+\mathscr{F}(k)\sum_{x_n\in S}\alpha(x_n^+)\alpha(x_n^-)\omega_n(k)p_-^2(k,x_n)\right|\lesssim\tfrac{1}{2}\varepsilon'\,. \tag{3.2}$$

If the unperturbed problem is a continuous one, the inversion formula[8] for the transform

$$2ik\tilde{g}(k)=\int_{-\infty}^{+\infty} ds\,f_-^2(k,s)g(s)ds, \tag{3.3}$$

where f_- is the Jost solution, is simply

$$-\int_x^{\infty} g(t)dt=\pi^{-1}\int_{-\infty}^{+\infty} dk\,f_+^2(k,x)\tilde{g}(k)\,. \tag{3.4}$$

Since, in our previous reference,[8] these formulas were reobtained by algebraic transformation methods, it is very likely that they hold even if the background is not continuous, providing that f_- is replaced by αp_- and f_+ by αp_+. We shall not try to clarify this point in the present application; we will simply assume that the background is such that the inversion formulas hold and that $\delta U(x)$ is given by

$$\int_x^{\infty} \delta U(t)dt = (2i\pi)^{-1} \lim_{\varepsilon\to 0^+} \times \int_{-\infty}^{+\infty} \frac{dk}{k+i\varepsilon} p_+^2(k,x)\alpha^2(x)\mathscr{F}(k) \times \sum_{x_n\in S} \alpha(x_n^+)\alpha(x_n^-)\omega_n(k)p_-^2(k,x_n). \tag{3.5}$$

It is clear that the filtering effect of $\mathscr{F}(k)$ amounts to a convolution of the distribution of singularities by a smoothing function that washes out details of length smaller than Δ. To be more precise, let us study the simplest case, where the background is zero, so that $\alpha(x)p_\pm(k,x)$ is simply $\exp[\pm ikx]$, and where we introduce perturbing discontinuities to α (the "hardest" discontinuities) so that $\omega_n(k) = -2ik\,\tilde{r}_n/\tilde{t}_n$. Hence the formula (3.5) becomes [if $\mathscr{F}(k)$ is the filter mentioned above]

$$-\int_x^{\infty} \delta U(t)dt = \pi^{-1}\sum_n \frac{\tilde{r}_n}{\tilde{t}_n}\int_{-\infty}^{+\infty} \exp[2ik(x-x_n) - k^2\Delta^2]dk = \pi^{-1/2}\Delta^{-1}\sum_n \frac{\tilde{r}_n}{\tilde{t}_n}\exp\left[\frac{-(x-x_n)^2}{\Delta^2}\right], \tag{3.6}$$

which makes clear that the discontinuities are integrated over length intervals of order Δ. It follows that

$$-\delta U(x) = 2\pi^{-1/2}\Delta^{-3}\sum_n (x-x_n)\frac{\tilde{r}_n}{\tilde{t}_n} \times \exp\left[\frac{-(x-x_n)^2}{\Delta^2}\right]. \tag{3.7}$$

For small Δ, the right-hand side of (3.7) would approach the δ' distribution supposedly equivalent to the hard discontinuities—the convolution effects being clear in (3.7). Hence allowing a filter and resonable measurement errors enables a first order homogenization between strong discontinuities by means of a continuous potential—the main condition being that the discontinuities, after they are integrated on intervals of length Δ, yield a function of smooth variations.

(2) However, the usual way of processing data assumes that the parameters are constant between strong discontinuities. The underlying idea obviously is that between what could be called "structural discontinuities," reflection factors (and others) of a given sign are balanced by reflection factors (or others) of the opposite sign, with all of them being distributed in a random way. It clearly follows from our analysis that this "constant" behavior between strong discontinuities may hold only in peculiar cases, where the random perturbations are, so to say, "unbiased."

Thus, to clarify this point, let us again assume that, starting from a reference model, we establish small discontinuities at fixed locations x_n. Since a continuously differentiable impedance would correspond to the choice $\tilde{r}_n = \tilde{s}_n = 0$, the number of fixed locations can always be supposed large and dense enough to be sufficiently general. One may even choose a so-called Goupillaud model characterized by a short discretization length h and $x_n = nh$. Now we obtain from (2.57), up to the third order of the additional discontinuities,

$$\frac{\tilde{R}(k) - R(k)}{T^2(k)} = \sum_n \tilde{R}_n p_-^2(k,x_n) - T(k)\sum_n \tilde{R}_n p_-^2(k,x_n) \times \sum_j \tilde{R}_j p_-(k,x_j)p_+(k,x_j) + \sum_n \tilde{R}_n \sum_{x_p > x_n} \tilde{R}_p g(k,x_n,x_p), \tag{3.8}$$

where

$$\tilde{R}_n = -(2ik)^{-1}\frac{\alpha^2}{\sigma^2}(x_n)\tilde{\sigma}(x_n^+)\tilde{\sigma}(x_n^-)\omega_n(k), \tag{3.9}$$

$$g(k,x_n,x_p) = p_-(k,x_n)[p_-(-k,x_n)p_-(k,x_p) - p_-(-k,x_p)p_-(k,x_n)]p_-(k,x_p). \tag{3.10}$$

We assume nothing about $\tilde{R}_n$ except that there exists for each location x_n a positive measure $\rho_n(x)dx$ such that $\int_{-\infty}^{+\infty}\rho_n(x)dx = 1$, and that $\tilde{R}_n$ has a probability $\rho_n(x)dx$ of having its value between x and $x+dx$. In addition, as it has been stated above, we assume that the reference model was chosen in such a way that all biases are removed, i.e., $\int_{-\infty}^{+\infty} x\rho_n(x)dx = 0$. This means that a first order continuous partial homogenization has been done "intra" layers. It then follows from (3.8) that the expectation value of $\tilde{R}(k) - R(k)$ does not vanish in general, being equal to

$$-T^3(k)\sum_n p_-^3(k,x_n)p_+(k,x_n)\int_{-\infty}^{+\infty} x^2\rho_n(x)dx. \tag{3.11}$$

This result can still be important if the mean squares of discontinuities are important. Notice that if we start from a continuous model $[\sigma(x) = 1]$, and introduce only hard discontinuities ($\tilde{r}_n \neq 0$, $\tilde{s}_n = 0$), the $\omega_n(k)$'s are simply related to the $\tilde{r}_n$'s (from 2.21):

$$\omega_n(k) = \frac{\tilde{r}_n}{\tilde{t}_n}\left[\frac{\alpha'(x_n^+)}{\alpha(x_n^-)} + \frac{\alpha'(x_n^-)}{\alpha(x_n^+)} + 2\frac{(\alpha^2 p'_-)(x_n)/p_-(x_n)}{\alpha(x_n^-)\alpha(x_n^+)}\right], \tag{3.12}$$

where the factor of $\tilde{r}_n/\tilde{t}_n$ reduces to $-2ik$ if the reference model is zero ($\alpha = 1$). Hence it is easily seen that a large number of small discontinuities can give a significant contribution. In geophysical problems, a strong discontinuity corresponds to $r_n \sim 0.1$, a small one to $r_n \sim 0.01$, and a completely negligible one (according to authors) to $r_n \sim 0.001$. But 10^5 negligible randomly distributed discontinuities can make as much of a contribution as a strong one. If there is a bias, and discontinuities concur, this figure may reduce down to 10^2!

IV. CONCLUSION

As stated in the Introduction, this paper is the beginning of a series in which we try to understand the importance of discontinuities in the one- or several-dimensional impedance problem. The first provisional conclusion is that even in the one-dimensional case, while taking into account a filter and measurement errors in data interpretation, only a continuous (but not constant) homogenization between "non-negligible" discontinuities can be justified in general, and to be careful, it should include an appraisal of second order terms as we have outlined above.

ACKNOWLEDGMENTS

The authors are indebted to Professor B. DeFacio for his reading of the manuscript and several stylistic improvements.

This work was performed under the auspices of Unité Associeé au CNRS No. 040768, Recherche Cooperative sur Programme No. 080264, and was supported in part by DRET, Contract No. 85/1435.

[1] See, for example, O. H. Hald, "Discontinuous inverse eigenvalue problems," Commun. Pure Appl. Math. **37**, 539 (1984).

[2] R. G. Newton, "Inversion of reflection data for layered media: a review of exact methods," Geophys. J. Roy. Astron. Soc. **65**, 191 (1981).

[3] M. S. Howard. "Inversion of the first-order equation governing sound propagation in a layered medium," Ph.D. thesis, Indiana University, 1981; "Inversion scattering for a layered acoustic medium using the first-order equations of motion," Geophysics **48**, 163 (1983).

[4] P. C. Sabatier, "Theoretical considerations for inverse scattering," Radio Sci. **18**, 1 (1983).

[5] P. C. Sabatier, *Rational reflection coefficients in one-dimensional inverse scattering problem and applications. Conference on Inverse Scattering: Theory and Application*, edited by J. Bee Bednar, R. Redner, E. Robinson, and A. Weglein (SIAM Philadelphia, 1983), pp. 75–100.

[6] P. C. Sabatier, *Well-posed questions and exploration of the space of parameters in linear and nonlinear inversion. Inverse Problems of Acoustic and Elastic Waves*, edited by F. Santosa, Yih-Hsing Pao, W. W. Symes, and C. Holland (SIAM, Philadelphia, 1986), pp. 82–103.

[7] A. Degasperis and P. C. Sabatier, "Extension of the one dimensional scattering theory, and ambiguities," Inverse Probl. **3**, 73 (1987).

[8] See Ref. 6, pp. 91–93.

Algebras of distributions suitable for phase-space quantum mechanics. I

José M. Gracia-Bondía and Joseph C. Várilly
Escuela de Matemática, Universidad de Costa Rica, San José, Costa Rica

(Received 14 July 1986; accepted for publication 11 November 1987)

The twisted product of functions on $\mathbb{R}^{2N}$ is extended to a *-algebra of tempered distributions that contains the rapidly decreasing smooth functions, the distributions of compact support, and all polynomials, and moreover is invariant under the Fourier transformation. The regularity properties of the twisted product are investigated. A matrix presentation of the twisted product is given, with respect to an appropriate orthonormal basis, which is used to construct a family of Banach algebras under this product.

I. INTRODUCTION

This is the first of two papers whose aim is to give a rigorous formulation to the Weyl–Wigner–Groenewold–Moyal or phase-space approach to quantum mechanics of spinless, nonrelativistic particles. (In a future article, we will show how spin may be incorporated also in this formalism.) In recent years, this approach has received increasing attention.[1–5] However, much remains to be done to unify its different strands. On the one hand, much useful quantum physics can be done using the distribution functions in the sense of Wigner.[4] On the other hand, most mathematical attention has centered on the Weyl operator calculus.[6–9] As Groenewold[10] and Moyal[11] have shown, one can work with functions on the classical phase space only, in a self-contained way, using the "twisted product" concept. Similarly, the practitioners of "deformation theory"[2] have given a promising axiomatic basis for quantum mechanics, but with mathematical tools rather different from the usual functional-analytic methods of quantum theory.

We attempt here to establish a mathematically rigorous and physically manageable formulation for quantum mechanics in phase space. To obtain the right mathematical context, we must, for example, specify those pairs of functions whose twisted product may be formed; and a suitable function space should include as many observables of physical interest as possible. To include the basic observables of position and momentum, we must leave aside the algebra of bounded operators on a Hilbert space: in rigorizing the phase-space approach, one soon finds that it is useful to work with locally convex topological vector spaces that are not necessarily Banach spaces.

The paper is organized as follows. In Sec. II, we review the properties of the twisted product and convolution in the Schwartz space $\mathscr{S}(\mathbb{R}^{2N})$. In Sec. III, we dualize these notions to the case where one or both factors are tempered distributions, and identify the Moyal algebra $\mathscr{M}$, that is, the largest *-algebra of distributions where these operations are defined and associative. We show that $\mathscr{M}$ is invariant under Fourier transformations. In Sec. IV, we consider the regularity properties of the twisted product and convolution, and show that distributions of compact support belong to the Moyal algebra. In Sec. V, we construct an orthonormal basis in $\mathscr{S}(\mathbb{R}^{2N})$; using this basis, we show that the twisted product may be presented as a matrix product of double sequences. As a consequence, we construct a net of Sobolev-like spaces of tempered distributions, some of which are Banach algebras with respect to the twisted product; these permit a more detailed examination of the Moyal algebra $\mathscr{M}$.

II. THE ALGEBRAS $(\mathscr{S}_2,\times)$ and $(\mathscr{S}_2,\Diamond)$

Throughout this paper, we work with certain spaces of functions and distributions over $\mathbb{R}^{2N}$, regarded as the phase space $T^*(\mathbb{R}^N)$. For $u, v \in \mathbb{R}^{2N}$, we write $u'v$ and $u'Jv$ for the ordinary and symplectic scalar products of u and v. Choosing and fixing an orthonormal symplectic basis for $\mathbb{R}^{2N}$, we write

$$u = (u_1,u_2,\dots,u_{2N}) = (q_1,\dots,q_N,p_1,\dots,p_N)$$

and

$$v = (v_1,v_2,\dots,v_{2N}) = (\tilde q_1,\dots,\tilde q_N,\tilde p_1,\dots,\tilde p_N),$$

where explicitly

$$u'v := \sum_{i=1}^{N} (q_i\tilde q_i + p_i\tilde p_i), \quad u'Jv := \sum_{i=1}^{N} (q_i\tilde p_i - p_i\tilde q_i),$$

where J is the matrix $\begin{pmatrix} 0 & I_N \\ -I_N & 0 \end{pmatrix}$ in the chosen basis. Note that $v'Ju = -u'Jv$ and $u'Ju = 0$.

We define $\mathscr{S}_2 := \mathscr{S}(\mathbb{R}^{2N})$ as the Schwartz space of smooth rapidly decreasing functions on $\mathbb{R}^{2N}$. If $f \in \mathscr{S}_2$, $s \in \mathbb{R}^{2N}$, and $1 \leqslant j \leqslant 2N$, we define $f^*(u) := \overline{f(u)}$, $\check f(u) := f(-u)$, and also

$$(\mu_j f)(u) := u_j f(u), \quad \partial_j f := \frac{\partial f}{\partial u_j},$$

$$(\tau_s f)(u) := f(u-s), \quad (\epsilon_s f)(u) := e^{is'Ju} f(u),$$

and

$$\hat\partial_j f := \begin{cases} \partial_{j+N} f, & \text{if } 1 \leqslant j \leqslant N, \\ -\partial_{j-N} f, & \text{if } N < j \leqslant 2N. \end{cases}$$

We make three normalizations which are a little unconventional. First, for integrals over $\mathbb{R}^{2N}$ we use the Haar measure $dx := (2\pi)^{-N} d^{2N}x$, where $d^{2N}x$ is the Lebesgue measure. (This gets rid of powers of 2π in Fourier transforms.[12]) In particular, $\int e^{-x^2/2}\,dx = 1$. Second, we use the *bilinear* form

$$\langle f,g \rangle := \int f(x) g(x)\,dx$$

and the *sesquilinear* form

 0022-2488/88/040869-11$02.50

$$(f|g) := 2^{-N}\langle f^*, g\rangle = 2^{-N}\int \overline{f(x)}g(x)\,dx$$

whenever the integrals converge. For $f\in L^2(\mathbb{R}^{2N})$, we will use the norm $\|f\| := (f|f)^{1/2}$. Third, for Planck's constant we take $\hbar = 2$ (rather than the usual $\hbar = 1$).

We define an ordinary Fourier transform $\mathscr{F}$ and two symplectic Fourier transforms[13] F and $\breve{F}$ by

$$(\mathscr{F}f)(u) := \int f(t)e^{-it'u}\,dt,$$

$$(Ff)(u) := \int f(t)e^{-it'Ju}\,dt,$$

$$(\breve{F}f)(u) := \int f(t)e^{it'Ju}\,dt.$$

The transforms $\mathscr{F}$, F, and $\breve{F}$ are commuting isomorphisms (of Fréchet spaces) of $\mathscr{S}_2$ onto $\mathscr{S}_2$, and satisfy the following formulas:

$$Ff(u) = \mathscr{F}f(Ju), \quad \breve{F}f(u) = \mathscr{F}f(-Ju),$$
$$F^2 = \breve{F}^2 = \mathrm{Id}, \quad \breve{F}f = (Ff)^\vee = F(\check{f}), \quad (Ff)^* = \breve{F}(f^*),$$
$$F(\tau_s f) = \epsilon_{-s}Ff, \quad F(\epsilon_s f) = \tau_{-s}Ff,$$
$$F(\hat{\partial}_j f) = -i\mu_j Ff, \quad F(\mu_j f) = i\hat{\partial}_j Ff,$$
$$\langle Ff, g\rangle = \langle f, Fg\rangle, \quad (Ff|g) = (f|Fg).$$

Definition 1: If $f,g\in\mathscr{S}_2$, the *twisted product* $f\times g$ is defined by

$$(f\times g)(u) := \iint f(v)g(w)\times\exp(i(u'Jv + v'Jw + w'Ju))\,dv\,dw$$
$$= \iint f(v+s)g(u+t)e^{is'Jt}\,ds\,dt. \qquad (1)$$

The *twisted convolution* $f\lozenge g$ is defined by

$$(f\lozenge g)(u) := \int f(u-t)g(t)e^{-iu'Jt}\,dt. \qquad (2)$$

Remarks: (1) It was von Neumann[14] who introduced the twisted convolution (although he gave it no name) in order to establish the uniqueness of the Schrödinger representation. It has been used by Kastler and others[13,15,16] to study the canonical commutation relations.

(2) The twisted product is nothing but the Weyl functional calculus[7] seen from another point of view. As in Ref. 1 and elsewhere, one may regard it as $f\times g = \mathscr{F}^{-1}(\mathscr{F}f\lozenge\mathscr{F}g)$; but perhaps a more natural motivation is the following. The pointwise product $f(u)g(u)$ is not suitable for quantum mechanics since the uncertainty principle forbids localization at a point in phase space. Following Sławianowski,[17] we seek to replace it by some other product that is translation and symplectic equivariant, associative, and nonlocal. In Ref. 3 it is shown that the only integral kernels satisfying translation and symplectic equivariance and associativity are $a\delta(s)\delta(t)$ (for the pointwise product) and $be^{ics'Jt}$, where a, b, and c are constants which we may set equal to 1.

Proposition 1: If $f,g\in\mathscr{S}_2$, then $f\times g\in\mathscr{S}_2$, $(f,g)\mapsto f\times g$ is a continuous bilinear operation on $\mathscr{S}_2$, and

$$\partial_j(f\times g) = \partial_j f\times g + f\times\partial_j g,$$
$$\mu_j(f\times g) = f\times\mu_j g + i\hat{\partial}_j f\times g = \mu_j f\times g - if\times\hat{\partial}_j g. \qquad (3)$$

Proof: The Leibniz formula follows by differentiating (1) under the integral sign, and (3) is a straightforward calculation. By induction on these formulas, $f\times g$ lies in $\mathscr{S}_2$. If $\alpha = (\alpha_1,\ldots,\alpha_{2N})\in\mathbb{N}^{2N}$, we write $\partial^\alpha = \partial_1^{\alpha_1}\cdots\partial_{2N}^{\alpha_{2N}}$ and similarly define μ^α, $\hat{\partial}^\alpha$. Then

$$\mu^\alpha\partial^\gamma(f\times g) = \sum_{\beta\leqslant\alpha}\sum_{\epsilon\leqslant\gamma}(-i)^{|\beta|}\binom{\alpha}{\beta}\binom{\gamma}{\epsilon}\times\mu^{\alpha-\beta}\partial^{\gamma-\epsilon}f\times\hat{\partial}^\beta\partial^\epsilon g.$$

From (1) we get $\|f\times g\|_\infty\leqslant\|f\|_1\|g\|_1$. Since the topology of $\mathscr{S}_2$ is given by the seminorms $p_{\alpha\gamma}(f) := \|\mu^\alpha\partial^\gamma f\|_\infty$ or by $q_{\alpha\gamma}(f) := \|\mu^\alpha\partial^\gamma f\|_1$, the estimates

$$p_{\alpha\gamma}(f\times g)\leqslant\sum_{\beta\leqslant\alpha}\sum_{\epsilon\leqslant\gamma}\binom{\alpha}{\beta}\binom{\gamma}{\epsilon}q_{\alpha-\beta,\gamma-\epsilon}(f)q_{0,\eta+\epsilon}(g),$$

with $\eta_j = \beta_{j\pm N}$ for all j, show that $(f,g)\mapsto f\times g$ is jointly continuous for the topology of $\mathscr{S}_2$. ■

The various Fourier transforms intertwine $\times$ and $\lozenge$, just as with "ordinary" products and convolutions. In fact, even more is true: by applying a symplectic Fourier transform to one side only, we can interchange the operations $\times$ and $\lozenge$. This allows us to work with the operation most convenient to any particular calculation, transferring the result to the other one by Fourier invariance of $\mathscr{S}_2$. Explicitly, we find

$$f\times g = Ff\lozenge g = f\lozenge\breve{F}g, \quad f\lozenge g = Ff\times g = f\times\breve{F}g, \qquad (4)$$

since, for example,

$$(f\times g)(u) = \int\int f(v)g(w)e^{-iv'J(u-w)}e^{iw'Ju}\,dv\,dw$$
$$= \int Ff(u-w)g(w)e^{-iu'Jw}\,dw = (Ff\lozenge g)(u).$$

We also find

$$\mathscr{F}(f\times g) = \mathscr{F}f\lozenge\mathscr{F}g, \quad \mathscr{F}(f\lozenge g) = \mathscr{F}f\times\mathscr{F}g, \qquad (5)$$

and exactly analogous formulas with $\mathscr{F}$ replaced by F or $\breve{F}$. Also,

$$(f\times g)\times h = f\times(g\times h), \quad (f\lozenge g)\lozenge h = f\lozenge(g\lozenge h), \qquad (6)$$

since

$$((f\lozenge g)\lozenge h)(u)$$
$$= \iint f(u-t-s)g(s)h(t)e^{-i(u'Jt+(u-t)'Js)}\,ds\,dt$$
$$= \iint f(u-v)g(v-t)h(t)e^{-i(u'Jv-t'Jv)}\,dt\,dv$$
$$= (f\lozenge(g\lozenge h))(u)$$

and applying (5) yields the associativity of $\times$. Next,

$$(f\times g)^* = g^*\times f^*, \quad (f\lozenge g)^* = g^*\lozenge f^*,$$

since, for instance,

$$(f\times g)^*(u)=\iint f^*(u+s)g^*(u+t)e^{-is'Jt}\,ds\,dt$$
$$=\iint g^*(u+t)f^*(u+s)e^{it'Js}\,dt\,ds$$
$$=(g^*\times f^*)(u).$$

A fact of fundamental importance is the following identity.

Proposition 2: If $f,g\in\mathscr{S}_2$, then

$$\int (f\times g)(u)du=\int (g\times f)(u)du=\int f(u)g(u)du. \tag{7}$$

Proof:

$$\int (f\times g)(u)du=\mathscr{F}(f\times g)(0)=(\mathscr{F}f\lozenge\mathscr{F}g)(0)$$
$$=\int \mathscr{F}f(-t)\mathscr{F}g(t)dt$$
$$=(\mathscr{F}f*\mathscr{F}g)(0)=\mathscr{F}(fg)(0)$$
$$=\int f(u)g(u)du,$$

where $*$ denotes ordinary convolution. ■

The cyclicity inherent in the tracial identity (7) is what allows us to push through the extension via duality. We note an important consequence of (7).

Proposition 3: If $f, g, h\in\mathscr{S}_2$, then

$$\langle f\times g,h\rangle=\langle f,g\times h\rangle=\langle g,h\times f\rangle, \tag{8}$$
$$\langle f\lozenge g,h\rangle=\langle f,\check{g}\lozenge h\rangle=\langle g,h\lozenge\check{f}\rangle, \tag{9}$$
$$(h\,|\,f\times g)=(f^*\times h\,|g)=(h\times g^*|\,f).$$

Proof: From (7) and the associativity (6), we find that all three expressions in (8) equal $\int(f\times g\times h)(u)du$. Now (9) follows from (4), and the third formula is immediate. ■

III. DUALITY AND THE MOYAL ALGEBRA

Having established a calculus for functions in $\mathscr{S}_2$ with twisted product and convolution, we now extend it to a larger algebra of tempered distributions. First we consider the twisted product or convolution of a tempered distribution and a test function.

For $T\in\mathscr{S}_2'$, $h\in\mathscr{S}_2$, we write $\langle T,h\rangle:=T(h)$. We also extend our previous notation in the usual way:

$$\langle\mathscr{F}T,h\rangle:=\langle T,\mathscr{F}h\rangle,\quad \langle\check{T},h\rangle:=\langle T,\check{h}\rangle,$$
$$(T\,|h):=2^{-N}\langle T^*,h\rangle:=2^{-N}\langle T,h^*\rangle^*,$$
$$\langle FT,h\rangle:=\langle T,\check{F}h\rangle,\quad \langle\check{F}T,h\rangle:=\langle T,Fh\rangle,$$
$$\langle\partial_jT,h\rangle:=-\langle T,\partial_jh\rangle,\quad \langle\mu_jT,h\rangle:=\langle T,\mu_jh\rangle.$$

Definition 2: For $T\in\mathscr{S}_2'$, $f,h\in\mathscr{S}_2$, we define $T\times f$, $f\times T$, $T\lozenge f$, $f\lozenge T$ in $\mathscr{S}_2'$ by

$$\begin{aligned}&\langle T\times f,h\rangle:=\langle T,f\times h\rangle,\quad \langle f\times T,h\rangle:=\langle T,h\times f\rangle,\\&\langle T\lozenge f,h\rangle:=\langle T,\check{f}\lozenge h\rangle,\quad \langle f\lozenge T,h\rangle:=\langle T,h\lozenge\check{f}\rangle.\end{aligned} \tag{10}$$

The continuity of $\times$ and $\lozenge$ in $\mathscr{S}_2$ implies that each right-hand side is continuous and linear in h and thus defines an element of $\mathscr{S}_2'$. By (8) and (9), these are extensions of the corresponding operations on $\mathscr{S}_2$.

Throughout this paper, every dual space E' of a locally convex space E is topologized by the strong dual topology, that of uniform convergence on bounded subsets of E. Then the four bilinear maps: $\mathscr{S}_2'\times\mathscr{S}_2\to\mathscr{S}_2'$ defined above are hypocontinous. Indeed, since $\mathscr{S}_2$ and $\mathscr{S}_2'$ are barreled it suffices[18,19] to check separate continuity. For example, for fixed T, the map $f\mapsto\langle T,f\times h\rangle$ is continuous, uniformly so for h in a bounded subset of $\mathscr{S}_2$ (by the joint continuity of f and h), so that $f\mapsto T\times f$ is continuous from $\mathscr{S}_2$ to $\mathscr{S}_2'$. For fixed f, $T\mapsto T\times f$ is the transpose of the continuous map $h\mapsto f\times h$ of $\mathscr{S}_2$ into $\mathscr{S}_2$, and as such is continuous from $\mathscr{S}_2'$ to $\mathscr{S}_2'$.

All formulas of Sec. II involving f and g extend to analogous formulas for T and f (e.g., $T\times f=FT\lozenge f=T\lozenge\check{F}f$). This is easily checked since $A=B$ in $\mathscr{S}_2'$ iff $\langle A,h\rangle=\langle B,h\rangle$ for all $h\in\mathscr{S}_2$, and we may reduce to the $\mathscr{S}_2$ case using (10).

We write $\mathbf{1}$ for the constant function with value 1, and δ for the Dirac measure of mass 1 supported at 0. These are the identities for the operations $\times$ and $\lozenge$:

$$\mathbf{1}\times f=f\times\mathbf{1}=f,\quad \delta\lozenge f=f\lozenge\delta=f.$$

This follows from $\langle\mathbf{1}\times f,h\rangle=\langle\mathbf{1},f\times h\rangle=\int(f\times h)(u)du=\langle f,h\rangle$ by (7): thus $\mathbf{1}\times f=f$ as elements of $\mathscr{S}_2'$. We show below that $\mathbf{1}\times f$ is continuous, so $\mathbf{1}\times f=f$ as functions in $\mathscr{S}_2$. The other half of the equation follows from (5), since $F\delta=\check{F}\delta=\mathscr{F}\delta=\mathbf{1}$. From (4) we also obtain the formulas

$$\mathbf{1}\lozenge f=\delta\times f=\check{F}f,\quad f\lozenge\mathbf{1}=f\times\delta=Ff. \tag{11}$$

Using the fact that $u_j=\mu_j\mathbf{1}$, we obtain from (3) the important identities

$$u_j\times f=\mu_jf+i\,\hat{\partial}_jf,\quad f\times u_j=\mu_jf-i\,\hat{\partial}_jf,$$

which, in the (q,p) notation, become

$$\begin{aligned}&q_j\times f=\Big(q_j+i\frac{\partial}{\partial p_j}\Big)f,\quad p_j\times f=\Big(p_j-i\frac{\partial}{\partial q_j}\Big)f,\\&f\times q_j=\Big(q_j-i\frac{\partial}{\partial p_j}\Big)f,\quad f\times p_j=\Big(p_j+i\frac{\partial}{\partial q_j}\Big)f.\end{aligned} \tag{12}$$

If $T\in\mathscr{S}_2'$ and $f\in\mathscr{S}_2$, the ordinary convolution $T*f$ is[18,20] a smooth function in $\mathscr{O}_C$, whereas the ordinary product Tf is a "rapidly decreasing distribution" in $\mathscr{O}_C'$ but need not be smooth. In contrast, (5) (extended to $\mathscr{S}_2'$) shows that the twisted product and twisted convolution have similar properties of smoothness and of growth at infinity. To see this, we first note that (2) may be rewritten as

$$(f\lozenge g)(u)=\langle\epsilon_{-u}\tau_u\check{f},g\rangle=\langle f,\epsilon_u\tau_u\check{g}\rangle.$$

Thus in convolution formulas, such as $(T*f)(u)=\langle T,\tau_u\check{f}\rangle$, the translations τ_u are replaced by $\epsilon_u\tau_u$ or $\epsilon_{-u}\tau_u$.

Theorem 1: If $T\in\mathscr{S}_2'$, $f\in\mathscr{S}_2$, then $T\times f$, $f\times T$, $T\lozenge f$, and $f\lozenge T$ are smooth functions on $\mathbb{R}^{2N}$, given by

$$\begin{aligned}&(T\times f)(u)=\langle T,\epsilon_u\tau_uFf\rangle,\\&(f\times T)(u)=\langle T,\epsilon_{-u}\tau_u\check{F}f\rangle,\end{aligned} \tag{13}$$
$$\begin{aligned}&(T\lozenge f)(u)=\langle T,\epsilon_u\tau_u\check{f}\rangle,\\&(f\lozenge T)(u)=\langle T,\epsilon_{-u}\tau_u\check{f}\rangle.\end{aligned} \tag{14}$$

Proof: If $h\in\mathscr{S}_2$, the maps $u\mapsto\tau_uh$, $u\mapsto\epsilon_uh$ are continuous from $\mathbb{R}^{2N}$ to $\mathscr{S}_2$, so the right-hand sides of these formulas are jointly continuous in u and f. By transposition, since $\langle T,\epsilon_u\tau_u \check{f}\rangle = \langle \epsilon_{-u}\tau_u \check{T}, f\rangle$, they are also continuous in T. These right-hand sides define separately continuous extensions to $\mathscr{S}'_2 \times \mathscr{S}_2$ of the twisted product and convolution on the dense subspace $\mathscr{S}_2 \times \mathscr{S}_2$; since these extensions are necessarily unique, they coincide with $T \times f$, etc., as defined earlier.

Now

$$\partial_j h(u) = \pm \lim_{c\to 0} c^{-1}(h(u+ce_j) - h(u)),$$

where e_k is the k th basis vector in $\mathbb{R}^{2N}$, whenever this limit exists. By calculation, we find that

$$\begin{aligned}&\lim_{c\to 0} c^{-1}\langle T,\epsilon_{u+ce_j}\tau_{u+ce_j}Ff - \epsilon_u\tau_u Ff\rangle \\ &\quad = (\partial_j T \times f)(u) + (T \times \partial_j f)(u)\end{aligned}$$

as expected, so by induction $T \times f$ is infinitely differentiable, with the Leibniz formula $\partial_j(T\times f) = \partial_j T \times f + T \times \partial_j f$ holding as an equality between smooth functions. The other three cases are similar. ■

Let $\mathscr{E}_2$ denote the space of smooth functions on $\mathbb{R}^{2N}$, with the topology of uniform convergence of all derivatives on compact sets. Ordinary convolution operators are precisely those that commute with translations; we may characterize the twisted convolution operators as those that commute with "twisted translations."

Theorem 2: Let $L: \mathscr{S}_2 \to \mathscr{E}_2$ be linear and continuous; then there is a unique $T\in\mathscr{S}'_2$ with $L(f) = T\times f$ for all $f\in\mathscr{S}_2$ iff L commutes with $\{\epsilon_u\tau_u : u\in\mathbb{R}^{2N}\}$.

Proof: From (14) we find that

$$\begin{aligned}\epsilon_v\tau_v(T\times f)(u) &= e^{iv'Ju}(T\times f)(u-v)\\ &= e^{iv'Ju}\langle T,\epsilon_{u-v}\tau_{u-v}Ff\rangle\\ &= e^{iv'Ju}\langle T, t\mapsto e^{i(u-v)'Jt}Ff(t-u+v)\rangle\\ &= \langle T, t\mapsto e^{i(u'Jt - v'J(t-u))}\epsilon_{-v}Ff(t-u)\rangle\\ &= \langle T, t\mapsto e^{iu'Jt}\tau_{-v}\epsilon_{-v}Ff(t-u)\rangle\\ &= \langle T,\epsilon_u\tau_u F(\epsilon_v\tau_v f)\rangle,\end{aligned}$$

so that $f\mapsto T\times f$ commutes with any $\epsilon_v\tau_v$.

On the other hand, given $L: \mathscr{S}_2\to\mathscr{E}_2$ which commutes with all $\epsilon_v\tau_v$, we define $T\in\mathscr{S}'_2$ by $\langle T,h\rangle := L(Fh)(0)$. (So T is unique.) For $u\in\mathbb{R}^{2N}$, $f\in\mathscr{S}_2$, we then obtain

$$\begin{aligned}(T\times f)(u) &= \langle T,\epsilon_u\tau_u Ff\rangle = \langle T, F(\epsilon_{-u}\tau_{-u}f)\rangle\\ &= L(\epsilon_{-u}\tau_{-u}f)(0) = \epsilon_{-u}\tau_{-u}(Lf)(0)\\ &= \tau_{-u}(Lf)(0) = (Lf)(u).\end{aligned}$$ ■

We can now define the Moyal *-algebra $\mathscr{M}$. We define it as the intersection of two spaces $\mathscr{M}_L$ and $\mathscr{M}_R$, which, to the best of our knowledge, were first considered by Antonets.[21]

Definition 3:

(1) $\mathscr{M}_L := \{\mathscr{S}\in\mathscr{S}'_2 : S\times f\in\mathscr{S}_2,\ \text{for all } f\in\mathscr{S}_2\}$;

(2) $\mathscr{M}_R := \{R\in\mathscr{S}'_2 : f\times R\in\mathscr{S}_2,\ \text{for all } f\in\mathscr{S}_2\}$;

(3) $\mathscr{M} := \mathscr{M}_L\cap\mathscr{M}_R$.

Note that $S\in\mathscr{M}_L$ iff $S^*\in\mathscr{M}_R$ since $(S\times f)^* = f^*\times S^*$. Since $\mathscr{S}_2$ is a Fréchet space, the maps $f\mapsto S\times f$, $f\mapsto f\times R$ are continuous from $\mathscr{S}_2$ to $\mathscr{S}_2$ by the closed graph theorem.

It is clear that $\mathscr{S}_2\subset\mathscr{M}$ and that $1\in\mathscr{M}$. Formulas (11) show that $\delta\in\mathscr{M}$. Now, if $S\in\mathscr{M}_L$, $f\in\mathscr{S}_2$, we have

$$\begin{aligned}\partial_j S\times f &= \partial_j(S\times f) - S\times\partial_j f,\\ \mu_j S\times f &= \mu_j(S\times f) + iS\times\hat{\partial}_j f,\end{aligned}$$

and so $\partial_j S\in\mathscr{M}_L$ and $\mu_j S\in\mathscr{M}_L$; thus $\mathscr{M}_L$, and similarly $\mathscr{M}_R$ and $\mathscr{M}$, is closed under partial differentiation and multiplication by polynomials. Hence, in particular, all polynomials lie in $\mathscr{M}$.

We now extend the twisted product to the case of one distribution in $\mathscr{M}$ and one in $\mathscr{S}'_2$ (so that $\mathscr{S}'_2$ is an $\mathscr{M}$ bimodule).

Definition 4: If $R\in\mathscr{M}_R$, $S\in\mathscr{M}_L$, $T\in\mathscr{S}'_2$, we define $T\times S$, $R\times T$ in $\mathscr{S}'_2$ by

$$\langle T\times S,h\rangle := \langle T,S\times h\rangle,\quad \langle R\times T,h\rangle := \langle T,h\times R\rangle, \tag{15}$$

for all $h\in\mathscr{S}_2$. Since the right-hand sides are continuous in h, $T\times S$ and $R\times T$ are defined in $\mathscr{S}'_2$.

If $R,S\in\mathscr{M}$, $T\in\mathscr{S}'_2$, and $f,g,h\in\mathscr{S}_2$, we may compute

$$\begin{aligned}\langle (T\times f)\times g,h\rangle &= \langle T\times f, g\times h\rangle = \langle T, f\times g\times h\rangle\\ &= \langle T\times(f\times g),h\rangle,\\ \langle (R\times S)\times f,h\rangle &= \langle R\times S, f\times h\rangle = \langle R, S\times f\times h\rangle\\ &= \langle R\times(S\times f),h\rangle.\end{aligned}$$

In particular, $(R\times S)\times f\in\mathscr{S}_2$ for $f\in\mathscr{S}_2$, so $R\times S\in\mathscr{M}_L$. Then $\langle (T\times R)\times S,h\rangle = \langle T\times R, S\times h\rangle = \langle T, R\times S\times h\rangle = \langle T\times(R\times S),h\rangle$. We conclude that $\mathscr{M}$ is an associative algebra; in fact, it is a *-algebra since, for $R,S\in\mathscr{M}$,

$$\begin{aligned}\langle (R\times S)^*,h\rangle &= \langle R\times S,h^*\rangle^* = \langle R, S\times h^*\rangle^*\\ &= \langle R^*, h\times S^*\rangle = \langle S^*\times R^*,h\rangle.\end{aligned}$$

We may also note that since $S\lozenge f = S\times\check{F}f$, $f\lozenge R = Ff\times R$, we have

$$\begin{aligned}\mathscr{M}_L &= \{S\in\mathscr{S}'_2 : S\lozenge f\in\mathscr{S}_2,\ \forall f\in\mathscr{S}_2\},\\ \mathscr{M}_R &= \{R\in\mathscr{S}'_2 : f\lozenge R\in\mathscr{S}_2,\ \forall f\in\mathscr{S}_2\},\end{aligned}$$

so $\mathscr{M}$ is also a *-algebra under $\lozenge$, where we define $\langle T\lozenge S,h\rangle := \langle T,\check{S}\lozenge h\rangle$, $\langle R\lozenge T,h\rangle := \langle T,h\lozenge\check{R}\rangle$. The invariance of $\mathscr{M}$ under the several Fourier transforms now follows easily. One easily checks that the formulas of Sec. II remain valid when f,g are replaced by $R,S\in\mathscr{M}$.

Remarks: (1) We show below that $\{f\times g: f,g\in\mathscr{S}_2\}$ equals $\mathscr{S}_2$. Thus $\mathscr{M}$ is the maximal *-algebra which we may define by duality. For, if $T\in\mathscr{S}'_2$ with $T\times f$, $f\times T\in\mathscr{M}$, $\forall f\in\mathscr{S}_2$, then by writing $f = g\times h$ we see that $T\times f$, $f\times T$ both lie in $\mathscr{S}_2$, since $T\times f = (T\times g)\times h$ and $f\times T = g\times(h\times T)$; hence $T\in\mathscr{M}$.

(2) $\mathscr{M}$, $\mathscr{M}_L$, $\mathscr{M}_R$, and $\mathscr{S}'_2$ are distinct spaces of distributions.

IV. REGULARITY PROPERTIES

In this section we consider in more detail the growth conditions on resultants of twisted products or convolutions. We identify a space of smooth functions, $\mathscr{O}_T$, which contains all functions defined by (13) and (14), and we show that $\mathscr{O}_T$ is a normal space of distributions. As a consequence its dual space $\mathscr{O}'_T$ contains all distributions of compact support and is contained in $\mathscr{M}$.

If $(E_i)_{i\in I}$ is a collection of locally convex spaces, the *projective topology* on the intersection $E := \bigcap_{i\in I} E_i$ is the weakest locally convex topology such that all inclusions $E\subset E_i$ are continuous. The *inductive topology* on the union $F := \bigcup_{i\in I} E_i$ is the strongest locally convex topology such that all inclusions $E_i\subset F$ are continuous. We shall use the projective topology on decreasing intersections, and the inductive topology on increasing unions, without further comment.

For $f\in C^m(\mathbb{R}^{2N})$, $k,m\in\mathbb{N}$, let

$$p_{k,m}(f) := \sup\{(1+u^2)^{-k-|\alpha|/2}|\partial^\alpha f(u)|:\ u\in\mathbb{R}^{2N},\ |\alpha|\leqslant m\} \tag{16}$$

(where $u^2 = u'u = u_1^2 + \cdots + u_{2N}^2$), and let $\mathscr{V}_k^m$ be the space of all $f\in C^m$ such that $(1+u^2)^{-k-|\alpha|/2}\,\partial^\alpha f(u)$ vanishes at infinity for all $|\alpha|\leqslant m$, normed by $p_{k,m}$. Now let

$$\mathscr{V}_k := \bigcap_{m\in\mathbb{N}} \mathscr{V}_k^m,\quad \mathscr{O}_T := \bigcup_{k\in\mathbb{N}} \mathscr{V}_k. \tag{17}$$

Let

$$q_{k,m}(f) := \sup\{(1+u^2)^{-k}|\partial^\alpha f(u)|:\ u\in\mathbb{R}^{2N},\ |\alpha|\leqslant m\};$$

then

$$\{f\in C^m:\ (1+u^2)^{-k}\,\partial^\alpha f(u)\ \text{vanishes at infinity for}\ |\alpha|\leqslant m\},$$

normed by $q_{k,m}$, is Horváth's space[18] $\mathscr{S}^m_{-k}$. Using the notations of Ref. 18, we have

$$\mathscr{S}_{-k} := \bigcap_{m\in\mathbb{N}} \mathscr{S}^m_{-k},\quad \mathscr{O}_C := \bigcup_{k\in\mathbb{N}} \mathscr{S}_{-k},$$
$$\mathscr{O}_C^m := \bigcup_{k\in\mathbb{N}} \mathscr{S}^m_{-k},\quad \mathscr{O}_M := \bigcap_{m\in\mathbb{N}} \mathscr{O}_C^m. \tag{18}$$

Here $\mathscr{O}_M$ consists of smooth functions for which each derivative is polynomially bounded; $\mathscr{O}_C$ is a subset of $\mathscr{O}_M$, wherein the degree of the polynomial bound is independent of the derivative; and $\mathscr{O}_T$ is the subset of $\mathscr{O}_M$ wherein that degree increases linearly with the order of the derivative. Indeed, this leads to the following proposition.

Proposition 4: $\mathscr{O}_C\subset\mathscr{O}_T\subset\mathscr{O}_M$ with continuous inclusions.

The proof is easy and will be omitted.

Theorem 3: If $T\in\mathscr{S}'_2$, $f\in\mathscr{S}_2$, then $T\times f$, $f\times T$, $T\Diamond f$, and $f\Diamond T$ all lie in $\mathscr{O}_T$. Moreover, these four bilinear maps of $\mathscr{S}'_2\times\mathscr{S}_2$ into $\mathscr{O}_T$ are separately continuous.

Proof: It suffices to consider the case of $T\Diamond f$.

Differentiating the equality $\epsilon_u\tau_u(T\Diamond f) = T\Diamond(\epsilon_u\tau_u f)$, with $u = tJe_j$, at $t=0$, we get

$$(\hat\partial_j + i\mu_j)(T\Diamond f) = T\Diamond(\hat\partial_j + i\mu_j)f,$$

so that

$$\hat\partial_j(T\Diamond f) = T\Diamond(\hat\partial_j + i\mu_j)f - i\mu_j(T\Diamond f),$$

and by induction we get, for $\alpha\in\mathbb{N}^{2N}$,

$$\hat\partial^\alpha(T\Diamond f)(u) = \sum_{\beta\leqslant\alpha} P_\beta(u)(T\Diamond f_\beta)(u), \tag{19}$$

where $P_\beta(u)$ is a polynomial of degree at most $|\beta|$, and $f_\beta\in\mathscr{S}_2$, for $\beta\leqslant\alpha$. From (16), we need only show that $T\Diamond f$ is polynomially bounded.

Any $T\in\mathscr{S}'_2$ can be written[20] as $T = \hat\partial^\gamma Q$, with $\gamma\in\mathbb{N}^{2N}$, where Q is a polynomially bounded continuous function on $\mathbb{R}^{2N}$. Recalling (3), $\hat\partial_j Q\Diamond f = \hat\partial_j(Q\Diamond f) + i\mu_j f$, and by iteration

$$T\Diamond f = \hat\partial^\gamma Q\Diamond f = \sum_{\epsilon\leqslant\gamma} \hat\partial^\epsilon(Q\Diamond g_\epsilon), \tag{20}$$

for certain $g_\epsilon\in\mathscr{S}_2$. Combining this with (19), we need only show that $Q\Diamond f$ is polynomially bounded.

If $|Q(t)|\leqslant C(1+t^2)^k$, then

$$\begin{aligned}|(Q\Diamond f)(u)| &= \left|\int Q(t)e^{iu'Jt}f(u-t)\,dt\right| \\ &\leqslant \int C(1+t^2)^k|f(u-t)|\,dt \\ &\leqslant \int 2^kC(1+u^2)^k(1+(u-t)^2)^k|f(u-t)|\,dt \\ &= K(1+u^2)^k,\end{aligned}$$

where

$$K = 2^kC\int(1+s^2)^k|f(s)|\,ds$$

is finite since $f\in\mathscr{S}_2$.

Since K depends continuously on f, the map $f\mapsto Q\Diamond f$ is continuous from $\mathscr{S}_2$ into $\mathscr{V}_{k+1}$. Since each g_ϵ in (20) depends continuously of f, and since $p_{k,m}(\partial_j f)\leqslant p_{k-1,m+1}(f)$, so that $\partial_j:\mathscr{V}_{k-1}\to\mathscr{V}_k$ is continuous, we conclude that $f\mapsto T\Diamond f$ is continous from $\mathscr{S}_2$ into $\mathscr{V}_{k+|\gamma|+1}$ and hence from $\mathscr{S}_2$ into $\mathscr{O}_T$.

Now fix $f\in\mathscr{S}_2$ and let T vary in $\mathscr{S}'_2$. Then K is a multiple of C, so if V is a neighborhood of zero in $\mathscr{O}_T$, then $V\cap\mathscr{V}_{k+|\gamma|+1}$ is a zero-neighborhood in $\mathscr{V}_{k+|\gamma|+1}$ and thus contains all $T\Diamond f$ with $T=\hat\partial^\gamma Q$, $|Q(t)|\leqslant C(1+t^2)^k$, for $C\leqslant c_{k\gamma}$ with $c_{k\gamma}$ small enough. Let

$$B := \left\{h\in\mathscr{S}_2:\ \int(1+u^2)^r|\partial^\alpha h(u)|\,du\leqslant\frac{1}{c_{r\alpha}},\ \forall r\in\mathbb{N},\ \forall\alpha\in\mathbb{N}^{2N}\right\}.$$

Then B is bounded in $\mathscr{S}_2$ and its polar B^0 is a neighborhood in $\mathscr{S}'_2$ such that $T\Diamond f\in V$ whenever $T\in B^0$. ■

Remark: The fact that $T\times f\in\mathscr{O}_M$ has been noted in Ref. 7.

A *normal space of distributions*[18] (on $\mathbb{R}^{2N}$) is a locally convex space $\mathscr{R}$, where $\mathscr{D}\subset\mathscr{R}\subset\mathscr{D}'$ with continuous inclusions and $\mathscr{D}$ is dense in $\mathscr{R}$. (Here $\mathscr{D}$ is the space of test functions of compact support on $\mathbb{R}^{2N}$.)

Lemma 1: $\mathscr{V}_k^m$ is a normal space of distributions.

Proof: We adapt the analogous proof of Horváth[18] for $\mathscr{S}^m_{-k}$. Take $g\in\mathscr{D}$ with $g(u)=1$ for $u^2\leqslant 1$ and $0\leqslant g(u)\leqslant 1$, $\forall u\in\mathbb{R}^{2N}$. Set $g_\epsilon(u) := g(\epsilon u)$ for $\epsilon>0$. Then for $f\in\mathscr{V}_k^m$ we have $fg_\epsilon\in\mathscr{D}^m$ (the space of C^m functions of compact support) and from (16) we get

$$p_{k,m}(f-fg_\epsilon)=\sup\left\{(1+u^2)^{-k-|\alpha|/2}\sum_{\beta\le\alpha}\binom{\alpha}{\beta}|\partial^{\alpha-\beta}(1-g(\epsilon u))\partial^\beta f(u)|:\ |\alpha|\le m,\ u\in\mathbb{R}^{2N}\right\}$$
$$\le C\sup\left\{(1+u^2)^{-k-|\alpha|/2}\sum_{\beta\le\alpha}\binom{\alpha}{\beta}|\partial^\beta f(u)|:\ |\alpha|\le m,\ u^2\ge\epsilon^{-2}\right\}$$
$$\le 2^m C\sup\{(1+u^2)^{-k-|\beta|/2}|\partial^\beta f(u)|:|\beta|\le m,\ u^2\ge\epsilon^{-2}\},$$

where we may take

$$C=1+\sup\{|\partial^\gamma g(u)|:|\gamma|\le m,\ u\in\mathbb{R}^{2N}\}.$$

Thus $fg_\epsilon\to f$ in $\mathscr{V}^m_k$ as $\epsilon\to 0$, and so $\mathscr{D}^m$ is dense in $\mathscr{V}^m_k$. Hence $\mathscr{D}$ is dense in $\mathscr{V}^m_k$ since it is dense in $\mathscr{D}^m$. On the other hand, since $q_{k+m,m}(f)\le p_{k,m}(f)\le q_{k,m}(f)$ for $f\in C^m(\mathbb{R}^{2N})$, we get a chain of continuous inclusions:

$$\mathscr{D}\subset\mathscr{S}^m_{-k}\subset\mathscr{V}^m_k\subset\mathscr{S}^m_{-k-m}\subset\mathscr{D}'. \qquad ■$$

Lemma 2: Let $(\mathscr{R}_k)_{k\in\mathbb{N}}$ be a sequence of normal spaces of distributions. Then (1) if $\mathscr{R}_{k+1}\subset\mathscr{R}_k$ with a continuous inclusion for all k, and if $\mathscr{R}:=\bigcap_{k\in\mathbb{N}}\mathscr{R}_k$ with the projective topology, or (2) if $\mathscr{R}_k\subset\mathscr{R}_{k+1}$ with a continuous inclusion for all k, and if $\mathscr{R}:=\bigcup_{k\in\mathbb{N}}\mathscr{R}_k$ with the inductive topology, then $\mathscr{R}$ is a normal space of distributions.

Proof: (1) We have $\mathscr{D}\subset\mathscr{R}\subset\mathscr{R}_k\subset\mathscr{D}'$, $\forall k\in\mathbb{N}$. The first inclusion is continuous since $\mathscr{R}$ has the projective topology and each $\mathscr{D}\subset\mathscr{R}_k$ is continuous; the continuity of the other inclusions is clear. If V is a neighborhood of 0 in $\mathscr{R}$ and if $f\in\mathscr{R}$, then $V=V_k\cap\mathscr{R}$, where V_k is a 0-neighborhood in some $\mathscr{R}_k$; then $f+V_k$ contains some $g\in\mathscr{D}$, and hence $g\in(f+V)$: so $\mathscr{D}$ is dense in $\mathscr{R}$.

(2) we have $\mathscr{D}\subset\mathscr{R}_k\subset\mathscr{R}\subset\mathscr{D}'$, $\forall k\in\mathbb{N}$. The third inclusion is continuous since $\mathscr{R}$ has the inductive topology and each $\mathscr{R}_k\subset\mathscr{D}'$ is continous; the continuity of the other inclusions is clear. If V is a neighborhood of 0 in $\mathscr{R}$ and if $f\in\mathscr{R}$, then $f\in\mathscr{R}_k$ for some k and $V\cap\mathscr{R}_k$ is a 0-neighborhood in some $\mathscr{R}_k$; then $f+(V\cap\mathscr{R}_k)$ contains some $g\in\mathscr{D}$, and hence $g\in(f+V)$: so $\mathscr{D}$ is dense in $\mathscr{R}$. ■

Already in Ref. 18, $\mathscr{O}_C$ has been shown to be a normal space of distributions, where the proof technique is essentially the application of Lemma 2 to the definition of $\mathscr{O}_C$ [(18)]. From (18) we also obtain the normality of $\mathscr{O}_M$. Combining the two lemmas with the definition (17) of $\mathscr{O}_T$, we get normality of $\mathscr{O}_T$. Thus each inclusion in the chain

$$\mathscr{D}\subset\mathscr{O}_C\subset\mathscr{O}_T\subset\mathscr{O}_M\subset\mathscr{D}'$$

is continuous and has dense image (since $\mathscr{D}$ is dense in all these spaces). Thus the transposed maps

$$\mathscr{D}\subset\mathscr{O}'_M\subset\mathscr{O}'_T\subset\mathscr{O}'_C\subset\mathscr{D}'$$

are one-to-one and continuous. We identify each dual space with its image in $\mathscr{D}'$. Since we can interpolate $\mathscr{S}_2$ and $\mathscr{S}'_2$ into both chains, the dual spaces consist of tempered distributions.

The space of distributions of compact support[20] on $\mathbb{R}^{2N}$ is the dual space $\mathscr{E}'_2$ of $\mathscr{E}_2$. We can now show that it is contained in the Moyal algebra.

Theorem 4: The spaces $\mathscr{E}'_2$, $\mathscr{O}'_M$, and $\mathscr{O}'_T$ are contained in $\mathscr{M}$.

Proof: Transposing $\mathscr{O}_T\subset\mathscr{O}_M\subset\mathscr{E}_2$, we get $\mathscr{E}'_2\subset\mathscr{O}'_M\subset\mathscr{O}'_T$, so we need only check that $\mathscr{O}'_T\subset\mathscr{M}$.

If $S\in\mathscr{O}'_T$ and $T\in\mathscr{S}'_2$, we may define $S\times T$, $T\times S$ by transposition:

$$\langle S\times T,h\rangle:=\langle S,T\times h\rangle,\quad \langle T\times S,h\rangle:=\langle S,h\times T\rangle, \tag{21}$$

for $h\in\mathscr{S}_2$; since the right-hand sides are continuous in h, by Theorem 3, $S\times T$ and $T\times S$ are defined in $\mathscr{S}'_2$.

Moreover, for a fixed $h\in\mathscr{S}_2$, the maps $T\mapsto T\times h$, $T\mapsto h\times T$ are continuous: $\mathscr{S}'_2\to\mathscr{O}_T$ by Theorem 3, so they transpose to continuous maps $S\mapsto h\times S$, $S\mapsto S\times h$ from $\mathscr{O}'_T$ into $(\mathscr{S}'_2)'=\mathscr{S}_2$, via

$$\langle h\times S,T\rangle:=\langle S,T\times h\rangle,\quad \langle S\times h,T\rangle:=\langle S,h\times T\rangle. \tag{22}$$

Combining (21) and (22), we get

$$\langle T\times S,h\rangle:=\langle T,S\times h\rangle,\quad \langle S\times T,h\rangle:=\langle T,h\times S\rangle, \tag{23}$$

for $T\in\mathscr{S}'_2$, $S\in\mathscr{O}'_T$, $h\in S_2$. We have shown that $S\times h\in\mathscr{S}_2$, $h\times S\in\mathscr{S}_2$, whenever $h\in\mathscr{S}_2$, $S\in\mathscr{O}'_T$; these are resultants of separately continuous extensions to $\mathscr{O}'_T\times\mathscr{S}_2$ of the twisted product on $\mathscr{S}_2$, and since $\mathscr{S}_2$ is dense in $\mathscr{O}'_T$ these extensions are unique. Since (23) is formally identical with (15), the twisted products (21), (22) are consistent with previous ones, and we conclude that $\mathscr{O}'_T\subset\mathscr{M}$. Since $\mathbf{1}\notin\mathscr{O}'_T$, we have $\mathscr{O}'_T\ne\mathscr{M}$. ■

Corollary: The space $\mathscr{O}_C$ is contained in $\mathscr{M}$. In particular, the (ordinary) convolution of any function in $\mathscr{S}_2$ with any tempered distribution belongs to $\mathscr{M}$.

Proof: Since $\mathscr{O}_C$ is reflexive,[22] it is enough to note that $\mathscr{F}(\mathscr{O}'_M)=\mathscr{O}_C$. Then use Theorem 4. Also,[18] $T*f$ belongs to $\mathscr{O}_C$ if $f\in\mathscr{S}_2$, $T\in\mathscr{S}'_2$. ■

If $R,S\in\mathscr{S}'_2$, their tensor product $R\otimes S\in\mathscr{S}'(\mathbb{R}^{4N})$ is given by

$$\langle R\otimes S,f\otimes g\rangle:=\langle R,f\rangle\langle S,g\rangle.$$

If $R,S\in\mathscr{M}$, we have

$$\langle R\lozenge S,h\rangle=\langle R,\check{S}\lozenge h\rangle=\langle R,u\mapsto\langle S,\epsilon_{-u}\tau_{-u}h\rangle\rangle$$
$$=\langle R\otimes S,(u,v)\mapsto e^{iu'Jv}h(u+v)\rangle.$$

Writing $h_2(u,v):=e^{-iu'Jv}h(u+v)$, we find that

$$\partial^\alpha_u\partial^\gamma_v h_2(u,v)$$
$$=\sum_{\beta\le\alpha}\sum_{\epsilon\le\gamma}i^{|\epsilon|-|\beta|}\binom{\alpha}{\beta}\binom{\gamma}{\epsilon}(Jv)^\beta(Ju)^\epsilon e^{-iu'Jv}\times(\partial^{\alpha-\beta}_u\partial^{\gamma-\epsilon}_v h)(u+v). \tag{24}$$

Thus $h\mapsto h_2$ is continuous: $\mathscr{E}_2\to\mathscr{E}(\mathbb{R}^{4N})$. Since $(R,S)\mapsto R\otimes S$ is jointly continuous: $\mathscr{E}'_2\times\mathscr{E}'_2\to\mathscr{E}'(\mathbb{R}^{4N})$, and $\langle R\lozenge S,h\rangle=\langle R\otimes S,h_2\rangle$, we find that $(R,S)\mapsto R\lozenge S$ is jointly continuous on $\mathscr{E}'_2$.

By the Paley–Wiener theorem, $F(\mathscr{E}'_2) = \mathscr{F}(\mathscr{E}'_2) =: \mathscr{O}_{\text{exp}}$ is the space of functions in $\mathscr{O}_M$ that extend to analytic functions of exponential type, and by Theorem 4 and the Fourier invariance of $\mathscr{M}$, $\mathscr{O}_{\text{exp}}$ is contained in $\mathscr{M}$. ($\mathscr{O}_{\text{exp}}$ carries the topology induced by $\mathscr{F}$ from $\mathscr{E}'_2$.)

For convenience, we write $\hat{\mu}_j f := \mu_{j+N} f$ if $j \leq N$, $\hat{\mu}_j f := -\mu_{j-N} f$ if $j > N$. If $h \in \mathscr{E}_2$, the expansion

$$e^{-iu'Jv} h(u+v) = \sum_{k=0}^{\infty} \frac{(-i)^k}{k!} (u'Jv)^k h(u+v)$$

converges uniformly on compact subsets of $\mathbf{R}^{2N}$, together with all derivatives on account of (24), and this convergence is uniform on bounded subsets of $\mathscr{E}_2$. Thus

$$\langle S \Diamond T, h \rangle = \sum_{k=0}^{\infty} \frac{(-i)^k}{k!} \langle S \otimes T, (u,v) \mapsto (u'Jv)^k h(u+v) \rangle.$$

Since

$$(u'Jv)^k = \sum_{|\alpha| = k} \frac{k!}{\alpha!} u^{\alpha'} (Jv)^{\alpha},$$

and $(\hat{\mu}_j f)(v) = (Jv)_j f(v)$ for all $f \in \mathscr{S}_2$ and each j, we derive

$$\begin{aligned}
&\langle S \Diamond T, h \rangle \\
&= \sum_{k=0}^{\infty} \sum_{|\alpha| = k} \frac{(-i)^k}{\alpha!} \langle S \otimes T, (u,v) \mapsto u^{\alpha'} (Jv)^{\alpha} h(u+v) \rangle \\
&= \sum_{\alpha \in \mathbf{N}^{2N}} \frac{(-i)^{|\alpha|}}{\alpha!} \langle \mu^{\alpha} S \otimes \hat{\mu}^{\alpha} T, (u,v) \mapsto h(u+v) \rangle \\
&= \sum_{\alpha \in \mathbf{N}^{2N}} \frac{(-i)^{|\alpha|}}{\alpha!} \langle \mu^{\alpha} S * \hat{\mu}^{\alpha} T, h \rangle, \qquad (25)
\end{aligned}$$

where the series converges uniformly for h in bounded subsets of $\mathscr{E}_2$. We are now able to expand the twisted product as a series of products of derivatives.

Theorem 5: If $S, T \in \mathscr{O}_{\text{exp}}$, then

$$S \times T = \sum_{\alpha \in \mathbf{N}^{2N}} \frac{i^{|\alpha|}}{\alpha!} (\partial^{\alpha} S)(\hat{\partial}^{\alpha} T)$$

with convergence in the topology of $\mathscr{O}_{\text{exp}}$.

Proof: We apply the Fourier transform $\mathscr{F}$ to (25), replacing S, T by $\mathscr{F}^{-1} S$, $\mathscr{F}^{-1} T$. By the continuity of $\mathscr{F}$ we get

$$\begin{aligned}
S \times T &= \sum_{\alpha} \frac{(-i)^{|\alpha|}}{\alpha!} \mathscr{F}(\mu^{\alpha} \mathscr{F}^{-1} S * \hat{\mu}_{\alpha} \mathscr{F}^{-1} T) \\
&= \sum_{\alpha} \frac{(-i)^{|\alpha|}}{\alpha!} i^{2|\alpha|} (\partial^{\alpha} S)(\hat{\partial}^{\alpha} T) \\
&= \sum_{\alpha} \frac{i^{|\alpha|}}{\alpha!} (\partial^{\alpha} S)(\hat{\partial}^{\alpha} T). \qquad \blacksquare
\end{aligned}$$

Corollary: If $S, T \in \mathscr{O}_{\text{exp}}$, then

$$\begin{aligned}
&S \times T - T \times S \\
&= 2i \sum_{r=0}^{\infty} \sum_{|\alpha| = 2r+1} \frac{(-1)^r}{\alpha!} (\partial^{\alpha} S)(\hat{\partial}^{\alpha} T). \qquad (26)
\end{aligned}$$

Remarks: (1) The restriction to $\mathscr{E}'_2$ or $\mathscr{O}_{\text{exp}}$ is only needed to guarantee convergence of the series indexed by α. If either S or T is a polynomial, these series are finite sums, and the expansions are valid.

(2) For $N = 1$, the leading term in (26) is the ordinary Poisson bracket $\partial_q S \, \partial_p T - \partial_p S \, \partial_q T$. As a differential operator on $S \otimes T$, we may formally write

$$\begin{aligned}
&S \times T - T \times S \\
&= 2i \left(\sum_{r=0} \frac{(-1)^r}{(2r+1)!} (\partial_q \otimes \partial_p - \partial_p \otimes \partial_q)^{2r+1} \right) (S \otimes T) \\
&= 2i (\sin(\partial_q \otimes \partial_p - \partial_p \otimes \partial_q))(S \otimes T),
\end{aligned}$$

an expression first derived by Moyal,[11] and called the "Moyal bracket."

V. THE MATRICIAL FORM OF THE TWISTED PRODUCT

By working in the Schwartz space $\mathscr{S}_2$, we avoid the usual continuity problems for the creation and annihilation operators for the harmonic oscillator, as has been observed before.[23] In the present context, these operators are represented by first-degree polynomials in $\mathscr{M}$. To avoid notational clutter, we will take $N = 1$ in this section, but the results go through in the general case with the systematic use of multi-indices. We write $u = (q,p)$ and use $q, p, \partial_q, \partial_p$ in place of $\mu_1, \mu_2, \partial_1, \partial_2$, respectively. We introduce the notation

$$\begin{aligned}
&a := \frac{q + ip}{\sqrt{2}}, \quad \bar{a} := \frac{q - ip}{\sqrt{2}}, \\
&\frac{\partial}{\partial a} := \frac{\partial_q - i\partial_p}{\sqrt{2}}, \quad \frac{\partial}{\partial \bar{a}} := \frac{\partial_q + i\partial_p}{\sqrt{2}}, \\
&H := a\bar{a} = \tfrac{1}{2}(q^2 + p^2) = \tfrac{1}{2}u^2, \quad f_0 = 2e^{-\bar{a}a} = 2e^{-H}.
\end{aligned}$$

From (12) we obtain the formulas

$$\begin{aligned}
&a \times f = af + \frac{\partial f}{\partial \bar{a}}, \quad f \times a = af - \frac{\partial f}{\partial \bar{a}}, \\
&\bar{a} \times f = \bar{a} f - \frac{\partial f}{\partial a}, \quad f \times \bar{a} = \bar{a} f + \frac{\partial f}{\partial a}.
\end{aligned} \qquad (27)$$

We get at once the following equalities in $\mathscr{M}$:

$$\bar{a} \times a = H - 1, \quad a \times \bar{a} = H + 1, \quad a \times \bar{a} - \bar{a} \times a = 2. \qquad (28)$$

The third equality is, of course, the canonical commutation relation for a and $\bar{a}$: recall that we have taken units in which $\hbar = 2$. We note also that $a \times a \times \cdots \times a$ (n times) $= a^n$.

The Gaussian function f_0 has several nice properties: it is (pointwise) positive, it is a fixed point for the various Fourier transforms, and it is an idempotent in $\mathscr{S}_2$ for both the twisted product and the twisted convolution. Moreover, it is a unit vector in $L^2(\mathbf{R}^2)$, because of our choices of normalization.

Notice that if $g \in \mathscr{O}_M$, (27) implies that

$$a \times (g f_0) = \left(\frac{\partial g}{\partial \bar{a}} \right) f_0, \quad \bar{a} \times (g f_0) = \left(2\bar{a} g - \frac{\partial g}{\partial a} \right) f_0. \qquad (29)$$

Taking $g = 2^m \bar{a}^m$, we find that

$$\begin{aligned}
\bar{a} \times (2^m \bar{a}^m f_0) &= \left(2^{m+1} \bar{a}^{m+1} - 2^m \frac{\partial \bar{a}^m}{\partial a} f_0 \right) \\
&= 2^{m+1} \bar{a}^{m+1} f_0,
\end{aligned}$$

so we get by induction that $\bar{a}^m \times f_0 = 2^m \bar{a}^m f_0$ if $m \in \mathbb{N}$. If $n > m$, this gives

$$a^n \times \bar{a}^m \times f_0 = a^n \times (2^m \bar{a}^m f_0) = 2^m \left(\frac{\partial^n}{\partial \bar{a}^n}\right)(\bar{a}^m) f_0 = 0,$$

and if $n < m$, then

$$f_0 \times a^n \times \bar{a}^m = (a^m \times \bar{a}^n \times f_0)^* = 0.$$

Also,

$$\begin{aligned} f_0 \times a^n \times \bar{a}^n \times f_0 &= f_0 \times a^n \times (2^n \bar{a}^n f_0) \\ &= f_0 \times \left(2^n \left(\frac{\partial^n}{\partial \bar{a}^n}\right)(\bar{a}^n) f_0\right) \\ &= 2^n n! f_0 \times f_0 = 2^n n! f_0. \end{aligned}$$

To summarize,

$$f_0 \times a^n \times \bar{a}^m \times f_0 = \delta_{mn} 2^n n! f_0, \quad \text{for } m,n \in \mathbb{N}. \tag{30}$$

We now introduce an orthonormal basis for $L^2(\mathbb{R}^2)$, which we declare as a doubly indexed family of functions in $\mathcal{S}_2$, since this family forms a system of matrix units with respect to the twisted product.

Definition 5: For $m,n \in \mathbb{N}$, we define $f_{mn} \in \mathcal{S}_2$ by

$$f_{mn} := (1/\sqrt{2^{m+n} m! n!}) \bar{a}^m \times f_0 \times a^n. \tag{31}$$

From (30) we get directly

$$\begin{aligned} &f_{mn} \times f_{kl} \\ &\quad = (2^{m+n+k+l} m! n! k! l!)^{-1/2} \bar{a}^m \times f_0 \times a^n \times \bar{a}^k \times f_0 \times a^l \\ &\quad = (\delta_{nk}/\sqrt{2^{m+l} m! l!}) \bar{a}^m \times f_0 \times a^l = \delta_{nk} f_{ml}. \end{aligned} \tag{32}$$

This implies that

$$\begin{aligned} 2(f_{mn} | f_{kl}) &= \langle f_{nm}, f_{kl} \rangle \\ &= \int (f_{nm} \times f_{kl})(u) du = \delta_{mk} \int f_{nl}(u) du \\ &= \frac{\delta_{mk}}{\sqrt{2^{n+l} n! l!}} \int (\bar{a}^n \times f_0 \times f_0 \times a^l)(u) du \\ &= \frac{\delta_{mk}}{\sqrt{2^{n+l} n! l!}} \int (f_0 \times a^l \times \bar{a}^n \times f_0)(u) du \\ &= \delta_{mk} \delta_{nl} \int f_0(u) du = 2 \delta_{mk} \delta_{nl}, \end{aligned}$$

so that $\{f_{mn}: m,n \in \mathbb{N}\}$ is orthonormal in $L^2(\mathbb{R}^2)$. This family is complete since [see (36)] all the Hermite functions on $\mathbb{R}^2$ are linear combinations of the f_{mn}.

Remark: The basis (f_{mn}) lies in $\mathcal{S}_2$, as is clear from (31) since $a, \bar{a} \in \mathcal{M}$. It has also the important property of diagonalizing the Fourier transforms; one readily checks that

$$\begin{aligned} &\mathcal{F}(f_{mn}) = (-i)^{m+n} f_{mn}, \\ &F(f_{mn}) = (-i)^n f_{mn}, \quad \breve{F}(f_{mn}) = (-i)^m f_{mn}, \end{aligned} \tag{33}$$

and hence also $\breve{f}_{mn} = (-1)^{m+n} f_{mn}$.

To present f_{mn} explicitly, we use polar coordinates: $q + ip =: \rho e^{i\alpha}$, noting that $\rho^2 = q^2 + p^2 = u^2$.

By induction, applying (29) to (31), we derive

$$\begin{aligned} f_{mn} = &\frac{1}{\sqrt{2^{m+n} m! n!}} \sum_{k=0}^{n} (-1)^k \binom{m}{k} \binom{n}{k} \\ &\times k! 2^{m+n-k} \bar{a}^{m-k} a^{n-k} f_0 \end{aligned}$$

and, since $a = (1/\sqrt{2}) \rho e^{i\alpha}$, $\bar{a} = (1/\sqrt{2}) \rho e^{-i\alpha}$, we get, after some rearrangement,

$$\begin{aligned} &f_{mn}(\rho, \alpha) \\ &\quad = 2(-1)^n \sqrt{(n!/m!)} e^{-i\alpha(m-n)} \rho^{m-n} L_n^{m-n}(\rho^2) e^{-\rho^2/2} \end{aligned} \tag{34}$$

and in particular

$$f_{nn}(\rho, \alpha) = 2(-1)^n L_n(\rho^2) e^{-\rho^2/2}, \tag{35}$$

where L_n and L_n^{m-n} are the usual Laguerre polynomials of order n.

Remark: Equation (35) agrees with the Wigner function of the nth energy level of the harmonic oscillator, obtained in Ref. 10; and (34) agrees with the "transition" between levels of the harmonic oscillator, first derived in Ref. 24; see also Refs. 3 and 25.

Now we represent $\mathcal{S}_2$, $\mathcal{S}'_2$, and $L^2(\mathbb{R}^2)$ as sequence spaces of coefficients after expansion in the twisted Hermite basis (f_{mn}). Our treatment is in the spirit of Simon's work[26] with the ordinary Hermite basis. Since here the coefficients form a doubly indexed family, we may consider their matrix product, which turns out to correspond to the twisted product of the associated functions or distributions.

The fundamental fact that underlies the sequence constructions is that the twisted Hermite basis "diagonalizes" the oscillator Hamiltonian H (its eigenvalues are odd integers rather than half-integers due to our convention that $\hbar = 2$).

Proposition 5: If $m,n \in \mathbb{N}$, then

$$H \times f_{mn} = (2m+1) f_{mn}, \quad f_{mn} \times H = (2n+1) f_{mn}.$$

Proof: From (31) we get at once

$$\begin{aligned} a \times f_{mn} &= \sqrt{2m} f_{m-1,n}, \quad f_{mn} \times a = \sqrt{2n+2} f_{m,n+1}, \\ \bar{a} \times f_{mn} &= \sqrt{2m+2} f_{m+1,n}, \quad f_{mn} \times \bar{a} = \sqrt{2n} f_{m,n-1} \end{aligned}$$

(with $f_{mn} = 0$ if m or n is -1). The result follows from (28). ■

Let us write $A(f) := H \times f \times H$. We could consider A as an operator on $L^2(\mathbb{R}^2)$ with domain $\mathcal{S}_2$; as such, A is symmetric and closable, and clearly unbounded. Indeed,

$$(A \pm iI) f_{mn} = ((2m+1)(2n+1) \pm i) f_{mn}$$

and hence $A \pm iI$ has dense range: thus A is essentially self-adjoint. Moreover, A^{-1} has finite-dimensional eigenspaces and

$$\sum_{m,n=0}^{\infty} (2m+1)^{-2} (2n+1)^{-2} = \left(\frac{\pi^2}{8}\right)^2$$

is finite, so A^{-1} is a Hilbert–Schmidt operator. It is not hard to check that the seminorms $f \mapsto \|A^k f\|$ $(k \in \mathbb{N})$ generate the topology of $\mathcal{S}_2$.

Remark: Let $B = u^2 - \Delta$ be the usual Hermite operator on $L^2(\mathbb{R}^2)$. We have $Bf = H \times f + f \times H$, so $Bf_{mn} = 2(m+n+1) f_{mn}$. If

$$h_k(x) := (2^{k-1} k!)^{-1/2} H_k(x) e^{-x^2/2}$$

is the usual Hermite function of degree k, we conclude that

$$f_{mn} = \sum_{k+l=m+n} c^{kl}_{mn} h_k \otimes h_l,$$
$$h_k \otimes h_l = \sum_{m+n=k+l} b^{mn}_{kl} f_{mn}, \tag{36}$$

for some constants c^{kl}_{mn}, b^{mn}_{kl}. In fact, we may compute that

$$c^{kl}_{mn} = 2^{(m-n)/2} i^{2m+l} \binom{m+n}{l}^{1/2} \binom{m+n}{m}^{-1/2} \times P^{l-m,k-m}_m(0),$$

where $P^{l-m,k-m}_m$ is the usual Jacobi polynomial. (This takes care of the completeness argument for the f_{mn}.)

We can now characterize $\mathscr{S}_2$ and $\mathscr{S}'_2$ as sequence spaces.

Theorem 6: Let **s** be the Fréchet space of rapidly decreasing double sequences $c = (c_{mn})$ such that

$$r_k(c) := \left[\sum_{m,n=0}^{\infty} (2m+1)^{2k}(2n+1)^{2k} |c_{mn}|^2 \right]^{1/2}$$

is finite for all $k \in \mathbb{N}$, topologized by the seminorms $(r_k)_{k\in\mathbb{N}}$. For $f \in \mathscr{S}_2$, let c be the sequence of coefficients in the expansion

$$f = \sum_{m,n=0}^{\infty} c_{mn} f_{mn}.$$

Then $f \mapsto c$ is an isomorphism of Fréchet spaces from $\mathscr{S}_2$ onto **s**.

Proof: If $f \in \mathscr{S}_2$, then $\|A^k f\| < \infty$, for all $k \in \mathbb{N}$, so that $r_k(c) = \|A^k f\|$ is finite for all k. It follows that $f \mapsto c$ is a one-to-one topological isomorphism of $\mathscr{S}_2$ into **s**.

Given any $c \in$ **s**, for $M,N \in \mathbb{N}$ let c^{MN} be the double sequence defined by $c^{MN}_{mn} := c_{mn}$, if $m \leqslant M$, $n \leqslant N$; $c^{MN}_{mn} := 0$, otherwise. Then $r_k(c^{MN} - c) \to 0$ as $M,N \to \infty$, for each k, so that the functions $\Sigma^M_{m=0} \Sigma^N_{n=0} c_{mn} f_{mn}$ form a Cauchy sequence in $\mathscr{S}_2$ and hence converge to a function f which maps onto c. ∎

For $T \in \mathscr{S}'_2$, $m,n \in \mathbb{N}$, define $b_{mn} := \langle T, f_{nm} \rangle$. Then

$$\langle T, f \rangle = \sum_{m,n=0}^{\infty} c_{nm} \langle T, f_{nm} \rangle = \sum_{m,n=0}^{\infty} c_{nm} b_{mn}, \tag{37}$$

where the series converges absolutely, for each

$$f = \sum_{m,n=0}^{\infty} c_{mn} f_{mn} \in \mathscr{S}_2.$$

Since $T \in \mathscr{S}'_2$, there exist $k \in \mathbb{N}$, $K > 0$ such that

$$|\langle T, f \rangle| \leqslant K \|A^k f\| = K r_k(c) \tag{38}$$

for all $f \in \mathscr{S}_2$, and since

$$\langle T, f \rangle = \sum_{m,n=0}^{\infty} (2m+1)^{-k}(2n+1)^{-k} \times b_{mn} (2n+1)^k (2m+1)^k c_{nm},$$

the Schwarz inequality gives $r_{-k}(b) \leqslant K$. Thus, whenever b_{mn} is a double sequence with $r_{-k}(b)$ finite for some k, the series $\Sigma^\infty_{m,n=0} b_{mn} f_{mn}$ converges weakly to T in $\mathscr{S}'_2$, and (38) shows that the convergence is uniform on bounded subsets of $\mathscr{S}_2$, so the series converges to T in the strong dual topology of $\mathscr{S}'_2$.

The main result of this section is now easy.

Theorem 7: If $a,b \in$ **s** correspond respectively to $f,g \in \mathscr{S}_2$ as coefficient sequences in the twisted Hermite basis, then the sequence corresponding to the twisted product $f \times g$ is the matrix product ab, where

$$(ab)_{mn} := \sum_{k=0}^{\infty} a_{mk} b_{kn}. \tag{39}$$

Proof: From (32) and the continuity of $\times$ in $\mathscr{S}_2$, we get

$$f \times g = \left(\sum_{m,k} a_{mk} f_{mk} \right) \times \left(\sum_{r,n} b_{rn} f_{rn} \right)$$
$$= \sum_{m,k,r,n} a_{mk} b_{rn} f_{mk} \times f_{rn} = \sum_{m,k,n} a_{mk} b_{kn} f_{mn}. \quad ∎$$

Corollary:

$$\{ f \times g : \ f,g \in \mathscr{S}_2 \} = \mathscr{S}_2.$$

Proof: It suffices to show that any $c \in$ **s** can be written as the matrix product of two sequences in **s**. We use Howe's argument[25] to show this.

Set $d_m := (\sup\{ |c_{jr}| : j \in \mathbb{N},\ r \geqslant m \})^{1/2}$ for $m \in \mathbb{N}$, and let d be the "diagonal" sequence with entries $d_m \delta_{mn}$. Then one verifies that $r_k(d)^2 \leqslant C_k r_{2k+2}(c)$ for some constants C_k, so that $d \in$ **s**. Now if we set $b_{mn} := c_{mn}/d_n$, we get

$$|b_{mn}| = |c_{mn}|/d_n \leqslant d^2_n / d_n = d_n$$

and thus $b \in$ **s** also. Clearly $bd = c$. ∎

Remark: The sequence $a \Diamond b$ corresponding to $f \Diamond g$ is

$$(a \Diamond b)_{mn} := \sum_{k=0}^{\infty} (-1)^k a_{mk} b_{kn}.$$

Since, by (4) and (11), $f \Diamond g = f \times \delta \times g$, it suffices to show that δ is represented by the diagonal matrix with entries $(-1)^m \delta_{mn}$; this follows from (33), since $\mathscr{F} 1 = \delta$. Thus the entire theory of the twisted product and convolution could be developed, at least formally, in the matrix language and without mention of the symplectic Fourier transforms; the basic transformation formulas [see (4)] are

$$R \Diamond S = R \times \delta \times S, \quad R \times S = R \Diamond 1 \Diamond S.$$

We now show that (39) gives a second way of defining the twisted product for many pairs of distributions, which lie in spaces of Sobolev type.

Definition 6: For $s,t \in \mathbb{R}$, we denote by $\mathscr{G}_{s,t}$ the Hilbert space obtained by completing $\mathscr{S}_2$ with respect to the norm

$$\|f\|_{s,t} := \left(\sum_{m,n=0}^{\infty} (2m+1)^s (2n+1)^t |c_{mn}|^2 \right)^{1/2}. \tag{40}$$

Observe that $\mathscr{G}_{0,0} = L^2(\mathbb{R}^2)$ with $\|f\|_{0,0} = \|f\|$. An orthonormal basis for $\mathscr{G}_{s,t}$ is given by the functions $(2m+1)^{-s/2}(2n+1)^{-t/2} f_{mn}$, and thus

$$f = \sum_{m,n=0}^{\infty} c_{mn} f_{mn}$$

with convergence in the (s,t) norm, for all $f \in \mathscr{G}_{s,t}$. Note that $\mathscr{S}_2 = \bigcap_{s,t\in\mathbb{R}} \mathscr{G}_{s,t}$ topologically. Since $\mathscr{S}_2 \subset \mathscr{G}_{s,t}$ is a continuous inclusion with dense image, $\mathcal{G}_{s,t}$ is a normal space of tempered distributions, and the transpose of $\mathcal{S}_2 \subset \mathcal{G}_{s,t}$ is the inclusion $\mathcal{G}_{-t,-s} \subset \mathcal{S}_2'$. Also, from (38), $\mathcal{S}_2' = \bigcup_{s,t\in\mathbb{R}} \mathcal{G}_{s,t}$ (topologically). Furthermore, $\mathcal{G}_{s,t} \subset \mathcal{G}_{q,r}$ with a continuous inclusion iff $s \geqslant q$ and $t \geqslant r$. Note also that $f^* \in \mathcal{G}_{t,s}$ whenever $f \in \mathcal{G}_{s,t}$.

If

$$g = \sum_{m,n=0}^{\infty} b_{mn} f_{mn} \in \mathcal{G}_{q,r},$$

we define formally

$$f \times g := \sum_{m,n=0}^{\infty} \left(\sum_{k=0}^{\infty} c_{mk} b_{kn} \right) f_{mn}. \tag{41}$$

Theorem 8: (1) The series (41) converges in $\mathcal{G}_{s,r}$ if $t + q \geqslant 0$, and in that case $\|f \times g\|_{s,r} \leqslant \|f\|_{s,t} \|g\|_{q,r}$.

(2) $\mathcal{G}_{s,t}$ is a Banach algebra under the twisted product (41) whenever $s + t \geqslant 0$; for $s \geqslant 0$, $\mathcal{G}_{s,s}$ is a Banach *-algebra.

(3) The Fourier transforms $F, \bar{F}, \mathcal{F}$ are unitary isometries of each $\mathcal{G}_{s,t}$ onto itself.

(4) The twisted product (41) is consistent with previous definitions.

Proof: (1) From the Schwarz inequality we get

$$\begin{aligned}
\|f \times g\|_{s,r}^2 &\leqslant \sum_{m,n=0}^{\infty} (2m+1)^s \left(\sum_{k=0}^{\infty} |c_{mk} b_{kn}| \right)^2 (2n+1)^r \\
&= \sum_{m,n=0}^{\infty} (2m+1)^s \left(\sum_{k=0}^{\infty} |c_{mk}| (2k+1)^{t/2} (2k+1)^{-t/2} |b_{kn}| \right)^2 (2n+1)^r \\
&\leqslant \sum_{m,k=0}^{\infty} (2m+1)^s |c_{mk}|^2 (2k+1)^t \sum_{l,n=0}^{\infty} (2l+1)^{-t} |b_{ln}|^2 (2n+1)^r = \|f\|_{s,t}^2 \|g\|_{-t,r}^2 \leqslant \|f\|_{s,t}^2 \|g\|_{q,r}^2,
\end{aligned}$$

whenever $q \geqslant -t$, which yields convergence of (41) in this case.

(2) It follows by taking $q = s$, $r = t$.

(3) The Fourier invariance of $\mathcal{G}_{s,t}$ is evident from (40) and (33).

(4) If

$$T = \sum_{m,n=0}^{\infty} d_{mn} f_{mn} \in \mathcal{S}_2',$$

then

$$\begin{aligned}
\langle T, f \times g \rangle &= \sum_{m,n=0}^{\infty} d_{mn} \left(\sum_{k=0}^{\infty} c_{mk} b_{kn} \right) \\
&= \sum_{k,n=0}^{\infty} \left(\sum_{m=0}^{\infty} d_{nm} c_{mk} \right) b_{kn} = \langle T \times f, g \rangle,
\end{aligned} \tag{42}$$

where the convergence of the double sums is absolute by (37) and that of the simple sums is also absolute, in the second case because $T \times f$ lies in some $\mathcal{G}_{s,t}$ and g lies in $\mathcal{S}_2 \subset \mathcal{G}_{-t,-s}$. Thus we may interchange the summations, obtaining

$$T \times f = \sum_{k,n=0}^{\infty} \left(\sum_{m=0}^{\infty} d_{nm} c_{mk} \right) f_{nk}.$$

If $f \in \mathcal{M}_L$, $g \in \mathcal{S}_2$, then by Definition 4,

$$\langle T \times f, g \rangle := \langle T, f \times g \rangle = \sum_{m,n=0}^{\infty} d_{nm} \left(\sum_{k=0}^{\infty} c_{mk} b_{kn} \right);$$

the double sum converges absolutely since $f \times g \in \mathcal{S}_2$, so we may interchange the order of summation to recover (42). ∎

Remarks: (1) We see that if $f, g \in L^2(\mathbb{R}^2)$, then $f \times g \in L^2(\mathbb{R}^2)$ and $\|f \times g\| \leqslant \|f\| \, \|g\|$. Moreover $f \times g$ lies in $C_0(\mathbb{R}^2)$: the continuity follows by adapting the analogous argument for (ordinary) convolution.

(2) Notice from (40) that $A(\mathcal{G}_{s,t}) = \mathcal{G}_{s-2,t-2}$, where we write $A(T) := H \times T \times H$ for any $T \in \mathcal{S}_2'$, which makes sense since $H \in \mathcal{M}$. Clearly $A(\mathcal{M}) \subset \mathcal{M}$, so that if $\mathcal{M}$ were to contain any $\mathcal{G}_{s,t}$, it would contain them all. However, $\mathcal{M} \neq \mathcal{S}_2'$, and thus $\mathcal{M}$ contains no $\mathcal{G}_{s,t}$; in particular, $L^2(\mathbb{R}^{2N}) \not\subset \mathcal{M}$. Hence, $\mathcal{M}$ really provides a *different* extension of $\times$ on $\mathcal{S}_2$ from the Banach algebras $\mathcal{G}_{s,t}$ $(s + t \geqslant 0)$.

VI. CONCLUSION AND OUTLOOK

We have been led to define the twisted product of a pair of distributions and to introduce the Moyal *-algebra $\mathcal{M}$ under the twisted product. A rigorous formulation of the phase-space approach to quantum theory, in the arena given by $\mathcal{M}$, should address the following problems: (a) find an equivalent of the spectral theorem; (b) solve the dynamical equations using such a spectral theorem; and (c) describe the algebraic structure of the state space corresponding to $\mathcal{M}$.

We plan to take up these problems. Our first task is to give $\mathcal{M}$ an appropriate topology: this we do in the following paper.[27]

ACKNOWLEDGMENTS

We are grateful for helpful correspondence from John Horváth and Peter Wagner. We would like to thank Professor Abdus Salam and the International Centre for Theoretical Physics, for their hospitality during a stay in which this work was completed.

We gratefully acknowledge support from the Vicerrectoría de Investigación of the Universidad de Costa Rica.

[1]A. Grossmann, G. Loupias, and E. M. Stein, Ann. Inst. Fourier (Grenoble) **18**, 343 (1968).

[2]F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, Ann. Phys. (NY) **111**, 61, 111 (1978).

[3]J. P. Amiet and P. Huguenin, *Mécaniques classique et quantique dans l'espace de phase* (Université de Neuchâtel, Neuchâtel, 1981).

[4]M. Hillery, R. F. O'Connell, M. D. Scully, and E. P. Wigner, Phys. Rep. **106**, 121 (1984).

[5]J. M. Gracia-Bondía, Phys. Rev. A **30**, 691 (1984).

[6]J. C. T. Pool, J. Math. Phys. **7**, 66 (1966).

[7]R. F. V. Anderson, J. Funct. Anal. **9**, 423 (1972).

[8]R. Cressman, J. Funct. Anal. **22**, 405 (1976).

[9]L. Hörmander, Commun. Pure Appl. Math. **32**, 359 (1979).
[10]H. J. Groenewold, Physica **12**, 405 (1946).
[11]J. E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99 (1949).
[12]W. Rudin, *Functional Analysis* (McGraw-Hill, New York, 1973).
[13]S. Miracle-Solé, thesis, Universidad de Barcelona, 1966.
[14]J. von Neumann, Math. Ann. **104**, 570 (1931).
[15]D. Kastler, Commun. Math. Phys. **1**, 175 (1964).
[16]G. Loupias and S. Miracle-Solé, Commun. Math. Phys. **2**, 31 (1966); Ann. Inst. H. Poincaré **6**, 39 (1967).
[17]J. J. Sławianowski, in *The Uncertainty Principle and Foundations of Quantum Mechanics* (Wiley, New York, 1977), p. 147.
[18]J. Horváth, *Topological Vector Spaces and Distributions I* (Addison–Wesley, Reading, MA, 1966).
[19]H. H. Schaefer, *Toplogical Vector Spaces* (Macmillan, New York, 1966).
[20]L. Schwartz, *Theorie des Distributions* (Hermann, Paris, 1966).
[21]M. A. Antonets, Lett. Math. Phys. **2**, 241 (1978); Teor. Mat. Fiz. **38**, 331 (1979).
[22]A. Grothendieck, *Produits tensoriels topologiques et espaces nucléaires* (Am. Math. Soc., Providence, RI, 1955).
[23]P. Kristensen, L. Mejlbo, and E. T. Poulsen, Commun. Math. Phys. **1**, 175 (1965).
[24]M. S. Bartlett and J. E. Moyal, Proc. Cambridge Philos. Soc. **45**, 545 (1949).
[25]R. Howe, J. Funct. Anal. **38**, 188 (1980).
[26]B. Simon, J. Math. Phys. **12**, 140 (1971).
[27]J. C. Várilly and J. M. Gracia-Bondía, J. Math. Phys. **29**, 880 (1988).

Algebras of distributions suitable for phase-space quantum mechanics. II. Topologies on the Moyal algebra

Joseph C. Várilly and José M. Gracia-Bondía
Escuela de Matemática, Universidad de Costa Rica, San José, Costa Rica

(Received 14 July 1986; accepted for publication 11 November 1987)

The topology of the Moyal *-algebra may be defined in three ways: the algebra may be regarded as an operator algebra over the space of smooth declining functions either on the configuration space or on the phase space itself; or one may construct the *-algebra via a filtration of Hilbert spaces (or other Banach spaces) of distributions. The equivalence of the three topologies thereby obtained is proved. As a consequence, by filtrating the space of tempered distributions by Banach subspaces, new sufficient conditions are given for a phase-space function to correspond to a trace-class operator via the Weyl correspondence rule.

I. INTRODUCTION

In the previous article (Ref. 1, hereinafter referred to simply as I), we laid the foundations of a promising mathematical mold for the phase-space formulation of quantum mechanics. In this paper we obtain some less straightforward results, in keeping with the preliminary stage of the program outlined in I.

A wealth of information may be gained by characterizing the Moyal *-algebra $\mathcal{M}$ as a suitably defined limit of a family of Banach spaces, which form a filtration of the space of tempered distributions on phase space. In fact, we introduce several variants of this filtration, depending on whether to use Hilbert algebras or some other kind of Banach algebras.

The definition of an adequate topology on $\mathcal{M}$ is obviously of great importance. For physical reasons, we should in principle consider two topologies on $\mathcal{M}$, depending on whether we wish to link the theory with ordinary quantum mechanics, or to study the dynamics in $\mathcal{M}$ in its natural context. A third topology on $\mathcal{M}$ is given by the filtration. We will show that these three topologies are equivalent.

The paper is organized as follows. In Sec. II we review the twisted product and the integral transformation of Wigner, which intertwines the twisted product with the composition of kernel functions. We show how this transformation and the kernel theorem establish a link between $\mathcal{M}$ and the algebra $\mathcal{L}_b(\mathcal{S}_1)$. We also sketch how the Wigner transformation and the twisted product yield a constructive proof of the Stone–von Neumann theorem. In Sec. III we introduce two topologies on $\mathcal{M}$, regarding $\mathcal{M}$ first as an operator algebra over a space of test functions on configuration space, and second as an operator algebra over functions on phase space; and we prove the equivalence of these two topologies. In Sec. IV we introduce the filtrations of the space of tempered distributions (on phase space) and characterize $\mathcal{M}$ in terms of these filtrations. From this characterization, we obtain the third topology on $\mathcal{M}$, and we prove its equivalence with the previous two. In Sec. V, we obtain conditions which imply that certain functions on phase space correspond to trace-class operators in the usual formulation of quantum mechanics.

II. TWISTED PRODUCTS AND THE KERNEL THEOREM

In the usual approach to phase-space quantum mechanics, via the Weyl correspondence between functions and operators,[2–4] L^2 functions correspond to Hilbert–Schmidt operators; since these have L^2 kernels, we may relate the twisted product to the composition of integral kernels by some transformation of $L^2(\mathbb{R}^{2N})$ onto itself. This transformation turns out to be the prescription introduced by Wigner[5,6] to associate a "distribution function" on phase space to a Schrödinger wave function. Moreover, it maps the twisted Hermite basis of I onto the ordinary Hermite basis for $L^2(\mathbb{R}^{2N})$. Also, we can build on the observation by Cressman[7] that the Wigner transformation allows us to transfer the kernel theorem to the twisted product calculus, and in this way we identify the Moyal *-algebra $\mathcal{M}$ in terms of more familiar spaces.

We use the same notations as I. Also, we write $\mathcal{S}_1 = \mathcal{S}(\mathbb{R})$ and $\mathcal{S}_2 = \mathcal{S}(\mathbb{R}^2)$ to denote the spaces of rapidly decreasing smooth functions on $\mathbb{R}$ and $\mathbb{R}^2$, and $\mathcal{S}'_1, \mathcal{S}'_2$ for their dual spaces of tempered distributions. When $\phi, \psi \in \mathcal{S}_1$, we define $\phi \otimes \psi \in \mathcal{S}_2$ by $(\phi \otimes \psi)(q,p) := \phi(q)\psi(p)$. If E,F are locally convex spaces, $\mathcal{L}(E,F)$ will denote the space of continuous linear maps: $E \to F$, which we abbreviate to $\mathcal{L}(E)$ in the case $E = F$.

If $f,g \in \mathcal{S}_2$ [or if $f,g \in L^2(\mathbb{R}^2)$], we write

$$(f \circ g)(x,y) := \frac{1}{\sqrt{4\pi}} \int_{\mathbb{R}} f(x,z) g(z,y)\, dz,$$

which is the "kernel product" of f and g. We also introduce the maps R, Φ, W from $\mathcal{S}_2$ onto $\mathcal{S}_2$ by

$$(Rf)(x,y) := \left(\frac{x+y}{\sqrt{2}}, \frac{x-y}{\sqrt{2}}\right),$$

$$(\Phi f)(x,y) := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x,z) e^{-iyz}\, dz,$$

$$(Wf)(x,y) := (R\Phi^{-1} f)(x,y)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f\left(\frac{x+y}{\sqrt{2}}, z\right) e^{i(x-y)z/\sqrt{2}}\, dz.$$

 0022-2488/88/040880-08$02.50

Clearly R and the "partial Fourier transform" Φ, and hence W, are Fréchet-space isomorphisms of $\mathscr{S}_2$ onto itself, and extend to unitary operators on $L^2(\mathbb{R}^2)$. We call W the *Wigner transformation* on $\mathscr{S}_2$. It intertwines the twisted product and the kernel product on $\mathscr{S}_2$:

$$W(f\times g) = Wf\circ Wg. \tag{1}$$

Indeed, since $W^{-1} = \Phi R^{-1} = \Phi R$, a straightforward computation shows that $W^{-1}(Wf\circ Wg) = f\times g$ for $f,g\in\mathscr{S}_2$. The identity (1) gives the connection between the Weyl operator formalism and the twisted product calculus.

Indeed, more is true: if $\{f_{mn}: m,n\in\mathbb{N}\}$ denotes the "twisted Hermite basis" of $\mathscr{S}_2$ discussed in I,

$$f_{mn}(q,p) = 2(-1)^n\sqrt{\frac{n!}{m!}}(q-ip)^{m-n} \times L_n^{m-n}(q^2+p^2)e^{-(q^2+p^2)/2} \tag{2}$$

(if $m\geqslant n$; $f_{mn} := f_{nm}^*$ otherwise), and if $h_m\in\mathscr{S}_1$ denotes the Hermite function,

$$h_m(x) := (1/\sqrt{2^{m-1}m!})H_m(x)e^{-x^2/2},$$

then a direct calculation shows that

$$W(f_{mn}) = h_m\otimes h_n. \tag{3}$$

We may extend W to $\mathscr{S}_2'$ by dualilty in the usual way. Indeed, writing $\overline{W} := R\Phi$, we find

$$\langle Wf,g\rangle = \langle R\Phi^{-1}f,g\rangle = \langle \Phi^{-1}f,Rg\rangle$$
$$= \langle f,\Phi^{-1}Rg\rangle = \langle f,\overline{W}^{-1}g\rangle, \quad \text{for } f,g\in\mathscr{S}_2,$$

so we define WT for $T\in\mathscr{S}_2'$ by

$$\langle WT,h\rangle := \langle T,\overline{W}^{-1}h\rangle,$$

for $h\in\mathscr{S}_2$. Similarly, we may define $\langle \overline{W}T,h\rangle := \langle T,W^{-1}h\rangle$. Since $\overline{W}^{-1}: \mathscr{S}_2\to\mathscr{S}_2$ is a topological isomorphism, so is its transpose $W: \mathscr{S}_2'\to\mathscr{S}_2'$ (where $\mathscr{S}_2'$ carries the strong dual topology[8,9]).

Now $\mathscr{S}_2$ acts on $\mathscr{S}_1$ by

$$(f\cdot\phi)(x) := \frac{1}{\sqrt{4\pi}}\int_{\mathbb{R}} f(x,y)\phi(y)dy.$$

Note that $\langle f\cdot\phi,\psi\rangle = (1/\sqrt{2})\langle f,\psi\otimes\phi\rangle$, for $f\in\mathscr{S}_2$, $\phi,\psi\in\mathscr{S}_1$. Thus we may define $T\cdot\phi$, for $T\in\mathscr{S}_2'$, $\phi\in\mathscr{S}_1$, by transposition:

$$\langle T\cdot\phi,\psi\rangle := (1/\sqrt{2})\langle T,\psi\otimes\phi\rangle.$$

Remark: We also observe that

$$(\phi|WT\cdot\psi) = \tfrac{1}{2}\langle T,\overline{W}^{-1}(\phi^*\otimes\psi)\rangle. \tag{4}$$

Indeed,

$$(\phi|WT\cdot\psi) = (1/\sqrt{2})\langle\phi^*,WT\cdot\psi\rangle$$
$$= \tfrac{1}{2}\langle WT,\phi^*\otimes\psi\rangle = \tfrac{1}{2}\langle T,\overline{W}^{-1}(\phi^*\otimes\psi)\rangle.$$

The identity (4) is Moyal's connection between the Weyl operator formalism and the calculus of Wigner functions. We may interpret it thus: to calculate the transition probabilities for an observable T between pure states represented by state vectors ψ,ϕ, one may compute the scalar product of the operator $WT\cdot$ with the ket $|\psi\rangle$ and the bra $\langle\phi|$ (in Dirac's terminology), or equivalently one may take the expected value of T with respect to the Wigner "distribution function"

$$\overline{W}^{-1}(\phi^*\otimes\psi) = \int_{\mathbb{R}}\phi^*\left(\frac{q+t}{\sqrt{2}}\right)\psi\left(\frac{q-t}{\sqrt{2}}\right)e^{ipt}\,dt.$$

Furthermore, if $f,g,h\in\mathscr{S}_2$ and if $\tilde{f}(q,p) := f(p,q)$, then

$$\langle f\circ g,h\rangle = \langle f,h\circ\tilde{g}\rangle = \langle g,\tilde{f}\circ h\rangle$$
$$= \frac{1}{\sqrt{2}}\int\!\!\int\!\!\int f(q,t)g(t,p)h(q,p)dt\,dq\,dp,$$

so $\langle T\circ f,h\rangle := \langle T,h\circ\tilde{f}\rangle$ defines $T\circ f$ by transposition, for $T\in\mathscr{S}_2'$, $f\in\mathscr{S}_2$.

Lemma 1: If $T\in\mathscr{S}_2'$, $f\in\mathscr{S}_2$, and $\phi,\psi\in\mathscr{S}_1$, then (i) $W(T\times f) = WT\circ Wf$; (ii) $T\circ(\phi\otimes\psi) = (T\cdot\phi)\otimes\psi$.

Proof: Since $W\overline{W}^{-1}h = R\Phi^{-2}Rh = \tilde{h}$, we get

$$\langle W(T\times f),h\rangle = \langle T,f\times\overline{W}^{-1}h\rangle$$
$$= \langle \overline{W}T,W(f\times\overline{W}^{-1}h)\rangle$$
$$= \langle \overline{W}T,Wf\circ\tilde{h}\rangle = \langle (WT)\tilde{},Wf\circ\tilde{h}\rangle$$
$$= \langle WT,h\circ(Wf)\tilde{}\rangle = \langle WT\circ Wf,h\rangle,$$

for $h\in\mathscr{S}_2$. Also,

$$\langle T\circ(\phi\otimes\psi),h\rangle = \langle T,h\circ(\psi\otimes\phi)\rangle = \langle T,(h\cdot\psi)\otimes\phi\rangle$$
$$= (1/\sqrt{2})\langle T\cdot\phi,h\cdot\psi\rangle = \langle (T\cdot\phi)\otimes\psi,h\rangle.$$

Writing $(ZT)(\phi) := T\cdot\phi$, the kernel theorem[9] for tempered distributions states that Z is an isomorphism from $\mathscr{S}_2'$ onto $\mathscr{L}(\mathscr{S}_1,\mathscr{S}_1')$. We now have the following theorem.

Theorem 1: $ZW(\mathscr{M}_L) = \mathscr{L}(\mathscr{S}_1)$; moreover, $\mathscr{M}_L$ is an algebra under the twisted product, and $ZW: \mathscr{M}_L\to\mathscr{L}(\mathscr{S}_1)$ is an algebra isomorphism.

Proof: $T\in\mathscr{M}_L$ iff $T\times f\in\mathscr{S}_2$, $\forall f\in\mathscr{S}_2$, iff $T\times f\in\mathscr{S}_2$, for all f of the form $W^{-1}(\phi\otimes\psi)$ with $\phi,\psi\in\mathscr{S}_1$, since $\mathscr{S}_1\otimes\mathscr{S}_1$ is dense in $\mathscr{S}_2$ and $f\mapsto T\times f$ is continuous (see I). Thus $T\in\mathscr{M}_L$ iff

$$W(T\times W^{-1}(\phi\otimes\psi)) = WT\circ(\phi\otimes\psi) = (WT\cdot\phi)\otimes\psi$$
$$= ZWT(\phi)\otimes\psi\in\mathscr{S}_2,$$

for all $\phi,\psi\in\mathscr{S}_1$, iff $ZWT(\phi)\in\mathscr{S}_1$ for all $\phi\in\mathscr{S}_1$. This last statement holds since the reduction map $\chi\otimes\psi\mapsto\langle R,\psi\rangle\chi$, for any $R\in\mathscr{S}_1'$, extends by linearity and continuity to the completed projective tensor product $\mathscr{S}_1\hat{\otimes}\mathscr{S}_1\cong\mathscr{S}_2$, and hence maps $\mathscr{S}_2$ into $\mathscr{S}_1$ continuously.

Taking T,f as before, and $\chi\in\mathscr{S}_1$, we have

$$ZW(T\times f)(\chi) = Z(WT\circ WF)(\chi) = Z(WT\circ(\phi\otimes\psi))(\chi)$$
$$= Z((WT\cdot\phi)\otimes\psi)(\chi) = ((WT\cdot\phi)\otimes\psi)\cdot\chi$$
$$= WT\cdot\phi\langle\psi,\chi\rangle = WT\cdot(Z(\phi\otimes\psi)(\chi))$$
$$= ZWT(ZWf(\chi)),$$

and [by density of $W(\mathscr{S}_1\otimes\mathscr{S}_1)$ in $\mathscr{S}_2$] we get $ZW(T\times f) = ZW(T)ZW(f)$ for any $f\in\mathscr{S}_2$, $T\in\mathscr{S}_2'$. If $S\in\mathscr{M}_L$, we then have

$$\begin{aligned}\langle ZW(T\times S)(\phi)\psi\rangle &= \langle W(T\times S)\cdot\phi,\psi\rangle\\ &= (1/\sqrt{2})\langle W(T\times S),\psi\otimes\phi\rangle\\ &= (1/\sqrt{2})\langle T,S\times\overline{W}^{-1}(\psi\otimes\phi)\rangle\\ &= (1/\sqrt{2})\langle T,S\times W^{-1}(\phi\otimes\psi)\rangle\\ &= (1/\sqrt{2})\langle \overline{W}T, WS\circ(\phi\otimes\psi)\rangle\\ &= (1/\sqrt{2})\langle \overline{W}T, ZWS(\phi)\otimes\psi\rangle\\ &= (1/\sqrt{2})\langle WT,\psi\otimes ZWS(\phi)\rangle\\ &= \langle ZWT(ZWS(\phi)),\psi\rangle,\end{aligned}$$

for any $\psi\in\mathcal{S}_1$. In particular, if $R\in\mathcal{M}_L$, we get $ZW(R\times S) = ZW(R)ZW(S)\in\mathcal{L}(\mathcal{S}_1)$ and since ZW is one-to-one: $\mathcal{S}_2'\to\mathcal{L}(\mathcal{S}_1,\mathcal{S}_1')$, we conclude that $R\times S\in\mathcal{M}_L$, so that $\mathcal{M}_L$ is in fact an algebra, and ZW: $\mathcal{M}_L\to\mathcal{L}(\mathcal{S}_1)$ is a bijective homomorphism. ■

Remark: By analogous arguments, or more directly by noting that $(WT)^* = \overline{W}(T^*)$, one can show that $Z\overline{W}$: $\mathcal{M}_R\to\mathcal{L}(\mathcal{S}_1)$ is an algebra isomorphism.

Before proceeding, we observe that the Wigner transformation gives a direct, constructive proof of the Stone–von Neumann theorem. We regard $(\mathcal{S}_2,\times)$ as an operator algebra and look at its left regular representation $\pi(f)g := f\times g$. If $f_0(u) := 2e^{-u^2/2}$, define ω: $\mathcal{S}_2\to\mathbb{C}$ by $\omega(g) := (f_0|g\times f_0)$. This ω is linear and continuous on $\mathcal{S}_2$, and $\omega(g^*\times g) = \|g\times f_0\|^2\geqslant 0$. Since the Gaussian function f_0 has the property that $f_0\times g\times f_0 = (f_0|g)f_0$, we have $\omega(g) := (f_0|g)$.

Using the positive functional ω, we can apply the Gelfand–Naimark–Segal construction to $\mathcal{S}_2$. We observe that $\mathcal{K} := \{g\in\mathcal{S}_2: g\times f_0 = g\}$ and $\mathcal{K}_0 := \{g\in\mathcal{S}_2: g\times f_0 = 0\}$ are closed left ideals in $\mathcal{S}_2$, and that if η: $\mathcal{S}_2\to\mathcal{S}_2/\mathcal{K}_0$ is the canonical projection, then $\mathcal{S}_2/\mathcal{K}_0$ becomes a pre-Hilbert space with inner product

$$(\eta(g)|\eta(f)) := \omega(g^*\times f) = (g\times f_0|f\times f_0),$$

whose completion is denoted $\mathcal{H}_\omega$, and that $\eta(g)\to\eta(f\times g)$ extends to an operator $\pi_\omega(f)\in\mathcal{L}(\mathcal{H}_\omega)$ with $\|\pi_\omega(f)\|\leqslant\|f\|$.

It is easily verified that $\eta(f_{mn}) = 0$ in $\mathcal{H}_\omega$ if $n\neq 0$, and that $\{\eta(f_{m0}): m\in\mathbb{N}\}$ is an orthonormal basis for $\mathcal{H}_\omega$. Now $\sum_{m,n=0}^\infty c_{mn}f_{mn}$ lies in $\mathcal{K}$ iff $c\in s$ (see I) and $c_{mn} = 0$ for $n\neq 0$. Thus η: $\mathcal{K}\to\mathcal{H}_\omega$ is isometric if $\mathcal{K}$ is given the norm of $L^2(\mathbb{R}^2)$, and so extends to a unitary map V: $\overline{\mathcal{K}}\to\mathcal{H}_\omega$, where $\overline{\mathcal{K}}$ is the L^2 closure of $\mathcal{K}$.

If

$$g = \sum_{m,n=0}^\infty c_{mn}f_{mn}\in\mathcal{S}_2$$

and $\pi_\omega(g) = 0$, then for all n, we have

$$0 = \eta(g\times f_{n0}) = \sum_{m=0}^\infty c_{mn}\eta(f_{m0}),$$

so that all $c_{mn} = 0$: hence π_ω: $\mathcal{S}_2\to\mathcal{L}(\mathcal{H}_\omega)$ is a faithful representation of $\mathcal{S}_2$. Using Schur's lemma, we can check that π_ω is irreducible.

Define the projector P: $L^2(\mathbb{R})\to L^2(\mathbb{R})$ by $P(\phi\otimes\psi) := (h_0|\psi)\phi$ and Q: $L^2(\mathbb{R})\to L^2(\mathbb{R})\otimes h_0$ by $Q(\phi) := \phi\otimes h_0$. Then PWV^{-1}: $\mathcal{H}_\omega\to L^2(\mathbb{R})$ is unitary, with inverse $VW^{-1}Q$, and $PWV^{-1}(\eta(f_{m0})) = h_m$. Then we may calculate that

$$PW(q\times f_{m0}) = \sqrt{m}\,h_{m-1} + \sqrt{m+1}\,h_{m+1} = \sqrt{2}qh_m,$$

and similarly $PW(p\times f_{m0}) = -\sqrt{2}i\,dh_m/dq$, so that

$$PW(q\times f) = \sqrt{2}qPWf,$$

$$PW(p\times f) = -\sqrt{2}i\frac{d}{dq}PWf, \quad \text{for } f\in\mathcal{K}.$$

Let us write

$$\pi_s(f) := PWV^{-1}\pi_\omega(f)VW^{-1}Q, \quad \text{for } f\in\mathcal{S}_2;$$

we may call π_s the *Schrödinger representation* of $\mathcal{S}_2$ on $L^2(\mathbb{R})$. This brings us to the Stone–von Neumann theorem, which in the present context states that any representation of the twisted product algebra $\mathcal{S}_2$ is equivalent to a multiple of π_s; we show this to be true for the left regular representation π, as a simple consequence of the GNS construction. The unitary equivalence that decomposes π is just the Wigner transformation.

Theorem 2:

$$W\pi(f)W^{-1} = \pi_s(f)\otimes I, \quad \text{for all } f\in\mathcal{S}_2.$$

Proof: From Lemma 1, for $\phi,\psi\in\mathcal{S}_1$ we obtain

$$\begin{aligned}W\pi(f)W^{-1}(\phi\otimes\psi) &= W(f\times W^{-1}(\phi\otimes\psi))\\ &= Wf\circ(\phi\otimes\psi) = (Wf\cdot\phi)\otimes\psi\end{aligned}$$

and also

$$\begin{aligned}\pi_s(f)\phi &= PWV^{-1}\pi_\omega(f)VW^{-1}(\phi\otimes h_0)\\ &= PWV^{-1}V(f\times W^{-1}(\phi\otimes h_0)) = Wf\cdot\phi,\end{aligned}$$

for all $\phi,\psi\in\mathcal{S}_1$. Since the functions $\phi\otimes\psi$ generate a dense subset of $L^2(\mathbb{R})$, we are done. ■

Remarks: (1) In the previous discussion, we may replace $L^2(\mathbb{R})$ by $\mathcal{S}_1$, $L^2(\mathbb{R}^2)$ by $\mathcal{S}_2$, and $\mathcal{S}_2$ by $\mathcal{M}_L$. Then PWV^{-1} is a Fréchet-space isomorphism from $\mathcal{S}^2/\mathcal{K}_0$ onto $\mathcal{S}_1$, and so we can extend the representations π_ω and π_s to $\mathcal{M}_L$ (acting on $\mathcal{S}^2/\mathcal{K}_0$ and $\mathcal{S}_1$, respectively). With π now denoting the left regular representation of $\mathcal{M}_L$ on its left ideal $\mathcal{S}_2$, Theorem 2 remains valid, with the added advantage that we can write $\pi_s(q)\phi = \sqrt{2}q\phi$, $\pi_s(p)\phi = -\sqrt{2}i\,d\phi/dq$ for $\phi\in\mathcal{S}_1$, thus displaying the Schrödinger representation π_s in its familiar form, modulo a normalization factor.

(2) From the proof, we see that

$$\begin{aligned}(\pi_s(f)\phi)(x) &= (WF\cdot\phi)(x)\\ &= \frac{1}{2\pi\sqrt{2}}\int\int_{\mathbb{R}^2} f\left(\frac{x+y}{\sqrt{2}},z\right)\\ &\quad\times e^{i(x-y)z/\sqrt{2}}\phi(y)\,dz\,dy. \qquad (5)\end{aligned}$$

Thus the operator $\pi_s(f)$ is the pseudodifferential operator (in the sense of Hörmander[10]) associated to the "symbol" f. This is precisely Weyl's quantization rule in more fashionable language. If $f = p^2/2m + V(q)$, we get easily from (5), at least formally, the "Schrödinger operator": $\pi_s(f) = -(2/m)\Delta + V(q)$.

III. OPERATOR TOPOLOGIES ON $\mathcal{M}$

Now $\mathcal{L}(\mathcal{S}_1)$ carries a natural topology, that of uniform convergence on bounded subsets of $\mathcal{S}_1$ [under which it is a complete, nuclear reflexive locally convex space,[11,12] usually denoted by $\mathcal{L}_b(\mathcal{S}_1)$]. This is the standard example of a locally convex algebra which is neither Fréchet nor DF. Let $\mathcal{T}_1$ be the unique locally convex topology on $\mathcal{M}_L$ so that ZW: $(\mathcal{M}_L,\mathcal{T}_1)\to\mathcal{L}_b(\mathcal{S}_1)$ is a homeomorphism.

A second method of topologizing $\mathcal{M}_L$ is as follows. If $(\epsilon_u\tau_u f)(v) := e^{iu'Jv}f(v-u)$, we have shown in I that $\epsilon_u\tau_u(T\times f) = T\times(\epsilon_u\tau_u f)$ for all $T\in\mathcal{S}'_2$, $f\in\mathcal{S}_2$, $u\in\mathbb{R}^2$; and moreover, that if L: $\mathcal{S}_2\to\mathcal{E}(\mathbb{R}^2)$ is a continuous linear map commuting with every $\epsilon_u\tau_u$, then $L(f) = T\times f$, $\forall f\in\mathcal{S}_2$, for some $T\in\mathcal{S}'_2$. Thus we find, writing $L_T(f) := T\times f$, that

$$\{L_S\colon S\in\mathcal{M}_L\} = \{L\in\mathcal{L}_b(\mathcal{S}_2)\colon L\epsilon_u\tau_u = \epsilon_u\tau_u L,\ \forall u\in\mathbb{R}^2\}.$$

Let $\mathcal{T}_2$ denote the topology on $\mathcal{M}_L$ so that $S\to L_S$ is a homeomorphism of $(\mathcal{M}_L,\mathcal{T}_2)$ onto this closed subspace of $\mathcal{L}_b(\mathcal{S}_2)$, where the subscript b denotes the topology of uniform convergence on bounded subsets of $\mathcal{S}_2$. Then, we have the following theorem.

Theorem 3: The topologies $\mathcal{T}_1$ and $\mathcal{T}_2$ on $\mathcal{M}_L$ coincide.

We next define the topology of $\mathcal{M}_R$ so that $S\to S^*$ is a topological isomorphism of $\mathcal{M}_L$ onto $\mathcal{M}_R$. Finally, we give $\mathcal{M}$ the natural topology of the intersection $\mathcal{M}_L\cap\mathcal{M}_R$, that is, the weakest locally convex topology so that both inclusions $\mathcal{M}\subset\mathcal{M}_L$, $\mathcal{M}\subset\mathcal{M}_R$ are continuous. From the definition of $\mathcal{T}_1$, it is already known that $\mathcal{M}$ is a complete, nuclear, semireflexive locally convex *-algebra with a separately continuous multiplication and a continuous involution; on the other hand, via $\mathcal{T}_2$, $\mathcal{M}$ may be regarded as an operator *-algebra on $\mathcal{S}_2$. This is particularly useful with a view to solving Schrödinger equations of the form

$$2i\frac{\partial U}{\partial t} = H\times U(t),\quad U(0)=\mathbf{1},$$

where $H\in\mathcal{M}$, $U(t)\in\mathcal{M}$, for $t\in\mathbb{R}$, using operator semigroup theory. Note that the semigroup property gives $U(s+t) = U(s)U(t)$. In this formula H denotes a time-independent Hamiltonian and $U(t)$ is the "twisted exponential"[13,14] or "evolution function"[15] associated with this Hamiltonian, which contains the dynamical information for the system analogously with the Green's function in the conventional formulation.

Proof of Theorem 3: For $i=1,2$ let $\mathcal{T}'_i$ be the locally convex topology on $W(\mathcal{M}_L)$ so that W: $(\mathcal{M}_L,\mathcal{T}_i)\to(W(\mathcal{M}_L),\mathcal{T}'_i)$ is a homeomorphism. It suffices to show that $\mathcal{T}'_1=\mathcal{T}'_2$.

A basic neighborhood of 0 for $\mathcal{T}'_1$ is of the form

$$(A;V)_1 := \{T\in W(\mathcal{M}_L)\colon T\cdot\phi\in V,\ \forall\phi\in A\},$$

where A is a bounded subset of $\mathcal{S}_1$ and V is a zero-neighborhood in $\mathcal{S}_1$. A basic neighborhood of 0 for $\mathcal{T}'_2$ is of the form

$$(B;U)_2 := \{T\in W(\mathcal{M}_L)\colon T\circ f\in U,\ \forall f\in B\},$$

where B is a bounded subset of $\mathcal{S}_2$ and U is a zero-neighborhood in $\mathcal{S}_2$. Since[9] $\mathcal{S}_2=\mathcal{S}_1\hat{\otimes}\mathcal{S}_1$, it suffices[16] to consider B of the form $B=A_1\otimes A_2$, where A_1,A_2 are bounded subsets of $\mathcal{S}_1$, and U of the form $U=\Gamma(V_1\otimes V_2)$ (the closed absolutely convex hull of $V_1\otimes V_2$), where V_1,V_2 are zero-neighborhoods in $\mathcal{S}_1$. Since A_2 is bounded, we can find $r>0$ so that $A_2\subset rV_2$; since $r^{-1}V_1\otimes rV_2 = V_1\otimes V_2$, we can also assume that $A_2\subset V_2$.

For such a given $(B;U)_2$, let $T\in(A_1:V_1)_1$, $\phi\in A_1$, $\psi\in A_2$; then from Lemma 1 we obtain

$$T\circ(\phi\otimes\psi) = (T\cdot\phi)\otimes\psi\in V_1\otimes A_2\subset V_1\otimes V_2\subset U$$

so that $(B;U)_2$ contains $(A_1;V_1)_1$.

On the other hand, let $(A;V)_1$ be given, with V absolutely convex. Choose $\psi\in\mathcal{S}_1$ and $R\in\mathcal{S}'_1$ such that $\langle R,\psi\rangle=1$, and set $V_2 := \{\phi\in\mathcal{S}_1\colon|\langle R,\phi\rangle|<1\}$. If $T\in(A\otimes\{\psi\}$; $\Gamma(V\otimes V_2))_2$ and $\phi\in A$, then $T\circ(\phi\otimes\psi)=(T\cdot\phi)\otimes\psi$ lies in $\Gamma(V\otimes V_2)$, and on applying the reduction map $\chi\otimes\omega\mapsto\langle R,\omega\rangle\chi$, we find that $T\cdot\phi\in V$. Thus $(A;V)_1$ contains a set of the form $(B;U)_2$.

We have shown that the zero-neighborhood bases for $\mathcal{T}'_1$,$\mathcal{T}'_2$ are equivalent, so $\mathcal{T}_1=\mathcal{T}_2$. ■

IV. FILTRATIONS OF $\mathcal{S}'_2$

In this section we introduce several filtrations of $\mathcal{S}'_2$, in terms of which $\mathcal{M}$ may be characterized. We start with a rigged Hilbert space structure which may be defined (in two-dimensional phase space) by a two-parameter family of Hilbert spaces. (See I, Sec V). In I we have shown that the basis functions f_{mn} of (2) are orthonormal with respect to the measure $(4\pi)^{-1}\,dq\,dp$, that $f_{mn}\times f_{kl}=\delta_{nk}f_{ml}$, and that

$$H\times f_{mn}=(2m+1)f_{mn},\quad f_{mn}\times H=(2n+1)f_{mn}. \tag{6}$$

We introduced the Hilbert spaces $\mathcal{G}_{s,t}$ (for $s,t\in\mathbb{R}$) in I as the completions of $\mathcal{S}_2$ with respect to the norm $\|\cdot\|_{s,t}$, where

$$\Bigg\|\sum_{m,n=0}^{\infty}c_{mn}f_{mn}\Bigg\|^2_{st} := \sum_{m,n=0}^{\infty}(2m+1)^s(2n+1)^t|c_{mn}|^2, \tag{7}$$

and we observed that

$$\mathcal{S}_2=\bigcap_{s,t\in\mathbb{R}}\mathcal{G}_{s,t},\quad \mathcal{S}'_2=\bigcup_{s,t\in\mathbb{R}}\mathcal{G}_{s,t} \tag{8}$$

topologically.

Recalling (3) that $W(f_{mn})=h_m\otimes h_n$, we get $ZW(f_{mn})=|h_m\rangle\langle h_n|$, which shows that $ZW(\mathcal{G}_{0,0}) = \mathcal{HS}(L^2(\mathbb{R}))$, the Hilbert–Schmidt operators on $L^2(\mathbb{R})$, on account of (7). Next we note that

$$\begin{aligned}ZWH(h_m)\otimes h_n &= (WH\cdot h_m)\otimes h_n = WH\circ(h_m\otimes h_n)\\ &= W(H\times f_{mn}) = (2m+1)W(f_{mn})\\ &= (2m+1)h_m\otimes h_n,\end{aligned}$$

for all $m,n = 0,1,2,\ldots$; thus $ZWH(h_m) = (2m+1)h_m$. This shows that $(ZWH)^{-1}$ exists and lies in $\mathscr{H}\mathscr{S}(L^2(\mathbb{R}))$.

Theorem 4:

$$\mathscr{M}_L = \bigcap_{s\in\mathbf{R}} \bigcup_{t\in\mathbf{R}} \mathscr{G}_{s,t}, \quad \mathscr{M}_R = \bigcap_{t\in\mathbf{R}} \bigcup_{s\in\mathbf{R}} \mathscr{G}_{s,t}.$$

Proof: We observe that $\mathscr{H}_k := \mathscr{D}((ZWH)^k)$ is a Hilbert space under the norm $\|\cdot\|_k$, where

$$\left|\left| \sum_{m=0}^{\infty} a_m h_m \right|\right|_k^2 := \sum_{m=0}^{\infty} (2m+1)^{2k} |a_m|^2,$$

and that $S_1 = \bigcap_{k=0}^{\infty} \mathscr{H}_k$ topologically.

We now notice that $T\in\mathscr{M}_L$ iff $ZWT\in\mathscr{L}(\mathscr{S}_1)$ iff for all $m\geqslant 0$, there exists $n\geqslant 1$ with $ZWT\in\mathscr{L}(\mathscr{H}_{n-1},\mathscr{H}_m)$ or equivalently

$$A := (ZWH)^m(ZWT)(ZWH)^{-n+1}\in\mathscr{L}(L^2(\mathbb{R})),$$
$$(ZWH)^m(ZWT)(ZWH)^{-n} = A(ZWH)^{-1}\in\mathscr{H}\mathscr{S}(L^2(\mathbb{R})),$$
$$H^{\times m}\times T\times H^{\times -n}\in L^2(\mathbb{R}^2) = \mathscr{G}_{0,0}, \quad T\in\mathscr{G}_{2m,-2n},$$

iff

$$T\in \bigcap_{m=0}^{\infty} \bigcup_{n=0}^{\infty} \mathscr{G}_{2m,-2n} = \bigcap_{s\in\mathbf{R}} \bigcup_{t\in\mathbf{R}} \mathscr{G}_{s,t}.$$

Note further that $S\in\mathscr{M}_L$ iff $S^*\in\mathscr{M}_L$ and $S^*\in\mathscr{G}_{t,s}$ iff $S\in\mathscr{G}_{s,t}$, so that $\mathscr{M}_R = \bigcap_{t\in\mathbf{R}} \bigcup_{s\in\mathbf{R}} \mathscr{G}_{s,t}$ as claimed. ∎

Remark: The observation that $A(ZWH)^{-1}$ is Hilbert–Schmidt for all m and some n is due to Unterberger,[17] who gives it in a slightly different form. (Similar ideas lie behind the fundamental approach to the generalized eigenvalue problem by van Eijndhoven and de Graaf.[18])

We may visualize the conclusions of Theorem 4 by means of the following diagram:

$$\begin{tikzcd}
\mathscr{M}_L \arrow[r] & \mathscr{G}_{s,-\infty} \arrow[r] & \mathscr{G}_{s-1,-\infty} \arrow[r] & \mathscr{S}_2' \\
\mathscr{M}_L' \arrow[u] & \mathscr{G}_{s,t-1} \arrow[u] \arrow[r] & \mathscr{G}_{s-1,t-1} \arrow[u] \arrow[r] & \mathscr{G}_{-\infty,t-1} \arrow[u] \\
\mathscr{G}_{+\infty,t} \arrow[u] \arrow[r] & \mathscr{G}_{s,t} \arrow[u] \arrow[r] & \mathscr{G}_{s-1,t} \arrow[u] \arrow[r] & \mathscr{G}_{-\infty,t} \arrow[u] \\
\mathscr{S}_2 \arrow[u] \arrow[r] & \mathscr{G}_{s,+\infty} \arrow[u] \arrow[r] & \mathscr{M}_R' \arrow[r] & \mathscr{M}_R \arrow[u]
\end{tikzcd}$$

Here we set $\mathscr{G}_{s,-\infty} := \bigcup_{t\in\mathbf{R}} \mathscr{G}_{s,t}$, $\mathscr{G}_{-\infty,t} := \bigcup_{s\in\mathbf{R}} \mathscr{G}_{s,t}$; then $\mathscr{M}_L = \bigcap_{s\in\mathbf{R}} \mathscr{G}_{s,-\infty}$, $\mathscr{M}_R = \bigcap_{t\in\mathbf{R}} \mathscr{G}_{-\infty,t}$. This gives us our third method of topologizing $\mathscr{M}_L$: let $\mathscr{G}_{s,-\infty}$ have the strongest locally convex topology such that all the inclusions $\mathscr{G}_{s,t}\subset\mathscr{G}_{s,-\infty}$ are continuous. Let $\mathscr{T}_3$ denote the weakest locally convex topology on $\mathscr{M}_L$ such that all the inclusions $\mathscr{M}_L\subset\mathscr{G}_{s,-\infty}$ are continuous. On the diagram, reflection in the principal diagonal [i.e., $(s,t)\mapsto(t,s)$] represents complex conjugation, and so the topology of $\mathscr{M}_R$ is defined in the analogous way. Thus every arrow in the diagram represents a continuous (and dense) inclusion.

[We include the left column and bottom row for completeness, although they are proved in a separate article.[19] We write $\mathscr{G}_{+\infty,t} := \bigcap_{s\in\mathbf{R}} \mathscr{G}_{s,t}$ (topologically) and in Ref. 19 we have shown that the dual space $\mathscr{M}_L' = \bigcup_{t\in\mathbf{R}} \mathscr{G}_{+\infty,t}$ (topologically) and that $\mathscr{M}_L'\subset\mathscr{M}_L$. Analogous results hold for the bottom row of the diagram.]

Note that the characterization of $\mathscr{M}$ given by Theorem 4 enables us to check whether a distribution given in the form

$$T = \sum_{m,n=0}^{\infty} c_{mn} f_{mn} \in \mathscr{S}_2'$$

(whenever this series converges in $\mathscr{S}_2'$) belongs to $\mathscr{M}$ or not. For example, if

$$T := \sum_{m,n=0}^{\infty} e^{-m} f_{mn},$$

then

$$\|T\|_{st}^2 = \sum_{m=0}^{\infty} (2m+1)^s e^{-2m} \sum_{n=0}^{\infty} (2n+1)^t,$$

which converges iff $t < -1$; thus $T\in\mathscr{M}_L$ but $T\notin\mathscr{M}_R$. (Also, $T^* = \sum_{m,n=0}^{\infty} e^{-n} f_{mn}$ lies in $\mathscr{M}_R$ but not in $\mathscr{M}_L$.) This shows that $\mathscr{M}$, $\mathscr{M}_L$, $\mathscr{M}_R$, and $\mathscr{S}_2'$ are distinct.

We may also filtrate $\mathscr{S}_2'$ by other types of Banach spaces, corresponding not to Hilbert–Schmidt operators but to trace-class or bounded operators on $L^2(\mathbb{R})$. Since ZW is an isomorphism between $\mathscr{G}_{0,0}$ and $\mathscr{H}\mathscr{S}(L^2(\mathbb{R}))$, we introduce

$$\mathscr{I}_{s,t} := \mathscr{G}_{s,0}\times\mathscr{G}_{0,t} := \{f\times g\colon f\in\mathscr{G}_{s,0},\ g\in\mathscr{G}_{0,t}\}.$$

From (6) and (7) we find that

$$\mathscr{I}_{s,t} = \{H^{\times(-s/2)}\times f\times g\times H^{\times(-t/2)}\colon f,g\in\mathscr{G}_{0,0}\}$$
$$= \{h\times k\colon h\in\mathscr{G}_{s,q},\ k\in\mathscr{G}_{-q,t}\}, \quad \text{for any } q\in\mathbb{R}.$$

We have $h\in\mathscr{I}_{0,0}$ iff ZWh is a trace-class operator, and so (by polar decomposition) we can find $u\in\mathscr{S}_2'$ with $u^*\times u = \mathbb{1}$ and $|h|\in\mathscr{I}_{0,0}$ so that $h = u\times|h|$ and $|h| = f^*\times f$ with $f\in\mathscr{G}_{0,0}$. Writing $\|h\|_{00,1} := \langle\mathbb{1},|h|\rangle = \|f\|_{00}^2$, we see that $\mathscr{I}_{0,0}$ is a Banach space and the inclusion $\mathscr{I}_{0,0}\subset\mathscr{G}_{0,0}$ is continuous. Since, by (7), $\|g\|_{st} = \|H^{\times(s/2)}\times g\times H^{\times(t/2)}\|_{00}$, for $g\in\mathscr{G}_{s,t}$, we may define

$$\|f\|_{st,1} := \|H^{\times(s/2)}\times f\times H^{\times(t/2)}\|_{00,1}$$

yielding the continuous inclusion $\mathscr{I}_{s,t}\subset\mathscr{G}_{s,t}$ for all $s,t\in\mathbb{R}$.

Next let $\mathscr{B}_{s,t}$ be the dual space of $\mathscr{I}_{-t,-s}$ under the dualilty $(T|f) := \frac{1}{2}\langle T^*,f\rangle$, for $T\in\mathscr{B}_{s,t}$, $f\in\mathscr{I}_{-t,-s}$. The space $\mathscr{B}_{s,t}$ is naturally imbedded in $\mathscr{S}_2'$. In fact, we may identify $\mathscr{B}_{s,t}$ with $\{T\in\mathscr{S}_2'\colon \langle T^*,f\rangle \text{ is finite for all } f\in\mathscr{I}_{-t,-s}\}$. Note also[20] that $\mathscr{B}_{0,0}$ is the space of tempered distributions whose twisted product with any function in $L^2(\mathbb{R}^2)$ lies in $L^2(\mathbb{R}^2)$. The norm

$$\|T\|_{\text{op}} := \sup\{\|T\times f\|_{00}/\|f\|_{00}\colon f\in L^2(\mathbb{R}^2),\ f\neq 0\}$$

on $\mathscr{B}_{0,0}$ coincides with its norm as the dual space of $\mathscr{I}_{0,0}$ (so we have incidentally proved that the dual of the space of trace-class operators on a separable Hilbert space is the space of bounded operators).

Since $\mathscr{G}_{s,t}$ and $\mathscr{G}_{-t,-s}$ are dual Hilbert spaces under the duality

$$\left(\sum_{m,n=0}^{\infty} c_{mn} f_{mn} \,\middle|\, \sum_{k,l=0}^{\infty} d_{kl} f_{kl}\right) := \sum_{m,n=0}^{\infty} c_{mn}^{*} d_{mn},$$

we obtain a chain of continuous inclusions

$$\mathscr{S}_2 \subset \mathscr{I}_{s,t} \subset \mathscr{G}_{s,t} \subset \mathscr{B}_{s,t} \subset \mathscr{S}'_2, \tag{9}$$

where we define the norm of $\mathscr{B}_{s,t}$ as the dual norm to $\|f\|_{-t,-s,1}$. Denoting this norm by $\|\cdot\|_{st,\infty}$, we obtain

$$\begin{aligned}\|f\|_{st,1} &= \|H^{\times(s/2)} \times f \times H^{\times(t/2)}\|_{00,1} \\ &\leqslant \|H^{\times(s/2)}\|_{00,1} \|f\|_{00,\infty} \|H^{\times(t/2)}\|_{00,1}\end{aligned} \tag{10}$$

by the standard properties of trace-class operators, since $\|f\|_{00,\infty}$ is the norm of ZWf as a bounded operator on $L^2(\mathbb{R})$. So, since

$$\|H^{\times(s/2)}\|_{00,1} = \langle 1, H^{\times(s/2)} \rangle = \sum_{n=0}^{\infty} (2n+1)^{s/2},$$

which converges iff $s < -2$, we conclude from (10) that $\mathscr{B}_{0,0} \subset \mathscr{I}_{s,t}$ if $s < -2$, $t < -2$, and thus that $\mathscr{B}_{p,q} \subset \mathscr{I}_{s,t}$ (continuously) if $p > s+2$, $q > t+2$. Interpolating $\mathscr{S}_2 \subset \mathscr{B}_{p,q} \subset \mathscr{I}_{s,t}$ in (9), we see that we can replace $\mathscr{G}_{s,t}$ by $\mathscr{I}_{s,t}$ or $\mathscr{B}_{s,t}$ in (8). In effect, if we define

$$\mathscr{I}_{s,-\infty} := \cup_{t\in\mathbb{R}} \mathscr{I}_{s,t}, \quad \mathscr{B}_{s,-\infty} := \cup_{t\in\mathbb{R}} \mathscr{B}_{s,t},$$

we find that

$$\mathscr{B}_{p,-\infty} \subset \mathscr{I}_{s,-\infty} \subset \mathscr{G}_{s,-\infty} \subset \mathscr{B}_{s,-\infty}, \quad \text{if } p > s+2,$$

and so

$$\mathscr{M}_L = \bigcap_{s\in\mathbb{R}} \mathscr{G}_{s,-\infty} = \bigcap_{s\in\mathbb{R}} \mathscr{B}_{s,-\infty} = \bigcap_{s,\in\mathbb{R}} \mathscr{I}_{s,-\infty}$$

topologically, if all intersections have the natural (projective) topologies. In particular, the topology $\mathscr{T}_3$ is induced by $\cap_{s\in\mathbb{R}} \cup_{t\in\mathbb{R}} \mathscr{B}_{s,t}$.

Theorem 5: The topologies $\mathscr{T}_2$ and $\mathscr{T}_3$ on $\mathscr{M}_L$ coincide.

Proof: Let us write

$$(B;U) := \{S \in \mathscr{M}_L\colon\ S \times f \in U,\ \forall f \in B\},$$

where U is a neighborhood of 0 in $\mathscr{S}_2$ and B is a bounded set in $\mathscr{S}_2$; then $(B;U)$ is a basic neighborhood of 0 for $\mathscr{T}_2$.

A basic neighborhood of 0 for $\mathscr{T}_3$ is given by a subset $V \subset \mathscr{M}_L$ such that $V = W \cap \mathscr{M}_L$, where W is a neighborhood of 0 in $\mathscr{G}_{s,-\infty}$ for some $s\in\mathbb{R}$. Then $W \cap \mathscr{G}_{s,-q}$ is a neighborhood of 0 in $\mathscr{G}_{s,-q}$ for all $q\in\mathbb{R}$ and so contains a set of the form $\{S \in \mathscr{G}_{s,-q}\colon\ \|S\|_{s,-q} < C(q)\}$, for all q.

Given $(B;U)$, a neighborhood of 0 for $\mathscr{T}_2$, we may suppose that $U = \{g \in \mathscr{S}_2\colon\ \|g\|_{st} \leqslant \delta\}$, for some $s,t\in\mathbb{R}$, $\delta > 0$, and that

$$B = \{f \in \mathscr{S}_2\colon\ \|f\|_{qp} \leqslant A(q,p),\ \forall q,p \in \mathbb{R}\},$$

where $A(q,p) > 0$ is a suitable function. Set

$$V := \bigcup_{q\in\mathbb{R}} \{S \in \mathscr{M}_L \cap \mathscr{G}_{s,-q}\colon\ \|S\|_{s,-q} \leqslant \delta / A(q,t)\}.$$

Then $V = W \cap \mathscr{M}_L$, where $W \cap \mathscr{G}_{s,-q}$ contains a ball of radius $\delta/A(q,t)$, for all q. Let $f \in B$, $S \in V \cap \mathscr{G}_{s,-q}$; then we get

$$\|S \times f\|_{st} \leqslant \|S\|_{s,-q} \|f\|_{qt} \leqslant (\delta / A(q,t)) A(q,t) = \delta,$$

so that $V \cap \mathscr{G}_{s,-q} \subset (B;U)$ for all q, and thus $V \subset (B;U)$.

On the other hand, we observe that each Hilbert space $\mathscr{G}_{s,t}$ is "strictly webbed" in the sense of de Wilde[21]; this property is preserved under countable inductive limits (here $\mathscr{G}_{s,-\infty} = \cup_{n\in\mathbb{N}} \mathscr{G}_{s,-n}$) and under countable projective limits; hence $(\mathscr{M}_L, \mathscr{T}_3) = \cap_{m\in\mathbb{N}} \mathscr{G}_{m,-\infty}$ is strictly webbed.

Since $(\mathscr{M}_L, \mathscr{T}_1)$ is a Montel space[11] and $\mathscr{T}_1 = \mathscr{T}_2$, $(\mathscr{M}_L, \mathscr{T}_2)$ is barreled and complete: furthermore, the identity map: $(\mathscr{M}_L, \mathscr{T}_3) \to (\mathscr{M}_L, \mathscr{T}_2)$ is continuous, by the first part of the proof. Now de Wilde's open mapping theorem[21] shows that this map is a homeomorphism, and so $\mathscr{T}_3 = \mathscr{T}_2$. ■

Remarks: (1) As a corollary, we find that the space $\mathscr{L}_b(\mathscr{S}_1)$ is strictly webbed.

(2) We may summarize the topological properties of the Moyal algebra as follows: (i) $\mathscr{M}_L$ and $\mathscr{M}_R$ are complete, nuclear, reflexive locally convex algebras with a hypocontinuous multiplication; (ii) $\mathscr{M}$ is a complete, nuclear, semireflexive locally convex $*$-algebra with a hypocontinuous multiplication and continuous involution; and (iii) using the technique outlined in Sec. IV of I, it is readily seen that $\mathscr{M}_L$, $\mathscr{M}_R$, and $\mathscr{M}$ are Fourier-invariant normal spaces of distributions.

(3) The technique of filtrating $\mathscr{S}'_2$ by Hilbert spaces used here to define $\mathscr{T}_3$ may be employed to show that the dual space $\mathscr{M}'_L$ can be represented as a dense ideal of $\mathscr{M}_L$ (with a continuous multiplication). Indeed, it can be shown that $\mathscr{M}'_L = \cup_{t\in\mathbb{R}} \cap_{s\in\mathbb{R}} \mathscr{G}_{s,t}$ and that

$$\mathscr{M}'_L = \{f \times T\colon\ f \in \mathscr{S}_2,\ T \in \mathscr{S}'_2\} \subset \mathscr{O}_T;$$

it follows that $\mathscr{S}_2 \subset \mathscr{M}'_L \subset \mathscr{M}_L$ and that $\mathscr{M}'_L$ is an ideal. This we do in a following paper.[18]

V. DISTRIBUTIONS CORRESPONDING TO TRACE-CLASS AND BOUNDED OPERATORS

The intermediate spaces $\mathscr{G}_{s,t}$ and $\mathscr{B}_{s,t}$ are useful for several purposes. We may, for example, obtain information about certain functions or tempered distributions by determining in which of these spaces they lie. For instance, the identity $\mathbf{1}$ for the twisted product lies in $\mathscr{B}_{0,0}$ (of course), but we may also compute from (7) that, since $\mathbf{1} = \Sigma_{n=0}^{\infty} f_{nn}$, then

$$\|\mathbf{1}\|_{st}^2 = \sum_{n=0}^{\infty} (2n+1)^{s+t}$$

and hence $\mathbf{1} \in \mathscr{G}_{s,t}$ iff $s + t < -1$.

The space $\mathscr{I}_{0,0}$ corresponds to the trace-class operators on $L^2(\mathbb{R})$. The question of which functions give rise to nuclear operators, via the Weyl correspondence, has been studied by Daubechies.[22,23] She has identified, for the present case of a two-dimensional phase space, a class of spaces $\mathscr{W}^r$ such that $\mathscr{W}^r \subset \mathscr{I}_{0,0}$ for $r > 1$, using a coherent-state representation of quantum mechanics. From our point of view, $\mathscr{W}^r$ essentially consists of functions f on $\mathbb{R}^2$ with $\langle f^*, A^r f \rangle$ finite, where $Af := H \times f + f \times H$. Since $f \mapsto H \times f$ and $f \mapsto f \times H$ are commuting positive operators (on $\mathscr{G}_{0,0}$, say), we find that

$$0 \leqslant \langle f^*, H^{\times r} \times f \rangle \leqslant \langle f^*, A^r f \rangle$$

and

$$0 \leqslant \langle f^*, f \times H^{\times r} \rangle \leqslant \langle f^*, A^r f \rangle, \quad \text{for } r \geqslant 1,$$

and thus $\mathscr{W}^r \subset \mathscr{G}_{r,0} \cap \mathscr{G}_{0,r}$. We can now show an improved result.

Theorem 6:

$$\mathscr{G}_{r,0} \cup \mathscr{G}_{0,r} \subset \mathscr{I}_{0,0} \quad \text{if } r > 1.$$

Proof: We need only show that $\mathscr{G}_{r,0} \subset \mathscr{I}_{0,0}$. Take $f \in \mathscr{G}_{r,0}$, and write

$$f = \sum_{m,n=0}^{\infty} c_{mn} f_{mn}$$

with

$$\|f\|_{r0}^2 = \sum_{m,n=0}^{\infty} (2m+1)^r |c_{mn}|^2$$

finite. Define d_m by $d_m \geqslant 0$,

$$d_m^2 := (2m+1)^r \sum_{n=0}^{\infty} |c_{mn}|^2.$$

Then

$$\sum_{m=0}^{\infty} d_m^2 = \|f\|_{r0}^2$$

so that

$$g := \sum_{m=0}^{\infty} d_m f_{mm} \in \mathscr{G}_{0,0}.$$

Define

$$h := \sum_{m,n=0}^{\infty} b_{mn} f_{mn},$$

where $b_{mn} := c_{mn}/d_m$. We now observe that

$$\sum_{m,n=0}^{\infty} |b_{mn}|^2 = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{|c_{mn}|^2}{d_m^2} = \sum_{m=0}^{\infty} (2m+1)^{-r} = (1-2^{-r})\zeta(r),$$

so that $h \in \mathscr{G}_{0,0}$ for $r > 1$. Thus $g \times h$ is defined and lies in $\mathscr{I}_{0,0}$, and it is clear that $f = g \times h$.

Furthermore, $\|g \times h\|_{00,1} \leqslant \|g\|_{00} \|h\|_{00}$ by a well-known property of trace-class operators (transferred via the isomorphism ZW to the present context), so we get the estimate

$$\|f\|_{00,1} \leqslant ((1-2^{-r})\zeta(r))^{1/2} \|f\|_{r0},$$

for $f \in \mathscr{G}_{r,0}$. (Replacing f by f^*, we see that the analogous estimate is valid for $f \in \mathscr{G}_{0,r}$.)

Remarks: (1) In the same vein, we observe that all distributions in $\mathscr{B}_{0,0}$ lie in $(\cup_{r>1} \mathscr{G}_{-r,0}) \cap (\cup_{r>1} \mathscr{G}_{0,-r})$ with estimates

$$\|T\|_{-r,0} \leqslant ((1-2^{-r})\zeta(r))^{1/2} \|T\|_{00,\infty},$$
$$\|T\|_{0,-r} \leqslant ((1-2^{-r})\zeta(r))^{1/2} \|T\|_{00,\infty}.$$

This is the tightest constraint of which we are aware, on the class of distributions corresponding to bounded operators by the Weyl rule. We remark that our proofs are simpler than those of Ref. 22 since they merely involve manipulation of the double series introduced in I.

(2) We have noted in I that the twisted product of two square integrable functions in $\mathbb{R}^2$ lies in $C_0(\mathbb{R}^2)$. Thus if $f \in \{\mathscr{G}_{s,t} : s \geqslant 0, t \geqslant 0, s+t \geqslant 2\}$, then $f \in C_0(\mathbb{R}^2)$. Then the Leibniz formula assures us that $f \times g \in C_0^m(\mathbb{R}^2)$ whenever

$$f, g \in \cup \{\mathscr{G}_{s,t} : s > 2m, \ t > 2m, \ s+t > 4m+2\};$$

analogously to what happens in the usual Sobolev spaces, the distributions in $\mathscr{G}_{s,t}$ grow more regular as s,t become larger in a suitable way.

VI. OUTLOOK

The formalism developed in I and the present paper puts forward a mathematical framework for phase-space quantum mechanics, in which the usual calculus of unbounded operators is replaced by a calculus of distributions on phase space, with some techniques of locally convex space theory hovering in the background. Using this framework, we have shown elsewhere[24] that the evolution functions corresponding to any quadratic Hamiltonian on phase space belong to the Moyal *-algebra $\mathscr{M}$. To obtain more general results, when the Schrödinger equation is not exactly solvable, we need an appropriate spectral theorem for $\mathscr{M}$ in order to apply semigroup theory. As a step in this direction, we have identified the dual space of $\mathscr{M}$ as a function space.[19] We hope to develop these aspects further in a forthcoming paper.

ACKNOWLEDGMENTS

We are grateful for helpful correspondence in connection with the present work from Seán Dineen, who brought Ref. 12 to our attention, and also from John Horváth and Peter Wagner. We would like to thank Professor Abdus Salam and the International Centre for Theoretical Physics, Trieste, for their hospitality during a stay in which work was completed.

We gratefully acknowledge support from the Vicerrectoría de Investigación of the Universidad de Costa Rica.

[1] J. M. Gracia-Bondía and J. C. Várilly, J. Math. Phys. **29**, 869 (1988).
[2] J. P. Amiet and P. Huguenin, *Mécaniques classique et quantique dans l'espace de phase* (Université de Neuchâtel, Neuchâtel, 1981).
[3] R. F. V. Anderson, J. Funct. Anal. **9**, 423 (1972).
[4] J. C. T. Pool, J. Math. Phys. **7**, 66 (1966).
[5] E. P. Wigner, Phys. Ref. **40**, 749 (1932).
[6] J. C. Várilly and J. M. Gracia-Bondía, J. Math. Phys. **28**, 2390 (1987).
[7] R. Cressman, J. Funct. Anal. **22**, 405 (1976).
[8] J. Horváth, *Topological Vector Spaces and Distributions I* (Addison–Wesley, Reading, MA, 1966).
[9] F. Treves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York, 1967).
[10] L. Hörmander, Commun. Pure Appl. Math. **32**, 359 (1979).

[11] A. Grothendieck, *Produits tensoriels topologiques et espaces nucleaires* (Am. Math. Soc., Providence, RI, 1955).
[12] D. Vogt, in *Functional Analysis: Surveys and Recent Results III* (North-Holland, Amsterdam, 1984), p. 349.
[13] F. Bayen and J. M. Maillard, Lett. Math. Phys. **6**, 491 (1982).
[14] P. Huguenin, Lett. Math. Phys. **2**, 321 (1978).
[15] J. M. Gracia-Bondía, Phys. Rev. A **30**, 691 (1984).
[16] H. H. Schaefer, *Topological Vector Spaces* (Macmillan, New York, 1966).
[17] J. Unterberger, C. R. Acad. Sci. Ser. **290** A, 1053 (1984).
[18] S. J. L. van Eijndhoven and J. de Graff, J. Funct. Anal. **63**, 74 (1985); *Lecture Notes in Mathematics*, Vol. 1162 (Springer, Heidelberg, 1985).
[19] J. M. Gracia-Bondía and J. C. Várilly, "The dual space of the algebra $\mathscr{L}_b(\mathscr{S})$," to appear.
[20] J.-B. Kammerer, J. Math. Phys. **27**, 529 (1986).
[21] G. Köthe, *Topological Vector Spaces II* (Springer, Heidelberg, 1979).
[22] I. Daubechies, J. Math. Phys. **24**, 1453 (1983).
[23] I. Daubechies, Commun. Math. Phys. **75**, 229 (1980).
[24] J. M. Gracia-Bondía, M. Gadella, L. M. Nieto-Calzada, and J. C. Várilly, "Feynman integrals, quadratic Hamiltonians, and Poincaré's generating function in the phase-space approach to Quantum Mechanics," to appear.

The Trotter–Lie product formula for Gibbs semigroups

V. A. Zagrebnov
Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, Dubna 141980, USSR

(Received 16 June 1987; accepted for publication 28 October 1987)

The trace-norm (and Hilbert–Schmidt-norm) convergence of the Trotter–Lie formula is proved for some classes of Gibbs semigroups.

I. INTRODUCTION

Since the discovery of the Trotter–Lie (TL) formula[1,2]

$$\lim_{n\to\infty} [\exp(-(t/n)A)\exp(-(t/n)B)]^n = \exp(-tC), \tag{1.1}$$

and, more particularly, of its connection with the path (functional) integration and the Feynman–Kac formula, it has permeated through many branches of mathematical physics.[3–5] Until now, the main problem was to "relax," or generalize, the hypotheses under which formula (1.1) holds,[6–9] or to reformulate (1.1) for applications to a modified Feynman integral.[10,11]

A solution of the first part of the problem generally implies that one ought to (i) find the topology in which the convergence in (1.1) will take place, (ii) establish the set of operator pairs A and B for which the limit in (1.1) exists, and (iii) identify the operator C and describe how it can be reconstructed from A and B.

As an example of realization of this program, one can refer to the following (important further) result for strongly continuous semigroups, to which we will return shortly.

Proposition 1.1 [8]: Let A,B be non-negative self-adjoint operators in a Hilbert space $\mathscr{H}$. Let $D(A^{1/2})\cap D(B^{1/2})$ be dense in $\mathscr{H}$. Then, for $t\geqslant 0$, the limit in (1.1) exists in the strong-operator topology, and C is the form sum of A and B:

$$\operatorname*{s-lim}_{n\to\infty} [\exp(-(t/n)A)\exp(-(t/n)B)]^n = \exp\{-t(A\dotplus B)\}. \tag{1.2}$$

The purpose of this paper is to prove some versions of the TL formula for Gibbs semigroups.

To formulate the problem more precisely, we recall some notation and definitions. Let $\mathscr{H}$ be a separable Hilbert space. Then the trace-class ($\mathscr{C}_1(\mathscr{H})$) and Hilbert–Schmidt ($\mathscr{C}_2(\mathscr{H})$) ideals in the Banach space of compact operators $\mathrm{Com}(\mathscr{H})$ are defined by the trace norm $\|A\|_1 = \mathrm{Tr}(A^*A)^{1/2}$ and the Hilbert–Schmidt norm $\|A\|_2 = (\mathrm{Tr}(A^*A))^{1/2}$, respectively. We have

$$\mathscr{C}_1(\mathscr{H})\subset\mathscr{C}_2(\mathscr{H})\subset\cdots\subset\mathrm{Com}(\mathscr{H})^1.$$

Definition 1.1 [12]: A strongly continuous semigroup Γ_t in a separable $\mathscr{H}$ is called a Gibbs semigroup if Γ_t: $\mathbb{R}^1_+ \to \mathscr{C}_1(\mathscr{H})$.

Remark 1.1: From the continuity of multiplication

$$A_nB_n \xrightarrow{\|\cdot\|_p} AB \text{ if } A_n \xrightarrow{s} A, \quad B_n \xrightarrow{\|\cdot\|_p} B, \quad \text{for } 1\leqslant p\leqslant\infty$$

(see Refs. 13 and 14), it follows that the Gibbs semigroup is $\|\cdot\|_1$-continuous for $t>0$.

These semigroups naturally arise in quantum statistical mechanics (QSM) as one-parameter, strongly continuous self-adjoint semigroups generated by the Hamiltonian H: $\Gamma_t = \exp(-tH)$. Here the parameter $t>0$ is nothing but the inverse temperature of a system described by H.

The $\|\cdot\|_1$-norm perturbation theory for Gibbs semigroups was developed in Refs. 15–17. The compactness and convergence of the families of Gibbs semigroups $\{\Gamma_t^{(\alpha)}\}_{\alpha\in M}$ in $\|\cdot\|_1$ topology were studied in Refs. 12, 18, and 19. As the TL formula is often used in QSM under the Tr, it is necessary to prove (1.1) for Gibbs semigroups in the natural (for this case) $\|\cdot\|_1$ topology (see Remark 1.1).

Therefore, in its full generality, the problem can be formulated as whether one can "lift" the convergence in, for example, Eq. (1.2) to a $\|\cdot\|_1$ norm when operator A, or B, or both, generate a Gibbs semigroup. The present paper contains only a partial solution of this problem.

II. TROTTER–LIE FORMULA FOR GIBBS SEMIGROUPS VIA PATH INTEGRAL

This section is inspired by a consideration of self-adjoint Gibbs semigroups which originated from QSM of continuous systems. (For spin systems the Hilbert space $\mathscr{H}$ is finite dimensional, and there is no difference between a $\|\cdot\|_1$ and an s topology.)

To explain the main idea, we start from a one-particle system enclosed in a box $\Lambda\subset\mathbb{R}^\nu$, which is a bounded, open, connected subset of ν-dimensional space with a smooth boundary $\partial\Lambda$. Hence the appropriate Hilbert space is $\mathscr{H}=L^2(\Lambda)$. Then the one-particle kinetic-energy operator $T_\Lambda^{(\sigma)}$ is a self-adjoint extension of the Laplacian $T_\Lambda = (-\Delta)$ with domain $D(T_\Lambda) = C_0^\infty(\Lambda)$. The domain $D(T_\Lambda^{(\sigma)})$ is specified by a boundary condition $\sigma\in C(\partial\Lambda)$. The interaction U is a self-adjoint operator of multiplication by a real-measurable function defined on the domain

$$D(U) = \{\psi\in\mathscr{H} \mid U\psi\in\mathscr{H}\}.$$

Proposition 2.1: Let $U=U_-+U_+$, where $U_-\leqslant 0$, $U_+\geqslant 0$. Assume that U_- is a $T_\Lambda^{(\sigma)}$-bounded operator with a relative bound $b<1$, and that the operator $U_+\in L^1_{\mathrm{loc}}(\Lambda)$. Then the following points obtain.

(a) The Hamiltonian $H_\Lambda^{(\sigma)} = (T_\Lambda^{(\sigma)} + U_-) \dotplus U_+$ is a self-adjoint operator on $\mathscr{H}$ which is semibounded from below.

(b) For $t>0$, $\Gamma_t = \exp(-tH_\Lambda^{(\sigma)})$ belongs to the trace-class operators $\mathscr{C}_1(\mathscr{H})$, and is a strongly continuous semigroup.

(c) The kernel $\Gamma_t(x,y)$, $x,y\in\bar{\Lambda}$, can be represented by the Feynman–Kac formula

 0022-2488/88/040888-04$02.50

$$\Gamma_t(x,y) = \int_{\Omega^t_{x,y}(\Lambda)} d\mu^{t,\sigma}_{x,y}(\omega)\exp\left[-\int_0^t d\tau\, U(\omega(\tau))\right]. \tag{2.1}$$

For the proof we note that point (a) is a well-known consequence of perturbation theory for linear operators and the definition of the sum of operators in the sense of quadratic forms.[20] Point (b) is a consequence of point (a) and of Weyl's min–max principle for $T^{(\sigma)}_\Lambda$ (for details see Ref. 17). Point (c) is one of the celebrated results of the Wiener path-integral theory.[3,4] Here $d\mu^{t,\sigma}_{x,y}(\omega)$ is the conditional Wiener measure on the space $\Omega^t_{x,y}(\Lambda)$ of continuous paths $\omega\subset\bar{\Lambda}$, with end points $\omega(0)=x$, $\omega(t)=y$, generated by the semigroup $\exp(-tT^{(\sigma)}_\Lambda)\in\mathscr{C}_1(\mathscr{H})$.

Theorem 2.1: Let $T^{(\sigma)}_\Lambda$ and U be operators on $\mathscr{H}$ as in Proposition 2.1. If $U(x)$ is almost everywhere continuous in $\bar{\Lambda}$ and is bounded from below, then

$$\lim_{n\to\infty} \mathrm{Tr}[\exp(-(t/n)T^{(\sigma)}_\Lambda)\exp(-(t/n)U)]^n = \mathrm{Tr}\exp(-tH^{(\sigma)}_\Lambda). \tag{2.2}$$

Proof: At the outset, we note that

$$F_t = \exp(-tT^{(\sigma)}_\Lambda)\exp(-tU)\in\mathscr{C}_1(\mathscr{H}).$$

We proceed by first proving (2.2) with U temporarily replaced by $W(x)\in C(\bar{\Lambda})$, and then we make the approximation arguments to obtain the general case. [We shall exploit the fact that, for operator

$$K\in\mathscr{C}_1(L^2(M,d\mu)),$$

there exists kernel

$$K(x,y)\in L^2(M\times M, d\mu\otimes d\mu),$$

such that

$$\mathrm{Tr}\,K = \int d\mu(x)\, K(x,x)$$

(see Ref. 1, VI.6 or Ref. 21).] Using formula (2.1) and the semigroup property for Green's function,

$$p^{(\sigma)}_\Lambda(t;x,y) = \mu^{t,\sigma}_{x,y}(\Omega^t_{x,y}(\Lambda)),$$

we obtain

$$\mathrm{Tr}(F_{t/n})^n - \mathrm{Tr}\,\Gamma_t = \int_{\bar{\Lambda}} dx \int_{\Omega^t_{x,x}(\Lambda)} d\mu^{t,\sigma}_{x,x}(\omega)\,[f^{(W)}_n(\omega) - f^{(W)}(\omega)], \tag{2.3}$$

where

$$f^{(W)}_n(\omega) = \exp\left[-\frac{t}{n}\sum_{j=1}^n W\left(\omega\left(\frac{t}{n}j\right)\right)\right],$$

$$f^{(W)}(\omega) = \exp\left[-\int_0^t d\tau\, W(\omega(\tau))\right].$$

Since, for each $x,y\in\bar{\Lambda}$, the corresponding paths are continuous and $W\in C(\bar{\Lambda})$,

$$\lim_{n\to\infty} f^{(W)}_n(\omega) = f^{(W)}(\omega)$$

pointwise everywhere on $\Omega^t_{x,y}(\Lambda)$. Furthermore,

$$f^{(W)}_n\in L^1(\Omega^t_{x,y}(\Lambda), d\mu^{t,\sigma}_{x,y}),$$

and we have

$$f^{(W)}_n(\omega)\leqslant\exp(t\|W\|_\infty).$$

Then, by the Lebesgue-dominated convergence theorem, we obtain

$$L^1(\Omega^t_{x,y}(\Lambda), d\mu^{t,\sigma}_{x,y})\text{-}\lim_{n\to\infty} f^{(W)}_n = f^{(W)}. \tag{2.4}$$

Just the same dominated-convergence-theorem argument can now be applied to the function $\mathscr{F}^{(W)}_n(x)$,

$$\mathscr{F}^{(W)}_n(x) = \int_{\Omega^t_{x,x}(\Lambda)} d\mu^{t,\sigma}_{x,x}(\omega)\,[f^{(W)}_n(\omega) - f^{(W)}(\omega)]\in L^1(\Lambda),$$

since

$$|\mathscr{F}^{(W)}_n(x)|\leqslant 2\exp(t\|W\|_\infty)p^{(\sigma)}_\Lambda(t;x,x)\in L^1(\Lambda),$$

and by (2.4) we have

$$\lim_{n\to\infty}\mathscr{F}^{(W)}_n(x) = 0$$

pointwise in $\bar{\Lambda}$. Hence, taking the limit $n\to\infty$ in (2.3), we finish the proof for $W\in C(\bar{\Lambda})$. To remove the continuity assumption, we introduce a sequence of approximations

$$U_L = \min(U,L)\in C(\bar{\Lambda}),$$

which monotonically (almost everywhere) converges to U when $L\to\infty$. Let $\{\Gamma^{(L)}_t\}_L$ and $\{F^{(L)}_t\}_L$ be sequences corresponding to the cutoff interaction U_L. Then the monotonicity implies

$$\|\cdot\|_1\text{-}\lim_{L\to\infty}\Gamma^{(L)}_t = \Gamma_t.$$

[Monotonic convergence $U_L\nearrow U$ for $L\to\infty$ implies the convergence

$$\lambda_n(\Gamma^{(L)}_t)\searrow\lambda_n(\Gamma_t)$$

for eigenvalues. This convergence and the relation

$$\mathrm{Tr}\,\Gamma^{(L)}_t = \sum_{n\geqslant 1}\lambda_n(\Gamma^{(L)}_t)$$

ensures the convergence $\Gamma^{(L)}_t\to\Gamma_t$ in $\|\cdot\|_1$ topology (see Refs. 12 and 18).] At the same time we obtain

$$\mathrm{Tr}(F^{(L)}_{t/n})^n\to\mathrm{Tr}(F_{t/n})^n,$$

for $L\to\infty$, uniformly in n. Thus, by the $\varepsilon/3$ argument, we have the desired result for U. ■

Corollary 2.1: The same arguments as above give

$$\lim_{n\to\infty}\mathrm{Tr}[(F_{t/n})^n(F^*_{t/n})^n] = \mathrm{Tr}\,\Gamma_{2t},\quad t>0. \tag{2.5}$$

Now we can combine (2.5) with very general arguments to "lift" the strong-operator convergence in the TL formula (1.2) to convergence in the Hilbert–Schmidt norm.

Theorem 2.2: Let $T_\Lambda^{(\sigma)}$ and U be operators as in Theorem 2.1. Then

$$\|\cdot\|_2\text{-}\lim_{n\to\infty} (F_{t/n})^n = \Gamma_t, \quad t>0. \tag{2.6}$$

Proof: Let us first note that (1.2) for F_t implies

$$\text{s-}\lim_{n\to\infty} (F_{t/n})^n = \text{s-}\lim_{n\to\infty} (F^*_{t/n})^n = \Gamma_t. \tag{2.7}$$

Since (2.5) can be read as

$$\|(F_{t/n})^n\|_2 \to \|\Gamma_t\|_2,$$

we obtain (2.6) by Grümm's convergence theorem.[13,14] ∎

Corollary 2.2: Since multiplication is continuous for sequences in the strong-operator topology, (2.7) gives the symmetrized TL formula

$$\text{s-}\lim_{n\to\infty} (F_{t/2n})^n (F^*_{t/2n})^n = \Gamma_t. \tag{2.8}$$

If we now read (2.5) as

$$\|(F_{t/2n})^n (F^*_{t/2n})^n\|_1 \to \|\Gamma_t\|_1,$$

then (2.8) gives[12,14,19] the symmetrized TL formula for Gibbs semigroups in $\|\cdot\|_1$ topology:

$$\|\cdot\|_1\text{-}\lim_{n\to\infty} (F_{t/2n})^n (F^*_{t/2n})^n = \Gamma_t, \quad t>0. \tag{2.9}$$

III. CONCLUDING REMARKS

Generalization of the results (2.6) and (2.9) to the n-particle case is obvious. It resembles what one should perform for the Feynman–Kac formula in the above case.[22] But quantum statistics causes some additional combinatorial complications.

The following remarks concern a more restrictive (from the viewpoint of QSM) class of Gibbs semigroups. Despite this, they elucidate the background of what we have proved in Sec. II.

Let A and B be self-adjoint operators on $\mathscr{H}$ bounded from below ($A\geqslant 0$, $B\geqslant b$), and let the intersection of the corresponding quadratic form domains, $Q(A)\cap Q(B)$, be dense in $\mathscr{H}$. Furthermore, for $t>0$, let $\exp(-tA)$ be trace class $\mathscr{C}_1(\mathscr{H})$, and let $\exp(-tB)$ be a compact operator whose eigenvectors are denoted by $\{e_j\}_{j=1}^\infty$. We can then prove the following.

Theorem 3.1: If

$$(e_i, \exp(-tA)e_j) \geqslant 0, \tag{3.1}$$

for $i,j = 1,2,\dots$, then ($t>0$)

$$\lim_{n\to\infty} \operatorname{Tr}[\exp(-(t/n)A)\exp(-(t/n)B)]^n = \operatorname{Tr}\exp\{-t(A+B)\}. \tag{3.2}$$

Proof: Since

$$S_t = \exp(-tA)\exp(-tB)\in\mathscr{C}_1(\mathscr{H}),$$

and $(S_{t/n})^n\in\mathscr{C}_1(\mathscr{H})$, we have

$$\operatorname{Tr}(S_{t/n})^n = \sum_{j=1}^{N} (e_j,(S_{t/n})^n e_j) + \sum_{j=N+1}^{\infty} (e_j,(S_{t/n})^n e_j). \tag{3.3}$$

Using (3.1), the last term in (3.3) can be estimated for arbitrary N uniformly in n:

$$\sum_{j=N+1}^{\infty} (e_j,(S_{t/n})^n e_j) \leqslant \exp(-tb) \sum_{j=N+1}^{\infty} (e_j, \exp(-tA)e_j). \tag{3.4}$$

Since

$$\exp\{-t(A+B)\}\in\mathscr{C}_1(\mathscr{H}),$$

then for arbitrary $\varepsilon>0$, there is N_ε large enough so that the strong convergence (1.2), the estimate (3.4), and the $\varepsilon/3$ argument give us the desired result (3.2). ∎

Corollary 3.1: From (3.2) one obtains the TL formula in $\|\cdot\|_2$ norm and the symmetrized TL formula in $\|\cdot\|_1$ norm (see Sec. II).

Remark 3.1: Conditions (3.1) are satisfied for positivity-preserving contraction semigroups.[2,23,24] Therefore (2.6) and (2.8) are extended to a class of positivity-preserving Gibbs semigroups.

In summary, this paper was concerned with the problem of whether the Trotter–Lie product formula (1.1) holds in the trace norm $\|\cdot\|_1$ for Gibbs semigroups. Our (not yet disproved) conjecture was that (1.1) does hold in $\|\cdot\|_1$ topology if at least one of the semigroups involved in (1.1) is a Gibbs semigroup. The paper presented arguments in favor of this conjecture as far as one considers some particular Gibbs semigroups—specifically, the one originating from quantum statistical mechanics.

ACKNOWLEDGMENTS

I would like to thank Michel L. Lapidus for useful correspondence, and J. L. van Hemmen for a careful reading of the manuscript and useful remarks.

[1] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. I, Functional Analysis* (Academic, New York, 1972).
[2] E. B. Davies, *One-Parameter Semigroups* (Academic, London, 1980).
[3] M. Kac, *Integration in Function Spaces and Its Applications*, Lezioni Fermiane (Scuola Normale Superiore, Pisa, 1980).
[4] B. Simon, *Functional Integration and Quantum Physics* (Academic, New York, 1979).
[5] J. Glimm and A. Jaffe, *Quantum Physics* (Springer, New York, 1981).
[6] P. R. Chernoff, "Product formulas, nonlinear semigroups and addition of unbounded operators," Mem. Am. Math. Soc. **140**, 1 (1974).
[7] T. Ichinose, "A product formula and its application to the Schrödinger equation," Publ. Res. Inst. Math. Sci. **16**, 585 (1980).
[8] T. Kato, "Trotter's product formula for an arbitrary pair of self-adjoint semigroups," in *Advances in Mathematics*, Vol. 3, edited by I. Gohberg and M. Kac (Academic, New York, 1978), pp. 185–195.
[9] M. L. Lapidus, "Generalization of the Trotter–Lie formula," Integral Eqs. Operator Theory **4**, 365 (1981).
[10] M. L. Lapidus, "Product formula, imaginary resolvents and modified Feynman integral," in *Proceedings of the Symposium on Pure Mathematics*, Vol. 44 (Am. Math. Soc., Providence, RI, 1985).
[11] M. L. Lapidus, "Product formula for imaginary resolvents with application to a modified Feynman integral," J. Funct. Anal. **63**, 261 (1985).
[12] V. A. Zagrebnov, "On the families of Gibbs semigroups," Commun. Math. Phys. **76**, 269 (1980).
[13] H. R. Grümm, "Two theorems about $\mathscr{C}_p$," Rep. Math. Phys. **4**, 211 (1973).
[14] B. Simon, *Trace Ideals and Their Applications*, London Math. Soc. Lecture Notes, Ser. 35 (Cambridge U.P., Cambridge, 1979).

[15]N. Angelescu, G. Nenciu, and M. Bundaru, "On the perturbation of Gibbs semigroups," Commun. Math. Phys. **42**, 29 (1975).
[16]H. D. Maison, "Analyticity of the partition function for finite quantum systems," Commun. Math. Phys. **22**, 166 (1971).
[17]V. A. Zagrebnov, "On statistical mechanics of systems with highly singular two-body potentials," Ann. Phys. (NY) **102**, 108 (1976).
[18]V. A. Zagrebnov, "On singular potential interactions in quantum statistical mechanics," Trans. Moscow Math. Soc. **41**, 101 (1980).
[19]A. Wehrl, "Three theorems about entropy and convergence of density matrices," Rep. Math. Phys. **10**, 159 (1976).
[20]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1976), 2nd ed.
[21]I. C. Gohberg and M. G. Krein, *Introduction to the Theory of Linear Nonselfadjoint Operators*, Transl. Math. Monographs, Vol. 18 (Am. Math. Soc., Providence, RI, 1969).
[22]J. Genibre, "Some applications of functional integration in statistical mechanics," *Statistical Mechanics and Quantum Field Theory*, edited by C. de Witt and R. Stora (Gordon and Breach, New York, 1971), pp. 327–427.
[23]E. B. Davies, "Properties of the Green's functions of some Schrödinger operators," J. London Math. Soc. **7**, 483 (1973).
[24]O. Bratteli, A. Kishimoto, and D. W. Robinson, "Positivity and monotonicity properties of C_0 semigroups I," Commun. Math. Phys. **75**, 67 (1980).

Exact analytic solutions for the quantum mechanical sextic anharmonic oscillator

A. K. Dutta and R. S. Willey
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 17 November 1987; accepted for publication 25 November 1987)

Exact analytic solutions for low-lying states of the $O(N)$ invariant sextic anharmonic oscillator are presented. The exact results are compared with Gaussian approximations and general $1/N$ expansions.

I. INTRODUCTION

In the course of some work on the ϕ_3^6 quantum field theory, we rediscovered that, for certain combinations of coupling constants, one can find exact analytic solutions for low-lying energy levels for the ϕ_1^6 theory, i.e., for the quantum mechanical sextic anharmonic oscillator. Although our ultimate interest is in the possible exploitation of these results in the field theory problem, we were sufficiently intrigued by the quantum mechanical results to undertake a detailed investigation. New results presented here include generalization of various approximation schemes developed for the quartic anharmonic oscillator to the sextic anharmonic oscillator and confrontation with the exact analytic results available in the latter case; new exact analytic energy values for the N-dimensional, $O(N)$ symmetric sextic anharmonic oscillator; and comparison of large N approximations with exact results available for any N.

II. EXACT SOLUTIONS

The Schrödinger equation in N space dimensions is

$$\left(-\frac{1}{2}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_i}+V(\mathbf{x})\right)\psi(\mathbf{x})=E\psi(\mathbf{x}). \tag{2.1}$$

We will be considering the $O(N)$ invariant sextic potential

$$V(\mathbf{x})=a\mathbf{x}^2+(b/N)(\mathbf{x}^2)^2+(c/N^2)(\mathbf{x}^2)^3. \tag{2.2}$$

Stability for arbitrary a, b requires $c>0$.

In the one-dimensional case ($N=1$, $-\infty\leqslant x\leqslant\infty$), by taking various combinations of positive and negative values for a and b, we generate single-well, double-well, or triple-well potentials (all reflection symmetric).

A. Simplest example: Exact ground state energy and wave function for $N=1$

Make the ansatz

$$\psi(x)=e^{f(x^2)},\quad f(x^2)=\tfrac{1}{2}\beta x^2-\tfrac{1}{4}\alpha x^4. \tag{2.3}$$

Substitute (2.3) into the $N=1$ Schrödinger equation to obtain

$$\begin{aligned}\psi''&=[\beta+(\beta^2-3\alpha)x^2-2\alpha\beta x^4+\alpha^2x^6]e^f\\&=2(V(x)-E)e^f\\&=[-2E+2ax^2+2bx^4+2cx^6]e^f.\end{aligned} \tag{2.4}$$

Matching powers of x^2 in (2.4) gives

$$\alpha=\sqrt{2c},\quad \beta=-b/\sqrt{2c}. \tag{2.5}$$

Equations (2.5) determine the ground state wave function (2.3):

$$(b^2/2c)-3\sqrt{2c}=2a. \tag{2.6}$$

Equation (2.6) is one constraint on the three parameters a, b, and c of the potential required for the ansatz to provide a solution and

$$E_0=-\beta/2=+b/2\sqrt{2c}. \tag{2.7}$$

The ansatz (2.3) does not provide the ground state energy for the general (a,b,c) reflection invariant sextic potential; rather, constraint (2.6) is required. We choose to regard a and c (and the sign of b) as the independent parameters that specify the potential. Then (2.6) determines $|b|$:

$$b=(\operatorname{sgn} b)\sqrt{2c}(2a+3\sqrt{2c})^{1/2} \tag{2.6$'$}$$

and

$$E_0=(\operatorname{sgn} b)\tfrac{1}{2}(2a+3\sqrt{2c})^{1/2}. \tag{2.7$'$}$$

We note[1] in passing that a generalization of the ansatz (2.3) can be used to generate solutions for potentials that are polynomials of degree $4n-2$; the quartic is conspicuously absent.

The ansatz (2.3) can be generalized[2] to $\psi(x)=\phi(x^2)\exp(f)$, which generates solutions with ϕ an nth-order polynomial in x^2 and gives the first n even parity energy levels when b satisfies the constraint $b_{(n)}=(\operatorname{sgn} b)\sqrt{2c}(2a+(4n-1)\sqrt{2c})^{1/2}$. However, to obtain explicit formulas for those n energy levels one has to solve for the n roots of an $n\times n$ determinant (the factored Hill determinant of Ref. 2). Thus the sextic oscillator potential, with coefficients satisfying the "solvability constraint," is intermediate between the case of a truly solvable potential such as the quadratic oscillator, for which one can write an analytic closed form expression for *all* of the energy levels, and the general case, in which one has to diagonalize an infinite matrix to find the *exact* value for *any* energy level. One can also generate odd parity solutions from the ansatz $\psi(x)=x\phi(x^2)\exp(f)$. We do not provide details here because they can be found in Ref. 2; we will calculate the case of general N in Sec. II B.

B. General N

We take advantage of the $O(N)$ symmetry of (2.2) and separate in N-dimensional spherical coordinates[3]

$$[-\tfrac{1}{2}\nabla^2+V(r)]\psi(\mathbf{x})=E\psi(\mathbf{x}), \tag{2.8}$$

 0022-2488/88/040892-09$02.50

$$\psi(\mathbf{x}) = R_l(r)H_l(\hat{x}), \tag{2.9}$$

$$\nabla^2\psi = \left[\frac{d^2}{dr^2} + \frac{N-1}{r}\frac{d}{dr} - \frac{l(l+N-2)}{r^2}\right] \times R_l(r)H_l(\hat{x}). \tag{2.10}$$

Let

$$R(r) = r^{(1-N)/2}\chi(r); \tag{2.11}$$

then

$$-\tfrac{1}{2}\chi'' + [(N-1)(N-3)/8r^2]\chi + [l(l+N-2)/2r^2]\chi + V(r)\chi = E\chi, \tag{2.12}$$

with

$$V(r) = ar^2 + (b/N)r^4 + (c/N^2)r^6. \tag{2.13}$$

For subsequent $1/N$ expansions it is useful to define a scaled radial coordinate

$$r = \sqrt{N}\rho, \quad \chi(r) = \tilde{\chi}(\rho), \tag{2.14}$$

$$\chi'' = \frac{d^2}{dr^2}\chi = \frac{1}{N}\frac{d^2}{d\rho^2}\tilde{\chi} = \frac{1}{N}\tilde{\chi}'', \tag{2.15}$$

and divide by N:

$$-(1/2N^2)\tilde{\chi}'' + \left[\frac{(N-1)(N-3)}{8N^2\rho^2} + \frac{l(l+N-2)}{2N^2\rho^2} + v(\rho)\right]\tilde{\chi} = e\tilde{\chi}, \tag{2.16}$$

$$v(\rho) = (1/N)V(r) = a\rho^2 + b\rho^4 + c\rho^6, \quad e = E/N. \tag{2.17}$$

This (exact) equation is the starting point for the general $1/N$ expansion in Sec. VI.

Now we continue exactly for general N. The ansatz (2.3) is generalized to

$$\tilde{\chi}(\rho) = \rho^\nu\phi(\rho)e^{f(\rho)}, \tag{2.18}$$

$$\nu = (N-1)/2 + l, \quad f(\rho) = N(\tfrac{1}{2}\beta\rho^2 - \tfrac{1}{4}\alpha\rho^4). \tag{2.19}$$

Substitution of ansatz (2.18) and (2.19) into (2.16) yields

$$\alpha = \sqrt{2c}, \quad \beta = -b/\sqrt{2c} \tag{2.20}$$

[same as (2.5)] and a differential equation for $\phi(\rho)$,

$$\phi'' + [(N-1+2l)/\rho + 2N(\beta\rho - \alpha\rho^3)]\phi' + N[(N+2l)\beta + 2E + (\beta^2 - (1 + (2l+2)/N)\alpha - 2a)N\rho^2]\phi = 0. \tag{2.21}$$

Next, substitute a power series for ϕ and derive a three-term recurrence relation satisfied by the coefficients

$$\phi(\rho) = \sum_{n=0}^{\infty} a_n\rho^n, \tag{2.22}$$

$$2(n+1)(N+2l+2n)a_{n+1} + N[2E + (N+2l+4n)\beta]a_n + N^2\alpha\{(\beta^2 - 2a)/\alpha - [1 + (4n+2l-2)/N]\}a_{n-1} = 0. \tag{2.23}$$

For $N=1$ and $l=0, 1$ (parity), the array of coefficients of this infinite set of simultaneous linear equations for the a_n is just the infinite Hill determinant of Ref. 2. Truly solvable potentials lead to two-term recurrence relations that are terminated after m terms by determination of the eigenvalue E_m. To terminate the three-term recurrence relation (2.23) after m terms requires determination of the first m E_k ($k = 0,1,\ldots,m-1$) and an additional solvability condition which is the generalization of the constraint (2.6). Let

$$\gamma \equiv (\beta^2 - 2a)/\alpha = [(b^2/2c) - 2a]/\sqrt{2c}. \tag{2.24}$$

The additional solvability condition required for termination of (2.23) after m terms is

$$\gamma = \gamma_{m,l} \equiv 1 + (4m + 2l - 2)/N. \tag{2.25}$$

For $N=1$ and $l=0,1$ (2.25) is just the condition found in Ref. 2 to factor the infinite Hill determinant into a finite factor times an infinite remainder. In the same way one can see easily that condition (2.25) gives rise to the factorization of the infinite Hill determinant formed from the coefficients of (2.23); this factorization plus the vanishing of the finite factored determinant is the condition for the existence of finite polynomial solutions of (2.21). Thus we see that we can just as well find exact solutions for any N (and l). As before we choose to treat a, sgn b, and c as independent parameters of the potential and write (2.24) and (2.25) as

$$b_{m,l} = (\operatorname{sgn} b)\sqrt{2c}(2a + \gamma_{m,l}\sqrt{2c})^{1/2}. \tag{2.26}$$

C. Particular solutions

Since we want exact solutions for later use and to illustrate the general procedure just described, we carry out the explicit solution for the first two $l=0$ solvability conditions and the first $l=1$ solvability condition. In all cases the initial conditions imposed on (2.23) are $a_{-1} = 0$, $a_0 = 1$. (This implies that all a_{-k} are zero.)

For $\gamma = \gamma_{1,0}$,

$$n = 0, \quad 2Na_1 + N(2E + N\beta)(1) = 0,$$

choose $2E = -N\beta$ and then $a_1 = 0$;

$n = 1$,

$$4(N+2)a_2 + N(2E + (N+4)\beta)(0) + N^2\alpha(\gamma - \gamma_{1,0})(1) = 0.$$

The condition $\gamma = \gamma_{1,0}$ gives $a_2 = 0$. The result is

$$E_{0,0} = N(-\beta/2) = N(\operatorname{sgn} b)\tfrac{1}{2}(2a + (1 + 2/N)\sqrt{2c})^{1/2}. \tag{2.27}$$

For $N=1$ this reduces to (2.7'). The large N expansion is discussed in Sec. III.

For $\gamma = \gamma_{2,0}$, the $n=0$ equation gives $a_1 = -(2E + N\beta)/2$. Then in the $n=1$ equation, E is chosen to make $a_2 = 0$ and in the $n=2$ equation the condition $\gamma = \gamma_{2,0}$ makes $a_3 = 0$, which terminates the recurrence. The resulting quadratic equation for E (from the $n=1$ equation and a_1) is

$$(2E + (N+4)\beta)(2E + N\beta) - 8\alpha = 0. \tag{2.28}$$

The solution of the quadratic equation gives the first two $l=0$ energy levels,

$$E_{0,0}(E_{1,0}) = (N+2)(-\beta/2) \mp \sqrt{\beta^2 + 2\alpha}, \tag{2.29}$$

$$-\beta = (\operatorname{sgn} b)(2a + (1 + 6/N)\sqrt{2c})^{1/2}, \quad \alpha = \sqrt{2c}, \tag{2.30}$$

For $\gamma = \gamma_{1,1}$, condition (2.25) only determines the lowest $l = 1$ energy level

$$\begin{aligned} E_{0,1} &= (N+2)(-\beta/2) \\ &= (N+2)(\operatorname{sgn} b)\tfrac{1}{2}(2a + (1 + 4/N)\sqrt{2c})^{1/2}. \end{aligned} \tag{2.31}$$

Note that we cannot compare directly this $E_{0,1}$ with $E_{0,0}$ from either (2.27) or (2.29) to determine the energy gap because they are exact solutions of *different* potentials ($b_{1,0} \neq b_{1,1} \neq b_{2,0}$). However, we can obtain a rigorous analytic upper bound on the energy gap $E_{0,1} - E_{0,0}$ for any potential that satisfies one of the $l = 0$ solvability conditions by using the exact $E_{0,0}$ and a variational upper bound for $E_{0,1}$, which is available because any $l = 1$ variational trial function for $E_{0,1}$ is strictly orthogonal to the exact $l = 0$ ground state wave function. We will see in Secs. V and VI that one feature of large N is that we can determine the energy gap analytically to leading nontrivial order.

III. LARGE *N* EXPANSION OF EXACT SOLUTIONS

Large N expansions are of considerable interest in both quantum mechanics and field theory. Usually substantial simplification occurs in the leading terms in this expansion. In Sec. II we have given some exact energy levels for general N; for these we can simply factor out the leading power in N and expand the remainder in powers of $1/N$. However, we have seen that for general N, to obtain the first energy levels (for any fixed l) requires finding the roots of an mth-order polynomial. The great simplification for large N is that the exact three-term recurrence relation (2.23) linearizes in n and can be solved to give an explicit closed formula for all $E_{k,l}$ such that $k + l \ll N/2$.

We return to (2.23) and make some additional definitions to facilitate the expansion in powers of $1/N$:

$$\begin{aligned} &2E = \epsilon, \quad \epsilon + N\beta = \epsilon', \\ &(\beta^2 - 2a)/\alpha = \gamma = 1 + (4\mu - 2)/N, \end{aligned} \tag{3.1}$$

where μ is not necessarily an integer. Then (2.23) becomes

$$\begin{aligned} &2(n+1)[1 + (2n+2l)/N]a_{n+1} \\ &\quad + [\epsilon' + (4n+2l)\beta]a_n \\ &\quad + 4\alpha(\mu - n - l/2)a_{n-1} = 0, \end{aligned} \tag{3.2}$$

which is still exact.

Now, for $N \gg 1$, we drop the $(2n + 2l)/N$ term from the coefficient of a_{n+1} (and require termination for n such that $2n + 2l \ll N$). This linearizes the recurrence relation (in n), so it can be solved by Laplace transform. This somewhat complicated calculation is relegated to Appendix A. The result is that the recurrence relation terminates after the mth term when $\mu = m + l/2$; this is just the m, l solvability condition

$$\begin{aligned} \gamma &\equiv \frac{1}{\sqrt{2c}}\left(\frac{b^2}{2c} - 2a\right) \\ &\equiv 1 + \frac{4\mu - 2}{N} = \gamma_{m,l} \equiv 1 + \frac{4m - 2 + 2l}{N}, \\ &\qquad m = 1,2,..., \end{aligned} \tag{3.3}$$

provided that ϵ' satisfies

$$\begin{aligned} &\epsilon' = -2(m - 1 + l)\beta + (4k - 2m + 2)\Delta, \\ &\quad k = 0,1,...,m-1, \end{aligned} \tag{3.4}$$

$$\Delta = \sqrt{\beta^2 + 2\alpha}. \tag{3.5}$$

Then the first m energy levels for each l and $2m + 2l \ll N$ are

$$\begin{aligned} E_{k,l} &= N(-\beta/2) + (m - 1 + l)(-\beta) \\ &\quad + (2k - m + 1)\Delta + O(1/N) \end{aligned} \tag{3.6}$$

provided that the m, l solvability condition is satisfied. This implies that β is a function of N, which must also be expanded:

$$\begin{aligned} -\beta &= -\beta_{(m,l)} \\ &= (\operatorname{sgn} b)(2a + [1 + (4m - 2 + 2l)/N]\sqrt{2c})^{1/2} \\ &= -\beta_0[1 + (2m - 1 + l)/N(\sqrt{2c}/\beta_0^2) + \cdots], \end{aligned} \tag{3.7}$$

$$-\beta_0 = (\operatorname{sgn} b)(2a + \sqrt{2c})^{1/2}. \tag{3.8}$$

Finally, this gives

$$\begin{aligned} E_{k,l} &= N\left(-\frac{\beta_0}{2}\right) + \frac{2m - 1 + l}{2}\frac{\sqrt{2c}}{(-\beta_0)} + (m - 1 + l) \\ &\quad \times(-\beta_0) + (2k - m + 1)\Delta + O(1/N), \\ &\quad k = 0,1,...,m-1 \quad (m + l \ll (N/2)), \\ &\quad \Delta = (\beta_0^2 + 2\sqrt{2c})^{1/2}. \end{aligned} \tag{3.9}$$

We emphasize that through order N^1, N^0 these are $1/N$ expansions of exact analytic solutions, i.e., the $1/N$ expansions of the $E_{0,0}$, $E_{2,0}$, and $E_{0,1}$ in Sec. II C give exactly (3.9) for the appropriate m, k, and l. For example, consider the $m = 1$, $k = 0$, and $l = 0$ exact solution (2.27). For simplicity, take $a = 0$. Then

$$E_{0,0} = \pm (2c)^{1/4}\tfrac{1}{2}N(1 + 2/N)^{1/2}, \quad \text{exact}, \tag{3.10}$$

$$= \pm (2c)^{1/4}\{\tfrac{1}{2}N + \tfrac{1}{2} + \cdots\}, \tag{3.11}$$

which is the same as (3.9) for $-\beta_0 = \pm (2c)^{1/4}$, $\Delta = |\beta_0|$, $m = 1, k = 0$, and $l = 0$. If we consider $N = 3$, then the exact E has a factor $(\frac{5}{3})^{1/2} = 1.29$ compared to the first term in the $1/N$ expansion, which gives 1 for the corresponding factor, or the first two terms in the $1/N$ expansion, which give 1.33.

IV. COMPARISON TO APPROXIMATE AND NUMERICAL CALCULATIONS, *N*=1

There is a vast literature on general properties, approximation schemes, and numerical calculations for the quantum mechanical one-dimensional anharmonic oscillator (usually quartic). There are no exact analytic solutions available for the quartic anharmonic oscillator, but most of the above-mentioned schemes and techniques can be extended to the sextic anharmonic oscillator as well; for this there are some exact analytic solutions available, as presented in Sec. II. We do not intend a systematic review and test of all the schemes discussed for the quartic oscillator. We have chosen two different representative approximation schemes to confront, with exact analytic results. Our criteria for selection were that the approximation scheme should be (i) analytic; (ii) simple, at least in lowest order of implementation; and (iii) intended for generalization to quantum field theory. Any reader who desires to confront any other scheme, or order of implementation, with the exact analytic results presented here is invited to do so. Our first choice is (a) perturbation about a Gaussian variational basis (GVP). Many variants of this approach have been proposed. We follow a recent simple analytic approach by Patnaik.[4] (The generalization of the GVP approach to field theory also has a long history. Most recently it has been urged and exploited in a series of papers by Stevenson[5].) Our second choice is (b) the finite element (FE) method of Bender *et al.*[6] (which they also applied to the field theory).

We have a choice of exact solutions from Sec. II. We will take the first two $l=0$ states for the solvability condition $\gamma=\gamma_{2,0}$ for $N=1$:

$$H=\tfrac{1}{2}p^2+ax^2+(\operatorname{sgn} b)\sqrt{2c}\,(2a+7\sqrt{2c})^{1/2}x^4+cx^6. \tag{4.1}$$

Let

$$E_{0,0}\equiv E_0,\quad E_{1,0}\equiv E_2. \tag{4.2}$$

Specializing (2.29) to $N=1$, we have

$$E_0,E_2=\tfrac{3}{2}(-\beta)\mp\sqrt{\beta^2+2\alpha}, \tag{4.3}$$

$$-\beta=(\operatorname{sgn} b)(2a+7\sqrt{2c})^{1/2},\quad \alpha=\sqrt{2c}. \tag{4.4}$$

In the weak coupling limit ($c\to 0$, $a>0$), (4.3) and (4.4) reduce to the quadratic oscillator with $-\beta=\sqrt{2a}\equiv\omega_0$ and

$$E_0=\tfrac{1}{2}\omega_0,\quad E_2=\tfrac{5}{2}\omega_0\quad \text{(weak coupling limit)}. \tag{4.5}$$

In the strong coupling regime, $\sqrt{2c}\gg 2a$. In order to have simple results depending on only one parameter we set $a=0$. Then we have, for $\operatorname{sgn} b=+1$, the exact results

$$\begin{aligned}
&E_0=(\tfrac{3}{2}\sqrt{7}-3)(2c)^{1/4}=(0.968\,627\ldots)(2c)^{1/4},\\
&E_2=(\tfrac{3}{2}\sqrt{7}+3)(2c)^{1/4}=(6.968\,627\ldots)(2c)^{1/4},\\
&E_2-E_0=6(2c)^{1/4},
\end{aligned} \tag{4.6}$$

and for $\operatorname{sgn} b=-1$,

$$\begin{aligned}
&E_0=(-\tfrac{3}{2}\sqrt{7}-3)(2c)^{1/4},\\
&E_2=(-\tfrac{3}{2}\sqrt{7}+3)(2c)^{1/4}.
\end{aligned} \tag{4.7}$$

For $a=0$ and $b<0$ the potential is a double well, minimum value

$$\begin{aligned}
V_0&=-(112\sqrt{7}/27)(2c)^{1/4},\\
&\text{at } x_0^2=(4\sqrt{7}/3)(2c)^{-1/4}.
\end{aligned} \tag{4.8}$$

Measuring from the bottom of the (double) well ($\hat{E}=E-V_0$)

$$\hat{E}_0=(4.006\,341\ldots)(2c)^{1/4},\quad \hat{E}_2-\hat{E}_0=6(2c)^{1/4}. \tag{4.9}$$

(a) In the simplest version[4] of the combined GVP approach, one starts with a Gaussian trial function for the ground state and makes a Bogoliubov transformation from the simple harmonic oscillator basis to a new basis with the minimized Gaussian as ground state. The exact Hamiltonian is reexpressed in this new basis. The c-number part is the Gaussian variational minimum for the ground state energy. The nontrivial diagonal part gives the zeroth approximation to the excited state energies and the off diagonal part is treated perturbatively (starting in second order; there is no first-order energy shift). The details for the quartic potential are found in Ref. 4. The extension to the sextic potential is given here as

$$\phi_0=Ne^{-(1/2)M(x-x_0)^2}, \tag{4.10}$$

$$H_{\text{eff}}=H_d+H_1, \tag{4.11}$$

$$\begin{aligned}
H_d={}&\underline{W_0}+\underline{M}B^+B\\
&+\left(\frac{3}{2}\frac{b}{\underline{M}^2}+\frac{45}{4}\frac{c}{\underline{M}^3}+\frac{45}{2}\frac{c}{M^2}x_{\underline{0}}^2\right)B^{+2}B^2\\
&+\frac{5}{2}\frac{c}{\underline{M}^3}B^{+3}B^3,
\end{aligned} \tag{4.12}$$

$$\begin{aligned}
W_0(M,x_0)={}&\langle\phi_0|H|\phi_0\rangle\\
={}&\frac{M}{4}+\frac{a}{2M}+\frac{3}{4}\frac{b}{M^2}+\frac{15}{8}\frac{c}{M^3}\\
&+\left(a+3\frac{b}{M}+\frac{45}{4}\frac{c}{M^2}\right)x_0^2\\
&+\left(b+\frac{15}{2}\frac{c}{M}\right)x_0^4+cx_0^6,
\end{aligned} \tag{4.13}$$

$$\underline{W_0}=\min_{(M,x_0)}W_0(M,x_0)=W_0(\underline{M},\underline{x_0}), \tag{4.14}$$

and in zeroth approximation (there is no first-order energy shift)

$$\begin{aligned}
&E_0=\underline{W_0},\quad E_1-E_0=\underline{M},\\
&E_2-E_0=2\underline{M}+3\frac{b}{\underline{M^2}}+\frac{45}{2}\frac{c}{\underline{M^3}}+45\frac{c}{\underline{M^2}}x_{\underline{0}}^2.
\end{aligned} \tag{4.15}$$

In the weak coupling limit (b, $c=0$, $a>0$) (4.10)–(4.15) reduce exactly to the harmonic oscillator. The results for the nontrivial choice of potential parameters ($a=0$, $b=\pm\sqrt{7}(2c)^{3/4}$, $c>0$) are given in Table I.

(b) In the FE scheme[6] the starting point is again a variationally improved Gaussian basis and the scheme then provides a set of difference equations which lead to a set of poly-

TABLE I. Low-lying energy levels, $N=1$, strong coupling.

		Exact	GVP	FE
	ϵ_0[b]	$\frac{3}{2}\sqrt{7}-3=0.968\,627$	0.995	···
V_+[a]	$\epsilon_2-\epsilon_0$	6	7.069	6.036
	$(\epsilon_1-\epsilon_0)$	(2.537)[c] <2.599[d]	2.715	2.715
	$\hat{\epsilon}_0$[e]	$-\frac{3}{2}\sqrt{7}-3+\frac{112}{27}\sqrt{7}$ $=4.006\,341$	4.116	···
V_-[a]	$\epsilon_2-\epsilon_0$	6	16.99	1.891
	$(\epsilon_1-\epsilon_0)$	(0.0003)[c] <4.341	8.0	0.694

[a] $V_\pm=\pm\sqrt{7}(2c)^{3/4}x^4+\frac{1}{2}2cx^6$.
[b] $\epsilon=E/(2c)^{1/4}$.
[c] From numerical integration of the Schrödinger equation.
[d] Rigorous bound $\underline{W_1}-E_0$, where $\underline{W_1}$ is a variational upper bound on the lowest odd parity level and E_0 is exact.
[e] $\hat{E}_0=E_0-V_0$, where V_0 is the minimum (negative) of the double well V_-.

nomial equations to solve for energy differences $\omega_n = E_{n+1} - E_n$. In the lowest approximation (one finite element), the scheme gives

$$\omega_1 = \underline{M}. \tag{4.16}$$

Here $\underline{M}$ is the variational parameter that minimizes the Gaussian ground state energy w_0, with x_0 constrained to be zero. The second energy difference ω_2 is the solution of

$$\omega_2^2 = 2a + 10b(1/\omega_2) + \tfrac{135}{2}c(1/\omega_2^2) \tag{4.17}$$

and

$$E_2 - E_0 = \omega_2 + \omega_1. \tag{4.18}$$

In the weak coupling limit (b, $c = 0$, $2a = \omega_0^2$), the FE scheme also reduces correctly to the harmonic oscillator — all $\omega_k = \omega_0$. The results for the nontrivial choice of anharmonic potential parameters are given in Table I.

We observe that the Gaussian variational calculation of the ground state energy is quite accurate: It is exact in the weak coupling limit and in the strong coupling regime has the correct ($C^{1/4}$) dependence on the coupling constant and is accurate numerically to within 3% for both the single well (V_+) and double well (V_-). However, we note that the accurate result for the double well required the use of the two-parameter (M,x_0) "shifted" Gaussian trial function (4.10). With x_0 constrained to be zero, the Gaussian variational calculation for V_- gives $\hat{\epsilon}_0 = 9.84$, in error by more than 100%. For the single well V_+, both GVP and FE, even in the lowest order of implementation, give reasonable approximations to the first two excitation energies. (However, GVP in lowest order is 15% off on $E_2 - E_0$, which is a substantially greater discrepancy then the corresponding result for the quartic anharmonic oscillator[4].) Higher order implementation presumably leads with sufficient effort to any desired accuracy.[4,7,8] For the double well, although the (shifted) Gaussian variation gives a numerically accurate value for the ground state energy, both GVP and FE in lowest order are far off the excitation energies. In particular, we comment on the energy gap $E_1 - E_0$, which is estimated in both GVP and FE by the value of the parameter M (inverse width) of the Gaussian trial function which minimizes W_0. For the convex potential V_+, this gives a reasonable numerical value (off by 4%), but even in this case this estimate violates the simple analytic upper bound. For the double well V_-, the estimation of $E_1 - E_0$ by $\underline{M}$ is totally inaccurate.

(i) The GVP and FE results are different because the $\underline{M}$ used by GVP is the $\underline{M}$ ($x_{\underline{0}}$) which provides the accurate $W_{\underline{0}}$ with $x_{\underline{0}} \neq 0$, while the equation to be solved for ω_1 in FE is identical to the equation for $\underline{M}$ which minimizes W_0 with $x_0 = 0$.

(ii) The striking result that E_1 is nearly degenerate with E_0 in the double-well potential is understood by observing that the odd parity wave function with its single node at $x = 0$ can fit itself more favorably to the negative regions of V_- than can the nodeless even parity wave function.

(iii) We will see in Sec. V that for large N, the estimate $E_1 - E_0 = \underline{M}$ is accurate up to order $1/N$ for all ranges of parameters in the potential.

V. GAUSSIAN APPROXIMATION FOR LARGE *N*

The expectation value of the Hamiltonian $\mathbf{p}^2/2 + V$ with

$$V = a\mathbf{x}^2 + (b/N)(\mathbf{x}^2)^2 + (c/N^2)(\mathbf{x}^2)^3 \tag{5.1}$$

in the (shifted) Gaussian trial function

$$\Phi = (M/\pi)^{N/4} e^{-(1/2)M(\mathbf{x}-\mathbf{x}_0)^2} \tag{5.2}$$

is

$$W_0(M,\mathbf{x}_0^2) = N\left\{\frac{M}{4} + a\left(\frac{1}{2M} + \xi^2\right) + b\left[\left(1+\frac{2}{N}\right)\frac{1}{4M^2} + \left(1+\frac{2}{N}\right)\frac{1}{M}\xi^2 + (\xi^2)^2\right]\right.$$
$$\left. + c\left[\left(1+\frac{6}{N}+\frac{8}{N^2}\right)\frac{1}{8M^3} + \left(1+\frac{6}{N}+\frac{8}{N^2}\right)\frac{3}{4M^2}\xi^2 + \left(1+\frac{4}{N}\right)\frac{3}{2M}(\xi^2)^2 + (\xi^2)^3\right]\right\}, \tag{5.3}$$

$\xi^2 = (1/N)\mathbf{x}_0^2$.

To leading order in N, the minimization conditions are

$$0 = \frac{1}{N}\frac{\partial W_0}{\partial M} = \frac{1}{4} - \frac{a}{2M^2} - \frac{b}{2M^3} - \frac{3}{8}\frac{c}{M^4} + \left(-\frac{b}{M^2} - \frac{3}{2}\frac{c}{M^3}\right)\xi^2 - \frac{3}{2}\frac{c}{M^2}(\xi^2)^2, \tag{5.4}$$

$$0 = \frac{1}{N}\frac{\partial W_0}{\partial \xi^2} = a + \frac{b}{M} + \frac{3}{4}\frac{c}{M^2} + \left(2b + 3\frac{c}{M}\right)\xi^2 + 3c(\xi^2)^2 = -2M^2\frac{1}{N}\frac{\partial W_0}{\partial M} + \frac{M^2}{2},$$

which cannot be satisfied simultaneously for $M > 0$ (required for a normalizable Φ). However, $W_0(m,\mathbf{x}_0^2)$ is bounded below for positive c, so it must have its lowest value at the end point value $\mathbf{x}_0^2 = 0$. (This is not the case for finite N, as we saw explicitly in the double-well example for $N = 1$ in Sec. IV.) Thus in the large N limit, the Gaussian minimization condition is

$$M^4 - 2aM^2 - 2bM - \tfrac{3}{2}C = 0, \tag{5.5}$$

which holds with no restriction on a, b, and c (except $c > 0$). We want to compare with the exact analytic solutions that exist when the parameters of the potential satisfy one of the sequence of solvability constraints (3.3). In the large N limit, for $2m + l - 1 \ll N/2$ these constraints all degenerate to the single condition

$$b = (\operatorname{sgn} b)\sqrt{2c}(2a + \sqrt{2c})^{1/2} + O(1/N). \tag{5.6}$$

In this case one can check by direct substitution that

$$\underline{M} = \tfrac{1}{2}(-\beta_0 + \Delta), \tag{5.7}$$

$$-\beta_0 = (\operatorname{sgn} b)(2a + \sqrt{2c})^{1/2}, \quad \Delta = (\beta_0^2 + 2\sqrt{2c})^{1/2} \tag{5.8}$$

is a solution of 5.5, which leads to the minimum value

$$\underline{W}_0 = N(-\tfrac{1}{2}\beta_0) + O(1). \tag{5.9}$$

Comparing with the $1/N$ expansion of the exact solutions (3.9), we see that the Gaussian variation gives the ground state energy exactly to leading order in N.

Consider now the energy gap $E_{0,1} - E_{0,0}$. From (3.9) this is of order 1, i.e., $(E_1 - E_0)/E_0 \sim 1/N$. However, the solvability conditions (3.3) also differ in order $1/N$, so again one cannot determine $E_{0,1} - E_{0,0}$ in a common potential directly from (3.9). The m, l solvability condition is (3.7):

$$b = (\operatorname{sgn} b)\sqrt{2c}(2a + [1 + (4m - 2 + 2l)/N]\sqrt{2c})^{1/2}. \tag{5.10}$$

In the large N limit there is an important simplification; the condition becomes linear in m and l:

$$b_{m,l} = \sqrt{2c}(-\beta_0)(1 + [(2m - 1 + l)/N] \times (\sqrt{2c}/\beta_0^2) + \cdots), \tag{5.11}$$

which suggests that we can obtain $E_{0,1}$ at $b_{m,0}$ by linear interpolation between the known exact $E_{0,1}$ at $b_{m-1,1}$ and $b_{m,1}$ $[b_{m,0} = \tfrac{1}{2}(b_{m-1,1} + b_{m,1}) + O(1/N)]$. Making this interpolation in (3.9) gives

$$E_{0,1} - E_{0,0} = \tfrac{1}{2}(-\beta_0 + \Delta) + O(1/N), \tag{5.12}$$

which is precisely $\underline{M}$ (5.7) to this order. Thus if this linear interpolation is correct for large N, then the Gaussian variational calculation also obtains the energy gap correctly to leading order, N^0 in this case, by the value of the single variational parameter which minimizes W_0 in order N. As we have seen in Sec. IV this is not the case for finite N; however, in Sec. VI we will provide a calculation which is a $1/N$ expansion from the start, requires no condition on the parameters a, b, and c of the potential, and gives (5.12).

VI. THE GENERAL 1/N EXPANSION TESTED ON NONTRIVIAL SOLVABLE EXAMPLES

The derivation of the unrestricted $1/N$ expansion given here is adapted from Mlodinow.[9] The starting point is the radial equation for the N-dimensional Schrödinger equation in the form (2.16), which we repeat here as

$$-\frac{1}{2N^2}\tilde{\chi}'' + \left[\frac{(N-1)(N-3)}{8N^2\rho^2} + \frac{l(l+N-2)}{2N^2\rho^2} + v(\rho)\right]\tilde{\chi} = e\tilde{\chi}, \quad e = (E/N), \tag{6.1}$$

$$(N-1)(N-3)/8N^2\rho^2 + l(l+N-2)/2N^2\rho^2 + v(\rho) \equiv v_{\text{eff}}(\rho). \tag{6.2}$$

The heuristic observation that is the basis of the derivation is that N^2 acts as an effective mass so that in the $N \to \infty$ limit the "particle" will sit at the classical minimum of v_{eff}. As N decreases from infinity one expects $\tilde{\chi}$ to relax to a Gaussian centered about ρ_0. Thus we write

$$v_{\text{eff}}(\rho) = v_{\text{eff}}(\rho_0) + 0 + \tfrac{1}{2}(\rho - \rho_0)^2 v''_{\text{eff}}(\rho_0) + \cdots. \tag{6.3}$$

Each $v_{\text{eff}}^{(n)}$ has a $1/N$ expansion, starting in order N^0, and for large N $\langle(\rho - \rho_0)^2\rangle \sim N^{-1}$ for the Gaussian $\exp[-N(\rho - \rho_0)^2]$. Thus one has

$$E = N v_{\text{eff}}(\rho_0) + (k + \tfrac{1}{2})\sqrt{v''_{\text{eff}}(\rho_0)} + O(1/N), \quad k = 0,1,2,\ldots. \tag{6.4}$$

For a general sextic potential $v(\rho) = a\rho^2 + b\rho^4 + c\rho^6$, the equation to determine ρ_0 is a quartic (in ρ^2) that cannot be solved analytically in general. Therefore we give up some generality in parameters and take $a = 0$ and b unrestricted, but written as

$$b = (\operatorname{sgn} b)(1 + 2s/N)^{1/2}(2c)^{3/4} = (\operatorname{sgn} b)(1 + s/N + \cdots)(2c)^{3/4} \quad (s \geqslant 0), \tag{6.5}$$

and we rescale

$$\rho = (u^4/2c)^{1/8}. \tag{6.6}$$

Then

$$v(\rho) = (2c)^{1/4}\{(\tfrac{1}{8} + (l-1)/2N + \cdots)(1/u) \pm (1 + s/N + \cdots)u^2 + \tfrac{1}{2}u^3\}. \tag{6.7}$$

Minimization of (6.7) leads to the quartic equation

$$u^4 \pm \tfrac{4}{3}u^3 - \tfrac{1}{12} = 0, \tag{6.8}$$

with the solution

$$u_0^{(\pm)} = (\sqrt{3} \mp 1)/2. \tag{6.9}$$

After some algebra Eqs. (6.4), (6.7), and (6.9) yield

$$E_{k,l}^{(+)} = (2c)^{1/4}\{N\tfrac{1}{2} + [s(1 - \sqrt{3}/2) + (l-1)(\tfrac{1}{2} + \sqrt{3}/2) + (k + \tfrac{1}{2})2\sqrt{3}] + O(1/N)\}, \tag{6.10a}$$

$$E_{k,l}^{(-)} = (2c)^{1/4}\{N(-\tfrac{1}{2}) + [-s(1 + \sqrt{3}/2) + (l-1)(-\tfrac{1}{2} + \sqrt{3}/2) + (k + \tfrac{1}{2})2\sqrt{3}] + O(1/N)\}. \tag{6.10b}$$

The $1/N$ expansions, (6.10a) and (6.10b), are derived for unrestricted $s \geqslant 0$. If we consider the special values $s = 2m - 1 + l$, then (6.5) is just the m, l solvability condition [Eqs. (3.7) and (5.10) for $a = 0$]. Substituting these values of s into (6.10) and setting $a = 0$ in the $1/N$ expansion of the exact analytic solutions (3.9), we find that both reduce to the common result

$$E_{k,l}^{(\pm)} = (2c)^{1/4}\{\pm N\tfrac{1}{2} + [\pm 2m + \tfrac{3}{2}(l-1) + (2k - m + 1)\sqrt{3}] + O(1/N)\}, \quad \text{for } a = 0,$$

$$b = \pm(1 + (4m - 2 + 2l)/N)^{1/2}(2c)^{3/4}, \tag{6.11}$$

thus verifying that the first two terms of the general $1/N$ expansion agree with the first two terms of the $1/N$ expansion of nontrivial exact analytic solutions for values of the parameters of the potential for which the exact analytic solutions are available. For clarity we repeat the point in a very specific example. Take the particular sextic anharmonic potential in arbitrary N dimensions:

$$V_{\pm} = \pm (1+2/N)^{1/2}(2c)^{3/4}(\mathbf{x}^2)^2 + c(\mathbf{x}^2)^3. \quad (6.12)$$

The general $1/N$ expansion, Eqs. (6.4)–(6.10), yields ($s=1$, $m=1$, $l=0$, and $k=0$)

$$E^{(\pm)}_{0,0} = \pm (2c)^{1/4}\{N\tfrac{1}{2} + \tfrac{1}{2} + O(1/N)\}, \quad (6.13)$$

while for this potential the exact ground state for arbitrary N is given by Eq. (2.27) for $a=0$:

$$E^{(\pm)}_{0,0} = \pm (2c)^{1/4}\tfrac{1}{2}N(1+2/N)^{1/2}. \quad (6.14)$$

On the other hand, having verified the correctness of the general $1/N$ expansion (6.10), which has no l-dependent restriction on the potential parameter b, we can immediately read off the leading term of the energy gap

$$E^{(\pm)}_{0,1} - E^{(\pm)}_{0,0} = (2c)^{1/4}\{\pm\tfrac{1}{2} + \sqrt{3}/2 + O(1/N)\}$$
$$= \tfrac{1}{2}(-\beta_0 + \Delta) + O(1/N) \quad (a=0), \quad (6.15)$$

which justifies the linear interpolation (5.12).

VII. SUMMARY AND CONCLUSIONS

We have exploited the availability of exact analytic solutions for low-lying energy levels of the one-dimensional sextic anharmonic oscillator $V = ax^2 + bx^4 + cx^6$ to confront some approximation schemes developed for the quartic anharmonic oscillator, for which no exact analytic solutions are available. Not surprisingly, the various lowest order approximations work well for a single well ($a,b \geqslant 0$), but are quite inaccurate (or slowly converging) for strong double-well potentials ($a=0$, $b<0$). We have shown that a large class of exact analytic solutions can also be obtained for the N-dimensional $O(N)$ symmetric generalization $V = a\mathbf{x}^2 + (b/N)(\mathbf{x}^2)^2 + (c/N^2)(\mathbf{x}^2)^3$, for any N. We have found that substantial simplifications occur for $N \gg 1$. For finite N (including $N=1$) to obtain the first m-energy levels (for fixed l), even in a "solvable" case, requires determination of the roots of an mth-order polynomial; however, for $N \gg 1$ we have given explicit closed form analytic expressions for $E_{k,l}$ for $k + l \ll N$, including terms of order N and order N^0, which are expansions of exact solutions. We have shown that general $1/N$ expansions (which are not guaranteed to be $1/N$ expansions of exact results in arbitrary N dimensions) do agree with these exact results. We have also shown by explicit calculation and comparison with exact results that the ground state energy $E_{0,0}$ is given exactly to leading order (N) by the Gaussian variational estimate with an unshifted Gaussian ($\mathbf{x}_0^2 = 0$ even for $a,b<0$) and the "energy gap" $E_{0,1} - E_{0,0}$ is given to leading order (N^0) by the value of the Gaussian variation parameter M which minimizes W_0.

As mentioned in the Introduction, our ultimate interest in these results is in their possible extension to the field theory problem. Two substantial difficulties are apparent immediately: (i) in going from $d = 0+1$ to $d = s+1$ spatial gradients in the s variables appear which have no counterpart in the $d = 0+1$ quantum mechanical problem; and (ii) ordinary integrals become functional integrals and ultraviolet divergences appear which require (infinite) renormalization of the parameters of the potential. Preliminary investigation indicates that these difficulties undermine the usefulness of exact results for the ground state energy in the quantum mechanical problem, but the question of the possibility of rigorous bounds on the energy gap is not yet settled.

APPENDIX A: LAPLACE TRANSFORM SOLUTION OF THE RECURRENCE RELATION FOR LARGE N

We have rescaled variables (2.14) and (2.22) so that in the large N limit all the a_n are of order N^0 and satisfy the linearized recurrence relation [Eq. (3.2) with the $1/N$ term dropped]

$$2(n+1)a_{n+1} + [\epsilon' + (4n+2l)\beta]a_n + 4\alpha(\mu - n - l/2)a_{n-1} = 0. \quad (A1)$$

The mth solvability condition is $\gamma = \gamma_{m,l}$ [(2.25)], or $\mu = m + l/2$. [See (3.1).]

We rewrite (A1) in a standard form:

$$(\alpha_0 + \beta_0 n)y_n + (\alpha_1 + \beta_1(n-1))y_{n-1} + (\alpha_2 + \beta_2(n-2))y_{n-2} = 0, \quad (A2)$$

where

$$\alpha_0 = 0, \quad \beta_0 = 1, \quad \alpha_1 = \epsilon'/2 + l\beta, \quad \beta_1 = -2\beta,$$
$$\alpha_2 = 2\alpha(\mu - 1 - l/2), \quad \beta_2 = -2\alpha. \quad (A3)$$

Substitute the Laplace transform

$$y_n = \int_\Gamma dz\, z^{n-1}v(z) \quad (A4)$$

into (A2) and integrate by parts to obtain

$$0 = \int_\Gamma dz\{[\alpha_0 z^{n-1} + \alpha_1 z^{n-2} + \alpha_2 z^{n-3}]v(z) - [\beta_0 z^n + \beta_1 z^{n-1} + \beta_2 z^{n-2}]v'(z)\}, \quad (A5)$$

provided the contour Γ is chosen such that

$$v(z)(\beta_0 z^n + \beta_1 z^{n-1} + \beta_2 z^{n-2})|_\Gamma = 0. \quad (A6)$$

Then the transform $v(z)$ satisfies the differential equation

$$\frac{v'(z)}{v(z)} = \frac{\alpha_0 z^{n-1} + \alpha_1 z^{n-2} + \alpha_2 z^{n-3}}{\beta_0 z^n + \beta_1 z^{n-1} + \beta_2 z^{n-2}} = \frac{\alpha_1 z + \alpha_2}{z(z^2 + \beta_1 z + \beta_2)} \quad (A7)$$

by the first two equations of (A3). Let

$$\frac{\alpha_1 z + \alpha_2}{z(z^2 + \beta_1 z + \beta_2)} = \frac{A}{z} + \frac{B}{z - z_1} + \frac{C}{z - z_2}. \quad (A8)$$

The roots of the quadratic expression in the denominator are

$$z_1 = -\beta_1/2 + \Delta, \quad z_2 = -\beta_1/2 - \Delta, \quad (A9)$$

$$\beta_1 = 2\beta, \quad \Delta = \sqrt{\beta^2 + 2\alpha}. \quad (A10)$$

Then some algebra leads to

$$A = \frac{\alpha_2}{z_1 z_2} = -\left(\mu - 1 - \frac{l}{2}\right),$$
$$B = \frac{\alpha_2 + \alpha_1 z_1}{z_1(z_1 - z_2)} = \frac{1}{2\Delta}\left(\alpha_1 - \left(\mu - 1 - \frac{l}{2}\right)z_2\right),$$
$$C = -\frac{\alpha_2 + \alpha_1 z_1}{z_2(z_1 - z_2)} = \frac{1}{2\Delta}\left(-\alpha_1 + \left(\mu - 1 - \frac{l}{2}\right)z_1\right). \quad (A11)$$

The solution of the differential equations (A7) and (A8) is

$$v(z)=v_0 z^A (z-z_1)^B (z-z_2)^C. \tag{A12}$$

By virtue of (A9), a contour Γ that satisfies the condition (A6) is a semicircle in the upper half-plane, centered at $-\beta$ on the real axis and intersecting the real axis at z_1 and z_2 (provided that $B,C>-1$). On Γ,

$$z=-\beta+\Delta e^{i\theta}, \quad 0\leqslant\theta\leqslant\pi. \tag{A13}$$

On this contour y_m is complex. Since the coefficients (A3) of (A2) are all real, the real and imaginary parts of y_m must satisfy (A2) separately. This will introduce two possible solutions: $a_{m+1}\alpha\,\mathrm{Re}\,y_m$ and $\mathrm{Im}\,y_m$. We will have to check which is the required solution.

Substitute (A3) and (A9)–(A13) into (A4) and change the integration variable $\theta=2\phi$ to obtain

$$y_n=v_0(2\Delta)^{B+C+1} i^{B+1}\int_0^{\pi/2} d\phi(-\beta+\Delta e^{2i\phi})^{A+n-1} \times e^{i(B+C+2)\phi}(\cos\phi)^C(\sin\phi)^B, \tag{A14}$$

$$\begin{aligned} A+n-1&=-\mu+n+l/2,\\ B+C&=\mu-1-l/2. \end{aligned} \tag{A15}$$

We want to terminate the recurrence by imposing the solvability condition $\mu=m+l/2$ to force $y_m=0$. For $n=m$ and $\mu=m+l/2$, Eq. (A14) simplifies to

$$\begin{aligned} y_m&=v_0(2\Delta)^m i^{B+1}\int_0^{\pi/2} d\phi(\cos(m+1)\phi+i\sin(m+1)\phi)(\cos\phi)^C(\sin\phi)^B\\ &=v_0(2\Delta)^m i^{B+1}\left(\cos\frac{(B+1)\pi}{2}+i\sin\frac{(B+1)\pi}{2}\right)\frac{\Gamma(C+1)\Gamma(B+1)}{\Gamma(B+C+2)}\\ &=v_0\frac{(2\Delta)^m}{m!}(\cos(B+1)\pi+i\sin(B+1)\pi)(m-1-B)(m-2-B)\cdots(1-B)\frac{\pi B}{\sin\pi B}. \end{aligned} \tag{A16}$$

Taking the solution $a_{m+1}\propto \mathrm{Im}\,y_m$, we obtain $\mathrm{Im}\,y_m=0$ by having B take on one of the m values $0,1,\ldots,m-1$. [Recall that (A6) requires $B>-1$.] Furthermore, starting again from (A14) with $n=m+r$, but still with $u=m+l/2$ so that $A+n-1=r$ and $B+C=m+1$, we can arrive at the factored result

$$\begin{aligned} y_{m+r}&\doteq y_m\Delta^r\left\{\sum_{p=0}^{r}\sum_{g=0}^{r-p}\binom{r}{p}\binom{r-p}{q}\left(\frac{-\beta}{\Delta}\right)^p(-1)^q\frac{\Gamma(C+1+r-p-q)}{\Gamma(C+1)}\frac{\Gamma(B+1+q)}{\Gamma(B+1)}\frac{\Gamma(m+1)}{\Gamma(m+1+r-p)}\right\}\\ &=y_m\times\text{real constant}. \end{aligned} \tag{A17}$$

Thus the condition $\mathrm{Im}\,y_m=0$ forces all subsequent $\mathrm{Im}\,y_{m+r}$ also to zero. Then (A3), (A9), and (A11) give, for $\mu=m+l/2$,

$$\begin{gathered} \epsilon'/2=(m-1+l)(-\beta)+(2k-m+1)\Delta,\\ k=0,1,\ldots,m-1, \end{gathered} \tag{A18}$$

which is (3.4).

APPENDIX B: GENERALIZATION TO *N* DIMENSIONS OF THE PERTURBATION ABOUT A GAUSSIAN VARIATIONAL BASIS

In Sec. IV we outlined briefly the extension of the treatment of the one-dimensional quartic anharmonic oscillator of Ref. 4 to the one-dimensional sextic anharmonic oscillator. In this Appendix we give the results of the extension to an N-dimensional [$O(N)$ invariant] sextic anharmonic oscillator. The simultaneous generalization from one to N dimensions and from an unshifted to shifted Gaussian variational function is complicated; thus we restrict ourselves to the generalization from one to N dimensions of the scheme of Patnaik[4] based on an unshifted Gaussian ($\mathbf{x}_0=0$). The results are

$$H=\frac{\mathbf{p}^2}{2}+a\mathbf{x}^2+\frac{b}{N}(\mathbf{x}^2)^2+\frac{c}{N^2}(\mathbf{x}^2)^3, \tag{B1}$$

$$\phi_0=e^{-(1/2)M\mathbf{x}^2}, \tag{B2}$$

$$W_0(M)=N\left\{\frac{1}{4}M+\frac{a}{2M}+\left(1+\frac{2}{N}\right)\frac{b}{4M^2}+\left(1+\frac{6}{N}+\frac{8}{N^2}\right)\frac{c}{8M^3}\right\}, \tag{B3}$$

$$\begin{aligned} H_{\text{eff}}&=\underline{W_0}+\underline{M}\,\mathbf{B}\cdot\mathbf{B}+\frac{b}{N}\frac{1}{4\underline{M}^2}:((\mathbf{B}^+ +\mathbf{B})^2)^2:\\ &\quad+\frac{c}{N^2}\frac{1}{8\underline{M}^3}[\,:((\mathbf{B}^+ +\mathbf{B})^2)^3:\\ &\quad+3(N+4):((\mathbf{B}^+ +\mathbf{B})^2)^2:]=H_d+H_1. \end{aligned} \tag{B4}$$

In lowest approximation ($E_0\equiv E_{0,0}$, $E_1\equiv E_{0,1}$, $E_2\equiv E_{1,0}$),

$$\begin{aligned} &E_0=\underline{W_0},\quad E_1-E_0=\underline{M},\\ &E_2-E_0=2\underline{M}+\left(1+\frac{2}{N}\right)\frac{b}{\underline{M}^2}+\left(1+\frac{6}{N}+\frac{8}{N^2}\right)\frac{3c}{2\underline{M}^3}. \end{aligned} \tag{B5}$$

In the weak coupling limit ($b,c=0$) (B1)–(B5) reduce correctly to the results for N independent harmonic oscillators (one in each dimension). For nonzero anharmonic couplings for finite N, we anticipate results qualitatively the same as in Sec. IV ($N=1$), i.e., accurate numerically for single-well potentials, but not for double-well potentials (even for the ground state, with $\mathbf{x}_0$ restricted to be zero). For large N, the Gaussian estimate of the ground state energy is minimized for $\mathbf{x}_0 = 0$, so we can reasonably compare (B5) for large N to the known large N results. We have seen in Sec. V that $E_0 = \underline{W_0}$ $[=O(N)]$ and $E_1 - E_0 = \underline{M}$ $[=O(1)]$ are correct in the large N limit. We check $E_2 - E_0$ for the potential parameter $a=0$ and $b = \pm (2c)^{3/4} + O(1/N)$. From (5.7), for large N,

$$\underline{M} = \tfrac{1}{2}(-\beta_0 + \Delta) = \tfrac{1}{2}(\pm 1 + \sqrt{3})(2c)^{1/4}, \quad a=0, \tag{B6}$$

which, substituted into (B5), gives

$$E_2 - E_0 = \begin{cases} (8\sqrt{3} - 10)(2c)^{1/4} & (b>0), \\ (8\sqrt{3}+10)(2c)^{1/4} & (b<0), \end{cases} \tag{B7}$$

which is nothing like the exact result $2\sqrt{3}\,(2c)^{1/4}$ for either sign of b from (3.9). This suggests that the perturbation expansion of Ref. 4, extended to (B4), is failing for large N. In fact, there are terms in every order of the perturbation expansion of (B4) which are the same order in $1/N$, namely $O(N^0)$, the same as (B7).

APPENDIX C: RELATION TO SUPERSYMMETRIC QUANTUM MECHANICS AND EXACTLY SOLVABLE POTENTIALS

In a brief but remarkable paper,[10] Gendenshtein[11] stated that for all known cases of potentials for which the Schrödinger equation is exactly solvable (admits a closed form solution for all the energy levels), one can formulate the problem in the supersymmetric quantum mechanics formalism of Witten[12] and then by elementary algebraic means determine the entire spectrum.

In particular, Gendenshtein[11] shows that if the potential can be written as

$$V(x;a) = \tfrac{1}{2}(W^2(x;a) - W'(x;a)) \equiv V_-(x;a), \tag{C1}$$

where a is a (set of) parameter(s) of the potential and if

$$\begin{aligned} V_+(x;a) &\equiv \tfrac{1}{2}(W^2(x;a) + W'(x;a)) \\ &= V_-(x;a_1) + R(a_1), \end{aligned} \tag{C2}$$

where $a_1 = f(a)$ for some function f, then

$$E_0 = 0, \quad E_n = \sum_{k=1}^{n} R(a_k), \tag{C3}$$

where a_k is the iterated value $a_k = f^{(k)}(a)$.

Gendenshtein[11] also pointed out that in order to find the exact ground state energy, condition (C1) alone is sufficient. This observation is based on the same ansatz used in Sec. II. Simply substitute

$$\psi_0(x) = \exp\left(-\int^x W(x')dx'\right) \tag{C4}$$

and (C1) into the Schrödinger equation to find the exact result $E_0 = 0$. As an example, Gendenshtein[11] considered potentials of the form

$$V = ax^2 + cx^6 \quad (b=0). \tag{C5}$$

To satisfy (C1) requires a one-parameter family $a = -A$, $c = 2A^2/9$, which is precisely the first even parity ($l=0$) solvability condition: $\gamma = \gamma_{1,0} = 3$, $b = \pm\sqrt{2c}(2a + \gamma\sqrt{2c})^{1/2} = 0$, with exact $E_0 = -\beta/2 = b/2\sqrt{2c} = 0$. For these values of a, c this potential also satisfies (C2) with $a_1 = f(a) = -a$, but trivial $R(a_1) = 0$; thus one cannot turn the supersymmetry crank (C3) to generate exact higher energy levels.

In fact, we know from Sec. II that there is a whole sequence of potentials of form (C5) corresponding to the mth $l=0$ solvability condition $\gamma = \gamma_{m,0} = 4m-1$, $a = -A$, and $c = 2A^2/\gamma^2$, for which we can find the exact E_0 by finding the zeros of an $m \times m$ determinant. These cases do not fit the supersymmetry mold; V does not satisfy (C1), ψ_0 is not of the form (C4), and $E_0 \neq 0$. [See, e.g., Eqs. (2.28)–(2.30) and the text.]

[1] We thank Dr. H. Neuberger for pushing us to this observation.

[2] P. M. Morse and E. C. G. Stuekelberg, Helv. Phys. Acta **4**, 337 (1931); V. Singh, S. N. Biswas, and K. Datta, Phys. Rev. D **18**, 1901 (1978).

[3] *Higher Transcendental Functions, Vol. II*, edited by A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi (McGraw-Hill, New York, 1953).

[4] P. K. Patnaik, Phys. Rev. D **33**, 3145 (1986). For implementation with the shifted Gaussian (4.10) see, also, C. Esebbag, J. Nunez, A. Plastino, and G. Bozzolo, Phys. Rev. D **32**, 522 (1985).

[5] P. M. Stevenson, Phys. Rev. D **32**, 1389 (1985).

[6] A concise description along with references to earlier papers is contained in C. M. Bender and K. A. Milton, Phys. Rev. D **34**, 3149 (1986).

[7] C. M. Bender, K. A. Milton, D. H. Sharp, L. M. Simmons, and R. Strong, Phys. Rev. D **32**, 1476 (1985).

[8] C. M. Bender and M. L. Green, Phys. Rev. D **34**, 3255 (1986).

[9] L. D. Mlodinow, International School of Nuclear Physics, Erice V, 1981.

[10] We thank Dr. H. Neuberger for calling this subject to our attention.

[11] L. E. Gendenshtein, JETP Lett. **38**, 299 (1983).

[12] E. Witten, Nucl. Phys. B **185**, 513 (1981).

Nonlinear "self-interaction" Hamiltonians of the form $H^{(0)}+\lambda\langle r^p\rangle r^q$ and their Rayleigh–Schrödinger perturbation expansions

Edward R. Vrscay
Department of Applied Mathematics, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

(Received 21 July 1987; accepted for publication 21 October 1987)

Rayleigh–Schrödinger perturbation expansions for eigenvalues $E(\lambda)$ of nonlinear Hamiltonians of the form $\hat{H}^{(0)}+\lambda\langle r^p\rangle r^q$, $p,q\geqslant 1$ are calculated using hypervirial (HV) and Hellmann–Feynman (HF) theorems. Such Hamiltonians are similar in form to those employed in the study of "self-interacting" systems, e.g., solute–solvent interactions. The specific cases considered for $H^{(0)}$ are one-dimensional harmonic oscillators and hydrogen atoms. The eigenvalue expansions for the nonlinear problems are compared with those of the linear problems where $p=0$, whose large-order behavior and summability properties are well-known. Also examined are the perturbation expansions for the expectation values $\langle r^k\rangle$, which are also products of the HVHF method.

I. INTRODUCTION

Nonlinear Schrödinger equations with the generic form

$$[\hat{H}^{(0)}+\hat{V}(\psi)]\psi=E\psi \tag{1.1}$$

have been used to describe quantum mechanical systems that interact with their environment.[1,2] Through its wave function ψ, the system interacts with its surroundings by, for example, inducing a net field which then acts back on the system itself. This "self-dependent" situation could be described by appropriate choice of the interaction operator $\hat{V}$ in (1.1). The linear Schrödinger equation,

$$\hat{H}^{(0)}\psi^{(0)}=E^{(0)}\psi^{(0)}, \tag{1.2}$$

is assumed to describe the system *in vacuo*, i.e., isolated and independent of its environment.

Such treatments have been employed in the studies of molecules immersed in polar solvents, a problem of prime importance in the study of biological systems. An electronic state of the molecule, described by a wave function ψ, induces an energetically favorable orientation of polar solvent molecules (e.g., water molecules) that surround it. This orientation produces a field that, in turn, acts on the molecule in question. The Hamiltonians used to describe such systems have assumed the form

$$[H^{(0)}+\lambda\langle\psi|\hat{A}|\psi\rangle\hat{B}]\psi=E(\lambda)\psi. \tag{1.3}$$

Again, the linear Schrödinger eigenvalue equation in (1.2) is usually assumed to describe the solute molecule *in vacuo*. Using the Kirkwood–Onsager model,[3] $\hat{A}=\hat{B}=\hat{M}$, the dipole moment operator. The perturbation parameter λ could describe the strength of the solute–solvent interaction. A variety of methods have been used to approximate the solutions of (1.3). Of course, even in the case of atoms or small molecules, the solution of (1.2) may be a formidable task.

This report was motivated by the paper of Surjan and Angyan[2] in which was developed a formal nondegenerate Rayleigh–Schrödinger (RS) perturbation theory for problems of the form in Eq. (1.3), assuming the solvability of the unperturbed problem in Eq. (1.2). The presence of the self-interaction introduces nonlinear contributions to the perturbation corrections to E and ψ. In the special case $\hat{A}=\hat{1}$, the identity operator, the perturbation formulas reduce to the usual Rayleigh–Schrödinger perturbation theory (RSPT). In principle, the method may be used to calculate perturbation corrections to arbitrary order.

Here, we examine eigenvalue perturbation expansions for simple radial oscillators, defined by the Hamiltonians

$$\hat{H}^{(p,q)}=\tfrac{1}{2}\hat{p}^2+\tfrac{1}{2}r^2+\lambda\langle r^{2p}\rangle r^{2q},\quad p,q=1,2,3,..., \tag{1.4}$$

where $\langle r^{2k}\rangle$ denotes the expectation value

$$\langle r^{2k}\rangle\equiv\frac{\int\psi^*(\underline{r})r^{2k}\psi(\underline{r})\,d\underline{r}}{\int\psi^*(\underline{r})\psi(\underline{r})\,d\underline{r}}, \tag{1.5}$$

ψ being an eigenstate of $\hat{H}^{(p,q)}$. Equation (1.4) can be considered to define such oscillators in arbitrary space dimensions, but the present analysis is restricted to one-dimensional problems. The eigenvalue expansions will assume the usual form of RSPT, i.e.,

$$E_K^{(p,q)}(\lambda)=\sum_{n=0}^{\infty}E_K^{(p,q)(n)}\lambda^n, \tag{1.6}$$

where $E_K^{(0)}=K+\frac{1}{2}$. Their behavior will be related to that of the well-known expansions associated with the "linear" perturbation problems where $p=0$, which correspond to the anharmonic oscillators studied in the context of quantum field theory.[4–9] In the spirit of our introductory remarks, the Hamiltonians in (1.4) could be viewed as defining self-interacting oscillators whose anharmonicities are directly proportional to the mean values of given powers of their vibrational amplitudes.

Specifically, we examine the large-order behavior of the RS coefficients $E_K^{(p,q)(n)}$. With the aid of numerical computations, the summability of these series is also conjectured. The coefficients are easily calculated by a method originally developed by Swenson and Danforth[10] to study perturbed oscillator problems. Their method, which employed the hypervirial (HV) and Hellmann–Feynman (HF) theorems, and which will henceforth be referred to as the HVHF method, permits a calculation of the (nondegenerate) eigenvalue series without a knowledge of wave functions. In short, no

 0022-2488/88/040901-11$02.50

matrix elements are needed and the only input into the algorithm is the unperturbed energy $E_K(0) = E_K^{(0)}$. A by-product of this approach is that it yields formal perturbation expansions for the expectation values $\langle r^k \rangle$. The HVHF method is reviewed for general N-dimensional problems in Sec. II. Section III is devoted to its application to oscillator problems. In addition to the eigenvalue expansions, the perturbation series for $\langle r^k \rangle$ will be examined in detail for the linear eigenvalue problems. These series also possess interesting large-order and summability properties which are useful for an understanding of the nonlinear expansions.

The HVHF perturbative method has been applied to hydrogenic problems by Killingbeck[11] and a number of other workers, e.g., Refs. 12–15. In these problems, the traditional difficulties posed by the continuum spectrum of the unperturbed hydrogen Hamiltonian operator are bypassed. In Sec. IV, the HVHF method is applied to the following nonlinear hydrogenic counterparts:

$$\hat{H} = \tfrac{1}{2}\hat{p}^2 - Z/r + \lambda \langle r^p \rangle r^q, \quad p,q = 1,2,3,\ldots . \tag{1.7}$$

For both oscillator and hydrogenic cases, the nonlinear expansions will be shown to be intimately related to the corresponding expansions for the linear problems, i.e., where $p = 0$. In Sec. V we apply a "renormalization method" to the RS series in Eq. (1.6) to accurately calculate $E(\lambda)$ for the entire infinite range of coupling constant values $0 < \lambda < \infty$. In addition, the eigenvalues of the infinite-field Hamiltonians

$$\hat{H}_\infty^{(p,q)} = \tfrac{1}{2}\hat{p}^2 + \langle r^{2p} \rangle r^{2q} \tag{1.8}$$

are calculated from the renormalized perturbation series.

II. HYPERVIRIAL AND HELLMANN–FEYNMAN (HVHF) THEOREMS AND PERTURBATION THEORY AT LARGE ORDER

In this section, we outline the essential aspects of the HVHF method as applied to N-space-dimensional eigenvalue equations of the form

$$\hat{H}\psi_n = (\hat{T} + \hat{V})\psi_n = E_n \psi_n , \tag{2.1}$$

where $\hat{T}$ is the kinetic energy operator and $\hat{V} = V(r)$ is the spherically symmetric potential energy operator. Since the method is relatively well-known and has been applied by many researchers, the following description is brief. The reader is referred to a new monograph on the subject by Fernandez and Castro.[16] The comprehensive review article by Marc and McMillan[17] is also recommended for a discussion of the viral theorem and its applications in both classical and quantum mechanics.

The case of radial potentials $\hat{V} = V(r)$ represents a relatively simple set of eigenvalue problems. A separation-of-variables approach, i.e., assuming $\psi(\underline{r}) = R(r)Y(\Omega)$, factors out the angular Ω dependence in terms of N-dimensional spherical harmonics. The result is the radial eigenvalue equation[18]

$$\hat{H}R_{nl}(r) = [\tfrac{1}{2}\hat{p}_r^2 + L^2/2r^2]R_{nl}(r) = E_n R_{nl}(r) , \tag{2.2}$$

where the operator

$$\hat{p}_r^2 = -[\hat{D}^2 + ((N-1)/r)\hat{D}]$$

is the square of the radial momentum operator in N-dimensional space, with $\hat{D} \equiv d/dr$. The indices n and l represent the radial and (one of the) angular momentum quantum numbers, respectively. The scalar $L^2 = l(l+N-2)$ represents the eigenvalue of the square of the N-dimensional angular momentum operator $\hat{L}^2$.

Now, given any linear operator $\hat{O}$, the following expectation values vanish ($[\hat{A},\hat{B}] \equiv \hat{A}\hat{B} - \hat{B}\hat{A}$):

$$\langle [\hat{O},\hat{H}] \rangle \equiv \int \psi^*(\underline{r})[\hat{O},\hat{H}]\psi(\underline{r})\, d\underline{r} = 0 , \tag{2.3}$$

for all eigenstates ψ_n of $\hat{H}$, by virtue of the self-adjointness of $\hat{H}$. The wave function ψ is always assumed to be normalized to unity. If we choose $\hat{O} = r^k\hat{D}$ and evaluate explicitly the commutator in Eq. (2.3), a set of recursion relations involving the expectation values, $\langle r^k \rangle \equiv \int \psi r^k \psi\, d\underline{r}$, are obtained. These equations are known as the *hypervirial relations.*[19] In order to evaluate the commutators, we begin with the following relations, which are easily derived from Eq. (2.2):

$$[\hat{D},\hat{H}] = (\hat{D}V) + ((N-1)/2r^2)\hat{D} - L^2/r^3 , \tag{2.4}$$

$$[r^k,\hat{H}] = kr^{k-1}\hat{D} + \tfrac{1}{2}k(k+N-2)r^{k-2} , \tag{2.5}$$

where $(\hat{D}V) \equiv dV/dr$.

The operator identity $[r^k\hat{D},\hat{H}] = r^k[\hat{D},\hat{H}] + [r^k,\hat{H}]\hat{D}$ is now used to rewrite the commutator in Eq. (2.3). Any appearance of $\hat{D}$ and $\hat{D}^2$ is then eliminated with the use of Eqs. (2.4) and (2.5). Taking expectation values with respect to the eigenstate ψ_n yields the following hypervirial relations for an arbitrary radial potential $V(r)$:

$$\begin{aligned} 2kE\langle r^{k-1}\rangle &= \langle r^k(\hat{D}V)\rangle + 2k\langle r^{k-1}V\rangle + (k-1)L^2\langle r^{k-3}\rangle \\ &\quad - \tfrac{1}{4}(k+N-3)[k(k-N)+N-1]\langle r^{k-3}\rangle , \\ &\quad k = \ldots, -2, -1,0,1,2,\ldots . \end{aligned} \tag{2.6}$$

The case $k = 1$ corresponds to the quantum mechanical virial theorem.[20] Since $V(r)$ will generally be a sum of powers of r, Eq. (2.6) will imply a recurrence relation between expectation values $\langle r^k \rangle$. Moreover, for perturbation problems, $V(r)$ will include the coupling constant λ. The essence of the HVHF perturbative method is to assume the following expansions for a given state in question:

$$E = \sum_{n=0}^{\infty} E^{(n)}\lambda^n , \tag{2.7}$$

$$\langle r^k \rangle = \sum_{n=0}^{\infty} C_k^{(n)}\lambda^n . \tag{2.8}$$

It is assumed that the unperturbed energy $E^{(0)}$ is known. Also, the normalization condition $\langle r^0 \rangle = 1$ is imposed, implying that

$$C_0^{(n)} = \delta_{0n} . \tag{2.9}$$

Substitution of these expansions into Eq. (2.6) and collection of like powers of λ^n yield a difference equation in the coefficients $C_i^{(n)}$ and $E^{(k)}$. An application of the Hellmann–Feynmann theorem,[21]

$$\frac{dE}{d\lambda} = \left\langle \frac{\partial H}{\partial \lambda} \right\rangle = \sum_{n=0} (n+1)E^{(n+1)}\lambda^n , \tag{2.10}$$

defines the relation between the $E^{(k)}$ and the $C_i^{(j)}$. Generally, the $C_i^{(j)}$ array is calculated "columnwise," starting from the $n=0$ (unperturbed) column, whose elements are computed recursively. The triangular nature of this computation will be seen in the examples that follow.

In traditional "textbook" presentations of Rayleigh–Schrödinger perturbation theory, little attention is paid to questions concerning the nature of the expansions obtained. Usually, it is naively assumed that for sufficiently small λ the series converges. Of course, these questions were addressed many years ago[22] and have continued to receive attention. Many eigenvalue expansions encountered in nonrelativistic quantum mechanics are divergent, yet asymptotic to $E(\lambda)$ on some sector in the complex λ plane. Their large-order behavior is typically given by

$$E^{(n)} \sim (-1)^{n+1} A\Gamma(mn+\alpha)b^n, \quad n\to\infty, \tag{2.11}$$

where A, B, α, and m are constants, with $m=1,2,3,\ldots$. Also, $E(\lambda)$ is usually analytic in an appropriate sector of the complex λ plane, which includes the real λ line. This ensures the existence of the asymptotic expansion in Eq. (2.10) within the sector.[23] In many cases, the above properties can be used to establish the Borel summability[24,25] of perturbation series to $E(\lambda)$ on some suitable sector of the complex λ plane which includes the positive real line.

Since we have been motivated in the past by the intimate relationship between continued fractions (CF) and RSPT,[26] this paper also considers the CF representations of the expansion given in Eqs. (2.7) and (2.8). These representations assume the form

$$E(\lambda) = E^{(0)} + \lambda C(\lambda), \tag{2.12}$$

where

$$C(z) = \frac{c_1}{1+}\,\frac{c_2\lambda}{1+}\,\frac{c_3\lambda}{1+}\cdots. \tag{2.13}$$

The reader is referred to Refs. 27 and 28 for comprehensive treatments of the analytic theory of continued fractions. The properties of continued fractions relevant to RSPT are given in Ref. 26.

The RS eigenvalue expansions for many standard perturbation problems, such as anharmonic oscillators, are negative Stieltjes for $n\geqslant 1$.[7] This implies that $C(z)$ in Eq. (2.13) is an S fraction, i.e., all coefficients c_n are positive. Moreover, when the Stieltjes coefficients behave asymptotically as in Eq. (2.11), then[26]

$$c_n = O(n^m), \quad \text{as } n\to\infty. \tag{2.14}$$

In particular, when $m=1$, then

$$c_n \sim \tfrac{1}{2}bn, \quad \text{as } n\to\infty. \tag{2.15}$$

When $m<2$, Carleman's condition[28] is satisfied, which is sufficient to guarantee Padé summability of the RS series.

The convergents of $C(z)$, denoted $w_n(z)$, are obtained by truncating $C(z)$, i.e., setting $c_{n+1}=0$. They are rational functions of z. The convergents $w_{2N}(z)$ and $w_{2N+1}(z)$ correspond, respectively, to the $[N-1,N]$ and $[N,N]$ Padé approximants[29] to the series being represented. If the series is a Stieltjes series,[28] then the sequences $\{w_{2N}(z)\}$ and $\{w_{2N+1}(z)\}$, $N=0,1,2,\ldots$, provide, respectively, lower and upper bounds to $E(z)$. If the series is Padé summable (corresponding to determinacy of the moment problem), then these sequences converge to $E(z)$ in the limit $N\to\infty$. The numerical calculations displayed in this report have used only continued fractions to "sum" the perturbation series.

III. SPECIFIC APPLICATION TO NONLINEAR OSCILLATORS

In this section, we examine the perturbation expansions associated with the one-dimensional oscillators

$$\hat{H}^{(p,q)} = -\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}x^2 + \lambda\langle x^{2p}\rangle x^{2q}, \tag{3.1}$$

with unperturbed energies $E_K^{(0)} = K+\frac{1}{2}$, $K=0,1,2,\ldots$. The hypervirial relations in Eq. (2.6) become

$$\begin{aligned}(2k+1)E\langle x^{2k}\rangle &= (k+1)\langle x^{2k+2}\rangle + \lambda(q+2k+1)\langle x^{2(k+q)}\rangle\langle x^{2p}\rangle \\ &\quad - \tfrac{1}{4}k(2k+1)(2k-1)\langle x^{2k-2}\rangle, \\ &\quad k=0,1,2,\ldots.\end{aligned} \tag{3.2}$$

Since expectation values of odd powers of x vanish, we let

$$\langle x^{2k}\rangle = \sum_{n=0}^{\infty} C_k^{(n)}\lambda^n. \tag{3.3}$$

The HF theorem implies that

$$\frac{dE}{d\lambda} = \langle x^{2q}\rangle \frac{d}{d\lambda}[\lambda\langle x^{2p}\rangle]. \tag{3.4}$$

Equations (3.2)–(3.4) then yield the following recurrence relations for the $C_k^{(n)}$ and the $E^{(n)}$:

$$\begin{aligned}kC_k^{(n)} &= (2k-1)\sum_{j=0}^{n} E^{(j)}C_{k-1}^{(n-j)} \\ &\quad -(q+2k-1)\sum_{j=0}^{n-1} C_p^{(j)}C_{q+k-1}^{(n-1-j)} \\ &\quad +\tfrac{1}{4}(k-1)(2k-1)(2k-3)C_{k-2}^{(n)},\end{aligned} \tag{3.5}$$

$$E^{(n+1)} = \frac{1}{n+1}\sum_{j=0}^{n}(j+1)C_p^{(j)}C_q^{(n-j)}. \tag{3.6}$$

In order to determine $E^{(n+1)}$, one calculates the columns $C_k^{(j)}$, where $j=0,1,\ldots,n$, and $k=1,2,\ldots,\max(p,q)+(n-j)q$. Note that the entries $C_k^{(0)}$ are the same for all perturbed oscillator problems, representing the unperturbed expectation values $\langle x^{2k}\rangle(0)$ associated with the harmonic oscillator eigenfunctions. These expectation values are functions of the unperturbed eigenvalue $E_K^{(0)}$. The first five entries are given for reference in Table I. The $C_i^{(j)}$ table and the RS coefficients $E^{(n)}$ may be calculated in algebraic or in rational number form using a symbolic manipulation algorithm (the MAPLE language being developed at Waterloo[30] has been used for these purposes), or in floating-point form. Since the $C_k^{(n)}$ will generally grow rapidly, especially as q in Eq. (3.1) increases, it may be necessary to avoid exponential overflow in floating-point calculations by scaling the coefficients. For example, define $D_k^{(n)} = C_k^{(n)}s^{k+n}$, where $0<s<1$ is a scaling factor, conveniently some power of 10. Then rewrite the recursion relations (3.5) and (3.6) in

TABLE I. Expectation values $\langle x^k \rangle$, $k = 0,1,...,5$, of the unperturbed harmonic oscillator eigenstates, expressed in terms of the unperturbed eigenvalue $E_K^{(0)} = K + \frac{1}{2}$. These entries define the first column of the $C_k^{(n)}$ table in Eq. (3.5).

k	$C_k^{(0)} = \langle x^k \rangle (0)$
0	1
1	$t \equiv E_K^{(0)} = K + \frac{1}{2}$
2	$\frac{3}{2} t^2 + \frac{3}{8}$
3	$\frac{5}{8} t(4t^2 + 5)$
4	$\frac{35}{8} t^4 + \frac{245}{16} t^2 + \frac{315}{128}$
5	$\frac{63}{8} t^5 + \frac{945}{16} t^3 + \frac{5607}{128} t$

terms of the $D_k^{(n)}$. In this way, a set of scaled perturbation coefficients $\overline{E}^{(n)} = s^n E^{(n)}$ is obtained.

A. Nonlinear harmonic oscillators ($q = 1$)

The Hamiltonians

$$\hat{H} = -\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}x^2 + \lambda \langle x^{2p} \rangle x^2 \tag{3.7}$$

could be considered as describing harmonic oscillators with force constants directly proportional to mean values of powers of the vibrational amplitudes. Since the eigenvalues of the oscillators may be found exactly as roots of polynomials, they serve as a good testing ground for approximation methods. The special case $p = q = 1$ has been used by Cioslowski[31] to demonstrate a method of connected moments, and by Handy[32] for another method of moments.

The quantum mechanical virial theorem [$k = 0$ in Eq. (3.2)] states that for the harmonic oscillator (i.e., $\lambda = 0$) in Eq. (3.7), $\langle x^2 \rangle = E = E_K^{(0)}$, $K = 0,1,2,...$. If the eigenvalue equation associated with (3.7) is scaled as $x \to \alpha^{1/2} x$, $\alpha \in \mathbb{R}$, then

$$\left[-\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}\alpha^2 (1 + 2\lambda \alpha^p \langle x^{2p} \rangle) x^2 \right] \psi = \alpha E \psi. \tag{3.8}$$

Choosing α so that

$$\alpha^2 (1 + 2\lambda \alpha^p \langle x^{2p} \rangle) = 1, \tag{3.9}$$

we have $\langle x^2 \rangle = \alpha E = E_K^{(0)}$. (When $\alpha = 1$, then $E = E_K^{(0)}$, our unperturbed state.) Equation (3.9) may be rewritten as

$$E^{p+2} - (E_K^{(0)})^2 E^p - 2\lambda (E_K^{(0)})^{p+2} \langle x^{2p} \rangle = 0. \tag{3.10}$$

Since the scaled Hamiltonian in (3.8) is a harmonic oscillator, $\langle x^{2p} \rangle$ in (3.10) is a function of $E_K^{(0)}$. It has been given explicitly in Table I for $p = 0,1,...,5$. The root of (3.10), which is a continuation of the unperturbed energy $E_K^{(0)}$ for $\lambda \neq 0$, is the desired eigenvalue $E(\lambda)$ of (3.7).

The RS expansions for $E(\lambda)$ could be obtained either directly from the polynomial equations (3.10) or by the HVHF method. The radius of convergence of each expansion, to be denoted as R_p (possibly dependent upon the quantum number K), will be the distance from $\lambda = 0$ to the nearest branch point singularities of (3.10), which will correspond to the multiple roots. The R_p are easily determined in closed form for $p = 0,1,2$. We summarize results below, presenting the first few terms of each RS expansion for $E(\lambda)$, its λ radius of convergence, and its continued fraction representation.

(i) $p=0$: The expansion is trivial here since $E(\lambda) = (1 + 2\lambda)^{1/2} E_K^{(0)}$, $R_p = \frac{1}{2}$. We have

RSPT: $E_K(\lambda)$

$$= E_K^{(0)} [1 + \lambda - \tfrac{1}{2}\lambda^2 + \tfrac{1}{2}\lambda^3 - \tfrac{5}{8}\lambda^4 + \tfrac{7}{8}\lambda^5 - \cdots],$$

$$\text{CF: } E_K(\lambda) = E_K^{(0)} + \frac{E_K^{(0)}\lambda}{1+}\,\frac{\frac{1}{2}\lambda}{1+}\,\frac{\frac{1}{2}\lambda}{1+}\,\frac{\frac{1}{2}\lambda}{1+} \cdots.$$

(ii) $p=1$: From (3.10), $R_p = (3\sqrt{3}\, E_K^{(0)})^{-1}$. We have

$$\text{RSPT: } E_K(\lambda) = E_K^{(0)} [1 + \beta - \tfrac{3}{2}\beta^2 + 4\beta^3 - \tfrac{105}{8}\beta^4 + 48\beta^5 - \cdots],$$

$$\text{CF: } E_K(\lambda) = E_K^{(0)} + \frac{E_K^{(0)}\beta}{1+}\,\frac{\frac{3}{2}\beta}{1+}\,\frac{\frac{7}{6}\beta}{1+}\,\frac{\frac{59}{32}\beta}{1+}\,\frac{\frac{1005}{826}\beta}{1+} \cdots,$$

where $\beta = \lambda E_K^{(0)}$.

(iii) $p=2$: $R_p = [8\langle x^4 \rangle (0)]^{-1}$. We have

$$\text{RSPT: } E_K(\lambda) = E_K^{(0)} [1 + g - \tfrac{5}{2}g^2 + \tfrac{21}{2}g^3 - \tfrac{429}{8}g^4 + \tfrac{2431}{8}g^5 - \cdots],$$

$$\text{CF: } E_K(\lambda) = E_K^{(0)} + \frac{E_K^{(0)}g}{1+}\,\frac{\frac{5}{2}g}{1+}\,\frac{\frac{17}{10}g}{1+}\,\frac{\frac{381}{170}g}{1+}\,\frac{\frac{23\,525}{12\,954}g}{1+} \cdots,$$

where $g = \lambda \langle x^4 \rangle (0) = \lambda\, [\frac{3}{2}(E_K^{(0)})^2 + \frac{3}{8}]$.

In all cases, the continued fractions are S fractions. Numerical asymptotic analysis of the coefficients shows that $c_n \to (4R_p)^{-1}$. This behavior is consistent with Van Vleck's theorem (see Ref. 33, p. 138): Let $C(z)$ be an S fraction such that $\lim_{n\to\infty} c_n = a \neq 0$, $a \in \mathbb{C}$. Then the continued fraction $C(z)$ converges to a function $f(z)$ that is meromorphic (or identically infinite) in the cut complex plane $C_a \equiv \{z: |\arg(az + \frac{1}{4})| < \pi\}$ (complex z plane with branch cut extending outward from $z = a$ toward $z = \infty$ on the line which is an extension of the line segment connecting $z = a$ and $z = 0$ in the plane). The behavior of the c_n suggests that $E(\lambda)$ is at least meromorphic in the complex λ plane with branch cut on $(-\infty, -R_p)$.

B. Quartic anharmonic oscillators ($q = 2$)

We now focus attention on the Hamiltonians

$$\hat{H}^{(p,2)} = -\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}x^2 + \lambda \langle x^{2p} \rangle x^4, \quad p \geqslant 1, \tag{3.11}$$

and relate their eigenvalue expansions with those of the corresponding "linear" Bender–Wu (BW) oscillators[4] (with different normalization), where $p = 0$:

$$\hat{H}^{(0,2)} = -\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}x^2 + \lambda x^4. \tag{3.12}$$

The BW expansions associated with Eq. (3.12) will be denoted

$$E_K(\lambda) = \sum_{n=0}^{\infty} A_K^{(n)} \lambda^n, \tag{3.13}$$

$$\langle x^{2k}\rangle(\lambda) = \sum_{n=0}^{\infty} a_k^{(n)}\lambda^n . \tag{3.14}$$

The first three BW coefficients will be useful in this section:

$$A_K^{(0)} = E_K^{(0)} = K + \tfrac{1}{2}, \quad A_K^{(1)} = \tfrac{3}{8}[4(E_K^{(0)})^2 + 1],$$
$$A_K^{(2)} = -\tfrac{1}{16}E_K^{(0)}[68(E_K^{(0)})^2 + 67] . \tag{3.15}$$

The large-order behavior of the $A_K^{(n)}$ was first determined in Refs. 4 and 5 using WKB techniques [the RS series in Eq. (3.13) coincides with that of the original BW anharmonic oscillators]:

$$A_K^{(n)} \sim (-1)^{n+1} D_K \Gamma(n + K + \tfrac{1}{2}) 3^n [1 + O(1/n)] , \tag{3.16}$$

where

$$D_K = (12^K/K!)(6/\pi^3)^{1/2} . \tag{3.17}$$

The series is Borel summable to $E(\lambda)$ over the complex plane with cut on $(-\infty, -B)$ for some $B > 0$.[9] The coefficients $A_K^{(n)}$ are negative Stieltjes for $n \geqslant 1$, and the RS series is Padé summable to $E(\lambda)$ on compact subsets of the cut plane $|\arg \lambda| < \pi$, the first Riemann sheet of $E(\lambda)$.[8] The continued fraction representation of the RS series is thus an S fraction, and its coefficients behave asymptotically as[26]

$$c_n \sim \frac{3}{2} n + \frac{3}{4} K + \frac{(-1)^n}{8}, \quad n \to \infty . \tag{3.18}$$

The hypervirial equations of (3.2) for the BW oscillator become

$$\begin{aligned}(2k+1)E\langle x^{2k}\rangle &= (k+1)\langle x^{2k+2}\rangle + \lambda(2k+3)\langle x^{2k+4}\rangle \\ &\quad - \tfrac{1}{4}k(2k+1)(2k-1)\langle x^{2k-2}\rangle , \\ &\quad k = 0,1,2,\dots .\end{aligned} \tag{3.19}$$

From the Hellmann–Feynman theorem, $dE/d\lambda = \langle x^4\rangle$, we have the result

$$a_2^{(n)} \sim (-1)^n D_K \Gamma(n + K + \tfrac{5}{2}) 3^{n+1}, \quad n \to \infty . \tag{3.20}$$

By setting $k = 0$ in Eq. (3.19) we also find that

$$a_1^{(n)} \sim (-1)^n D_K \Gamma(n + K + \tfrac{3}{2}) 3^{n+1}, \quad n \to \infty . \tag{3.21}$$

From these results, and from a repeated application of the difference equation for the $C_i^{(j)}$ associated with Eq. (3.19), we arrive at the following general result for the large-order behavior of the expansion coefficients in (3.14):

$$a_k^{(n)} \sim (-1)^n B_k D_K \Gamma(n + K + \tfrac{1}{2} + k) 3^{n+1}, \quad n \to \infty , \tag{3.22}$$

where $B_1 = B_2 = 1$, and

$$B_k = \prod_{j=3}^{k} \frac{(j-1)}{(2j-1)}, \quad k \geqslant 3 . \tag{3.23}$$

This formula has also been verified by numerical asymptotic analysis of the $a_k^{(n)}$, for $k = 1,2,\dots,5$. The analyticity of $E(\lambda)$ along with the recursion relation (3.19) ensures analyticity of $\langle x^{2k}\rangle(\lambda)$, $k = 1,2,3,\dots$, on the cut plane, $|\arg \lambda| < \pi$. The large-order behavior of the $a_k^{(n)}$ in (3.22) establishes Borel summability of the expansions in (3.14) to $\langle x^{2k}\rangle(\lambda)$ on a strip which contains the real λ line.

The continued fraction representations of the $\langle x^{2k}\rangle$ series, having the form

$$\langle x^{2k}\rangle(\lambda) = C_k(\lambda) = \frac{c_{k1}}{1+}\,\frac{c_{k2}\lambda}{1+}\,\frac{c_{k3}\lambda}{1+}\cdots , \tag{3.24}$$

have also been computed to order $n = 70$ for $k = 1,2,\dots,5$. In all cases, the $C_k(\lambda)$ are observed to be S fractions. Numerical asymptotic analysis indicates that

$$c_{kn} \sim \tfrac{3}{2}n + O(1) , \quad n \to \infty . \tag{3.25}$$

On the basis of this numerical evidence, we conjecture the Stieltjes nature of the $\langle x^{2k}\rangle$ expansions in (3.14). The $n!$ growth of the coefficients $a_k^{(n)}$ for $k \geqslant 1$ suggests the Padé summability of the series over compact subsets of the cut plane $|\arg \lambda| < \pi$, in accordance with Carleman's condition.[28]

The properties of the BW oscillator expansions will now be useful for an understanding of the nonlinear problems in Eq. (3.11). For notational convenience, indices referring to p and q as well as to the quantum number K will be omitted unless there may be an ambiguity. The nonlinear problems may be considered to define a new coupling constant β,

$$\beta = \lambda \langle x^{2p}\rangle = \sum_{n=0}^{\infty} C_p^{(n)} \lambda^{n+1} . \tag{3.26}$$

To lowest order in λ, $\beta \sim C_p^{(0)}\lambda$, so that we would expect the geometric factor 3^k in Eqs. (3.15) and (3.22) to be replaced by $(3C_p^{(0)})^k$. This will indeed be verified below. A formal relationship between the "nonlinear" RS coefficients $E^{(n)}$ and the BW coefficients $A^{(n)}$ may be obtained by equating powers of λ^n in the relation

$$E(\lambda) = \sum_{n=0}^{\infty} E^{(n)}\lambda = \sum_{j=0}^{\infty} A^{(j)}\beta^j . \tag{3.27}$$

If we set $g_n \equiv C_p^{(n)}$, then the first few relations become

$$\begin{aligned} &E^{(0)} = A^{(0)}, \quad E^{(1)} = A^{(1)} g_0 , \\ &E^{(2)} = A^{(1)} g_1 + A^{(2)} g_0^2 , \\ &E^{(3)} = A^{(1)} g_2 + 2A^{(2)} g_0 g_1 + A^{(3)} g_0^3 . \end{aligned} \tag{3.28}$$

The general formula for $E^{(n)}$, $n \geqslant 1$, in (3.28) may be written as

$$\frac{E^{(n)}}{A^{(n)} g_0^{(n)}} = \left[\sum_{k=0}^{n} \binom{n-1}{k} \frac{A^{(n-k)}}{A^{(n)}} \left(\frac{g_1}{g_0^2}\right)^k \right] + \frac{A^{(1)}}{A^{(n)}} \frac{g_{n-1}}{g_0^n} + \frac{2A^{(2)}}{A^{(n)}} \frac{g_{n-2}}{g_0^{n-1}} + \cdots . \tag{3.29}$$

The pattern exhibited in these equations indicates that the asymptotic behavior of the $E^{(n)}$ will be dependent upon $g_n = C_p^{(n)}$ as well as the $A^{(n)}$. From Eq. (3.16), the partial sum in square brackets becomes, in the limit $n \to \infty$, $\exp(-g_1/(3g_0^2))$. It now remains to determine the asymptotics implied by the remaining terms, which will be done below for the particular cases $p = 1,2$. Keeping in mind the behavior of the "linear" expectation series coefficients $a_k^{(n)}$ in Eq. (3.22) and the remarks following Eq. (3.26), two assumptions on the asymptotic behavior of the g_n will be made in the analysis to follow:

$$C_{p+1}^{(n)}/C_p^{(n)} \sim O(n)\,, \quad \text{as } n\to\infty\,, \tag{3.30a}$$

$$\lim_{n\to\infty} (g_{n+1}/ng_n) = -3g_0\,, \tag{3.30b}$$

the latter implying an $n!$ growth of the form (3.26), replacing the geometric term 3 by $3g_0$.

Case 1: $p=1$, $g_n \equiv C_1^{(n)}$. First, rewrite (3.30) as

$$\frac{E^{(n)}}{A^{(n)}g_0^n}\left[1 - A^{(1)}\frac{g_{n-1}}{E^{(n)}}\right] \sim e^{-g_1/(3g_0^2)}\,, \tag{3.31}$$

where we have used (3.29) to ignore the contributions of the remaining terms in (3.30). The relevant parameters are $g_0 = E^{(0)}$ and $g_1 = -2E^{(0)}A^{(1)}$. It now remains to determine the asymptotic behavior of the second term in square brackets. From the hypervirial relation [$k=0$ in Eq. (3.2)]

$$E = \langle x^2\rangle + 3\lambda\,\langle x^2\rangle\langle x^4\rangle\,, \tag{3.32}$$

it follows that

$$\frac{E^{(n)}}{C_1^{(n-1)}} = \frac{C_1^{(n)}}{C_1^{(n-1)}} + 3C_1^{(1)}\frac{C_2^{(n-1)}}{C_1^{(n-1)}} + 3\frac{C_2^{(n-2)}}{C_1^{(n-1)}} + 3C_2^{(0)} + O\left(\frac{1}{n}\right). \tag{3.33}$$

From Eq. (3.6) we also have

$$\frac{E^{(n)}}{C_1^{(n-1)}} = \frac{C_1^{(0)}C_2^{(n-1)}}{nC_1^{(n-1)}} + C_2^{(0)} + O\left(\frac{1}{n}\right). \tag{3.34}$$

From the two assumptions in Eq. (3.30), the first two terms on the rhs of (3.33) behave as $O(n)$. Equating (3.33) and (3.34), and using (3.29), reveals that

$$C_2^{(n-1)}/nC_1^{(n-1)} \sim 1\,, \quad \text{as } n\to\infty\,. \tag{3.35}$$

From Eq. (3.34), and the fact that $C_2^{(0)} = A^{(1)}$,

$$E^{(n)}/C_1^{(n-1)} = A^{(0)} + A^{(1)}\,, \tag{3.36}$$

which, when substituted into (3.31), gives

$$E_K^{(n)} \sim [1 + A_K^{(1)}/A_K^{(0)}] \times \exp[2A_K^{(1)}/3A_K^{(0)}](A_K^{(0)})^n A_K^{(n)}\,, \quad \text{as } n\to\infty\,. \tag{3.37}$$

This formula has been verified by numerical asymptotic analysis of the expansions corresponding to $K = 0,1,\ldots,6$.

Case 2: $p=2$, $g_n \equiv C_2^{(n)}$. From Eq. (3.6), we have

$$E^{(n)} = \frac{1}{n}\sum_{j=0}^{n-1}(j+1)g_j\,g_{n-1-j}\,, \tag{3.38}$$

so that

$$\frac{E^{(n)}}{g_{n-1}} = g_0 + \left[g_0 - \frac{g_1}{3g_0}\right]\frac{1}{n} + O\left(\frac{1}{n}\right). \tag{3.39}$$

Also,

$$\frac{E^{(n)}}{g_{n-2}} = \frac{E^{(n)}}{g_{n-1}}\frac{g_{n-1}}{g_{n-2}} = \frac{E^{(n)}}{g_{n-1}}[-3ng_0 + O(1)]\,. \tag{3.40}$$

Equation (3.29) is then rewritten as

$$\frac{E^{(n)}}{A^{(n)}g_0^n}\left[1 - A^{(1)}\frac{g_{n-1}}{E^{(n)}} - 2A^2g_0\frac{g_{n-2}}{E^{(n)}}\right] \sim e^{-g_1/(3g_0^2)}\,, \quad n\to\infty\,. \tag{3.41}$$

[Note the difference between (3.41) and (3.31).] For this case we compute $g_0 = A^{(1)}$ and $g_1 = 2A^{(2)}A^{(1)}$, so that the term in square brackets reduces to $A^{(1)}/(g_0n) = 1/n$. The net result is

$$E_K^{(n)} \sim n\exp[-2A_K^{(2)}/3A_K^{(1)}](A_K^{(1)})^n A_K^{(n)}\,, \quad n\to\infty\,, \tag{3.42}$$

which we may write as

$$E_K^{(n)} \sim (-1)^{n+1}\exp[-2A_K^{(2)}/3A_K^{(1)}] \times D_K\Gamma(n+K+\tfrac{1}{2})(3A_K^{(1)})^n\,, \quad n\to\infty\,. \tag{3.43}$$

A final note must be made concerning the relations in (3.37) and (3.43). The exponential factors occurring in these relations can easily be obtained by substituting the relevant form of Eq. (3.26) into the Bender–Wu formula for the asymptotics of $\operatorname{Im} E(\lambda)$, $\lambda\to 0^-$ [cf. Eq. (A8), p. 1635 of Ref. 4(b)]. However, this naive treatment ignores the contribution of terms such as g_{n-1} in Eq. (3.29) or (3.31).

IV. APPLICATIONS TO RADIAL HYDROGENIC PROBLEMS

The application of the HVHF and renormalization methods to the (N-dimensional) hydrogenic problems

$$\hat{H}^{(p,q)} = \frac{1}{2}\hat{p}_r^2 + \frac{\hat{L}^2}{2r^2} - \frac{Z}{r} + \lambda\,\langle r^p\rangle\, r^q \tag{4.1}$$

is quite straightforward. Our treatment will be restricted to three-dimensional problems. The cases $p=0$ correspond to generalized charmonium problems.[34] After a factorization of angular terms, the hypervirial relations of Eq. (2.6) become

$$\begin{aligned}2(k+1)E\langle r^k\rangle &= -(2k+1)Z\langle r^{k-1}\rangle + (2k+q+2)\lambda\,\langle r^p\rangle\langle r^{k+q}\rangle \\ &\quad + k[L(L+1) - \tfrac{1}{4}(k+1)(k-1)]\langle r^{k-2}\rangle\,, \\ &\quad k\in\mathbb{Z}\,,\end{aligned} \tag{4.2}$$

where $E = E_{NLM}(\lambda)$ represents the eigenvalue arising from the perturbation of the bound state hydrogenic eigenfunction $\psi_{NLM}^{(0)}$ with eigenvalue $E_{NLM}^{(0)} = -Z^2/(2N^2)$ [N, L, and M will denote the usual hydrogenic quantum numbers, so $L(L+1)$ replaces the scalar L^2 in Eq. (2.3)]. When the expansions

$$\langle r^k\rangle = \sum_{n=0}^{\infty} C_k^{(n)}\lambda^n \tag{4.3}$$

are assumed, the difference equations for the $C_k^{(n)}$ and $E^{(n)}$ become

$$\begin{aligned}2(k+1)E^{(0)}C_k^{(n)} &= -2(k+1)\sum_{j=1}^{n}E^{(j)}C_k^{(n-j)} - (2k+1)ZC_{k-1}^{(n)} \\ &\quad + (2k+1+2)\sum_{j=1}^{n-1}C_p^{(j)}C_{q+k}^{(n-1-j)} \\ &\quad + k[L(L+1) - \tfrac{1}{4}(k+1)(k-1)]C_{k-2}^{(n)}\,,\end{aligned} \tag{4.4}$$

$$E^{(n+1)} = \frac{1}{n+1} \sum_{j=0}^{n} (j+1) C_p^{(j)} C_q^{(n-j)} . \quad (4.5)$$

The calculation of the $C_k^{(n)}$ array proceeds columnwise as for the oscillator problems, with the exception that the row corresponding to $k = -1$ must now be included. At the beginning of each order of calculation n, we use the formula [$k = 0$ in (4.2)]

$$c_{-1}^{(n)} = \frac{1}{Z} \left[-2E^{(n)} + (q+2) \sum_{j=0}^{n-1} c_p^{(j)} c_q^{(n-1-j)} \right].$$

We shall consider in particular the Hamiltonians ($Z = 1$)

$$\hat{H} = \frac{1}{2} \hat{p}_r^2 + \frac{\hat{L}^2}{2r^2} - \frac{1}{r} + \lambda \langle r^p \rangle r , \quad p = 0,1,2 , \quad (4.6)$$

and relate their perturbation expansions with those of the "charmonium" problem,[34]

$$\hat{H} = \frac{1}{2} \hat{p}_r^2 + \frac{\hat{L}^2}{2r^2} - \frac{1}{r} + \lambda r . \quad (4.7)$$

The expansions associated with the "linear" eigenvalue problem (4.7) will be denoted as

$$E_{NLM}(\lambda) = \sum_{n=0}^{\infty} A_{NLM}^{(n)} \lambda^n , \quad (4.8)$$

$$\langle r^k \rangle (\lambda) = \sum_{n=0}^{\infty} a_k^{(n)} \lambda^n , \quad k \geqslant -1 . \quad (4.9)$$

The first three coefficients of the eigenvalue expansion are given by

$$\begin{aligned} A_{NLM}^{(0)} &= -1/2N^2 , \\ A_{NLM}^{(1)} &= \tfrac{3}{2}N^2 - \tfrac{1}{2}L(L+1) , \\ A_{NLM}^{(2)} &= -\tfrac{7}{8}N^6 + \tfrac{3}{8}[L(L+1)]^2 N^2 - \tfrac{5}{8}N^4 . \end{aligned} \quad (4.10)$$

The large-order behavior of the $A_{NLM}^{(n)}$ is given by[35]

$$A_{NLM}^{(n)} \sim (-1)^{n+1} D_{NLM} \Gamma(n+2N) (\tfrac{3}{2}N^3)^n , \quad n \to \infty , \quad (4.11)$$

where

$$D_{NLM} = \frac{3^{2N} 2^{2N-1} e^{-3N + L(L+1)/N}}{\pi N^3 (N+L)! (N-L-1)!} . \quad (4.12)$$

The RS coefficients are negative Stieltjes for $n \geqslant 1$, and the series is Padé summable to $E(\lambda)$ on the cut plane $|\arg \lambda| < \pi$. The coefficients of the S fraction representation to $E(\lambda)$ behave asymptotically as

$$c_n \sim \tfrac{3}{4}N^3 n + K^{(i)} , \quad i = \begin{cases} 1, & n \text{ even}, \\ 2, & n \text{ odd}, \end{cases} \quad (4.13)$$

where

$$K^{(1)} = \tfrac{3}{2}N^4 - \tfrac{1}{2}N^3 , \quad K^{(2)} = \tfrac{3}{2}N^4 - \tfrac{1}{4}N^3 .$$

We now determine the asymptotic behavior of the expansion coefficients $a_k^{(n)}$ in Eq. (4.9), in a manner similar to that used for the oscillator problems. The hypervirial equations to be used for the linear charmonium problem are

$$\begin{aligned} &2(k+1)E \langle r^k \rangle \\ &\quad = -(2k+1)\langle r^{k-1} \rangle + (2k+3)\lambda \langle r^{k+1} \rangle \\ &\qquad + k[L(L+1) - \tfrac{1}{4}(k+1)(k-1)] \langle r^{k-2} \rangle , \\ &\qquad k \geqslant -1 . \end{aligned} \quad (4.14)$$

From the Hellmann–Feynman theorem, $dE/d\lambda = \langle r \rangle$, it follows that

$$a_1^{(n)} \sim (-1)^n D_{NLM} \Gamma(n+N+2) (\tfrac{3}{2}N^3)^{n+1} , \quad n \to \infty . \quad (4.15)$$

This implies that, for $k \geqslant 1$,

$$a_{k+1}^{(n-1)} / a_k^{(n)} \sim [2(k+1)/(2k+3)] \, E^{(0)} , \quad n \to \infty . \quad (4.16)$$

Repeated application of this property leads to the following asymptotic formulas:

$$a_k^{(n)} \sim (-1)^n B_k D_{NLM} \Gamma(n+N+k+1) (\tfrac{3}{2}N^3)^{n+1} , \quad (4.17)$$

where $B_1 = 1$ and

$$B_k = \left(\frac{3}{4} N \right)^{k-1} \prod_{j=1}^{k} \frac{2(j+1)}{2j+3} , \quad k \geqslant 2 . \quad (4.18)$$

This formula has been verified by numerical asymptotic analysis of the $a_k^{(n)}$ for $k = 1,2,...,5$.

Analyticity of the functions $\langle r^k \rangle (\lambda)$, $k \geqslant 1$, in the cut plane $|\arg \lambda| < \pi$ follows from the analyticity properties of $E(\lambda)$ and the recursion relation in (4.14). We also conjecture that their series expansions in (4.9) are Stieltjes. This is based on the numerical evidence that the continued fraction representations of $\langle r^k \rangle$ having the same form as in Eq. (3.24) are S fractions. The coefficients c_{kn} have been computed accurately to order $n = 70$ for $k = 1,2,...,5$. In all cases, the generic asymptotic growth of Eq. (4.13) is observed. The $n!$ growth of the $a_k^{(n)}$ satisfies Carleman's condition which would imply Padé summability on compact subsets of the cut plane $|\arg \lambda| < \pi$. Convergence of $[N-1,N]$ and $[N,N]$ Padé approximants to $\langle r^k \rangle$ for $k = -1,1,2$ were first observed by Austin.[14]

The behavior of the series expansion for $\langle r^{-1} \rangle$ stands apart from those described above. By setting $k = 0$ in Eq. (4.14) and comparing Eqs. (4.11) and (4.15), it follows that $a_{-1}^{(n)} \sim 3nE^{(n)}$, or

$$a_{-1}^{(n)} \sim (-1)^{n+1} 3 D_{NLM} \Gamma(n+N+1) (\tfrac{3}{2}N^3)^n , \quad n \to \infty . \quad (4.19)$$

Since $a_{-1}^{(1)}$ is positive, we consider CF representations of the form

$$\langle r^{-1} \rangle (\lambda) = 1/N^2 + C_{-1}(\lambda) , \quad (4.20)$$

where $C(\lambda)$ has the usual form in Eq. (2.13). All CF representations are observed to be S fractions. Their coefficients behave asymptotically as in Eq. (4.13), in accordance with the large-order behavior in Eq. (4.19).

The analysis of perturbation expansions for the nonlinear problems in (4.1) now proceeds as in Sec. III B. The nonlinear problems define the new coupling constant β,

$$\beta = \lambda \langle r^p \rangle = \sum_{n=0}^{\infty} C_p^{(n)} \lambda^{n+1} \equiv \sum_{n=0}^{\infty} g_n \lambda^{n+1} . \quad (4.21)$$

The formal relations of Eqs. (3.27) and (3.28) apply, and the assumptions similar to Eq. (3.30) of the oscillator problems are made, with the following minor modification:

$$\lim_{n \to \infty} (g_{n+1}/n g_n) = -c g_0 , \quad (4.22)$$

where $c = \tfrac{3}{2}N^3$.

We shall now employ Eq. (3.29) and elucidate the asymptotics for the particular case $p = 1$. The relevant parameters are $g_0 = A^{(1)}$ and $g_1 = 2A^{(1)}A^{(2)}$. This case is seen to be analogous to case 2 of Sec. III B. The net result is [cf. Eq. (3.42)]

$$E^{(n)} \sim \exp[-g_1/3g_0^2]\, nA^{(n)}g_0^n, \quad n\to\infty . \tag{4.23}$$

This may be written as

$$E_{NLM}^{(n)} \sim (-1)^{n+1} D_{NLM} \exp[-\tfrac{2}{3}A_{NLM}^{(2)}/A_{NLM}^{(1)}] \times \Gamma(n+2N+1)[\tfrac{3}{2}N^3 A_{NLM}^{(1)}]^n, \quad n\to\infty . \tag{4.24}$$

V. RENORMALIZED RSPT AND EIGENVALUES OF INFINITE FIELD HAMILTONIANS

An examination of the region of analyticity of eigenvalues $E(\lambda)$ and the nature of their RS perturbation expansions can establish the theoretical summability of these expansions. From a practical aspect, however, RSPT, being a "low-field" expansion, can be relied upon to furnish accurate estimates of $E(\lambda)$ only for small values of the coupling constant λ. Methods for accelerating the convergence of these summability methods may increase this region of λ values by perhaps an order of magnitude. Recently,[36] a "renormalized" perturbation theory has been devised to permit accurate perturbative calculations of $E(\lambda)$ over an infinite range of λ values. It has been applied successfully to the "linear" oscillator and hydrogenic problems and will now be applied to the nonlinear oscillator problems introduced above. The goal is to calculate (1) the eigenvalues $E_K^{(p,q)}(\lambda)$ of the Hamiltonians in Eq. (3.1) accurately over the entire real interval $0\leqslant\lambda<\infty$, and (2) the eigenvalues $F_K^{(p,q)(0)}$ of the "infinite-field" Hamiltonians corresponding to the nonlinear oscillators of above, i.e.,

$$\hat{H}_\infty^{(p,q)} = -\frac{1}{2}\frac{d^2}{dx^2} + \langle x^{2p}\rangle x^{2q} . \tag{5.1}$$

The technical discussions of this method of renormalization, presented in Ref. 30, are omitted. Again it is emphasized that only continued fractions were employed for the following numerical calculations. Padé summability is expected to break down for $q>2$, however. An examination of other summability methods, including Borel, is currently in progress.

The eigenvalues $F_K^{(p,q)(0)}$ of (5.1) are significant for the following reason. A scaling $x\to\alpha^{1/2}x$, where $\alpha = \lambda^{-1/(p+q+1)}$, $\lambda>0$ (representing a unitary transformation), may be applied to the eigenvalue problems associated with (3.1) to give

$$E_K^{(p,q)}(\lambda) \sim F_K^{(p,q)}\lambda^{1/(p+q+1)}, \quad \text{as } \lambda\to\infty . \tag{5.2}$$

[Under this coordinate transformation, $\langle x^{2p}\rangle$ scales as α^p, by its definition in Eq. (1.5).] This is, in fact, but the leading term in the infinite-field expansion

$$E_K^{(p,q)}(\lambda) = \lambda^{1/(p+q+1)} \sum_{k=0}^{\infty} F_K^{(p,q)(k)} \lambda^{-2k/(p+q+1)}, \tag{5.3}$$

which, for $q\geqslant 2$, should be convergent for $|\lambda|>R>0$.[7]

We now outline a method to calculate the $F_K^{(p,q)(0)}$. First, construct the "renormalized" Schrödinger equation [dropping the (p,q) and K indices for notational convenience]:

$$\hat{H}_R(\beta)\psi = \left[-\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}x^2 + \beta\left(\langle x^{2p}\rangle x^{2q} - \frac{1}{2}x^2\right)\right]\psi = G(\beta), \tag{5.4}$$

so that the $G^{(p,q)}(1)$ correspond to the eigenvalues of $H_\infty^{(p,q)}$ in (5.1). Now assume an RS perturbation expansion of the form

$$G(\beta) = \sum_{n=0}^{\infty} G^{(n)}\lambda^n \tag{5.5}$$

for each eigenstate. By scaling the coordinates in (5.4) as $x\to\tau^{1/2}x$, where $0\leqslant\tau^2 = 1-\beta\leqslant 1$, the eigenvalues of (5.4) and (3.1) are related as

$$G_K(\beta) = (1-\beta)^{1/2} E(\beta/(1-\beta)^{(p+q+1)/2}) . \tag{5.6}$$

This relation effectively defines a renormalization map $R:\beta\to\lambda$ which, restricted to the nonnegative real line, maps $\beta\in[0,1)$ onto $\lambda\in[0,\infty)$. By equating the series in (1.6) and (5.5) termwise, and using the general binomial expansion

$$(1-\beta)^{-\alpha} = \sum_{k=0}^{\infty} \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)(k+1)}\beta^k, \tag{5.7}$$

we find the relation

$$G^{(n)} = \sum_{k=0}^{n} \frac{\Gamma(k(p+q-1)/2 + n - \frac{1}{2})}{\Gamma(k(p+q+1)/2 - \frac{1}{2})\Gamma(n-k-1)} E^{(n)} . \tag{5.8}$$

A similar type of renormalization and linear transformation (5.8) occurs in the case of Wick-ordered perturbation series associated with simple field theories.[7,37] The renormalized series coefficients $G^{(n)}$ could, in principle, be calculated from the $E^{(k)}$ by the above relation. Practically speaking, however, the HVHF method is easily applied to the perturbation problem in Eq. (5.4), requiring but a minor modification of the difference equations (3.5) and (3.6). An asymptotic analysis of Eq. (5.8) reveals that $G^{(n)} = O(E^{(n)})$ as $n\to\infty$.

Again, we have not rigorously established the summability of the RS λ series or the renormalized β series for the nonlinear oscillator problems. In Ref. 36, it is shown, using a theorem of Sokal,[38] that if the λ series is Borel summable, then the β series is Borel summable in [0,1].

In Table II are presented estimates of the eigenvalues $F_K^{(p,q)(0)}$ of the Hamiltonians $H_\infty^{(p,q)}$ in Eq. (5.1) for $p = 0,1,2$, $q = 2$ (quartic anharmonic oscillators). These estimates were obtained from the convergents $w_{49}(1)$ and $w_{50}(1)$ to the continued fraction representation of the renormalized β series in Eq. (5.5). A slight "trick" was employed here, as described in Ref. 36, so that the continued fractions were S fractions: for oscillator problems such as (4.4), the perturbation $W(x) = \langle x^{2p}\rangle x^{2q} - \frac{1}{2}x^2$ is not positive definite, since $W(x)<0$ for $x\ll 1$. To overcome this difficulty,

TABLE II. Lower and upper bounds to the eigenvalues $F_K^{(p,2)(0)}$ of the infinite field Hamiltonian $\hat{H}^{(p,2)}$ as defined in Eq. (5.1). The lower and upper bounds are obtained from the convergents $w_{50}(1)$ and $w_{49}(1)$, respectively, of the continued fraction representation of the renormalized β series in Eq. (5.5). The actual numerical entries are the lower bounds. Replacing the final n digits of the entry by the n digits in parentheses gives the upper bound. The entries in parentheses below the $p=0$ row are numerical values calculated by Bell *et al.*[39] and Reid[40] (scaled appropriately due to a different normalization). The estimate $0.489\,05 < F_1^{(1,2)(0)} < 0.489\,07$ has been obtained by Handy[32] using a method of moments.

	K		
p	0	1	2
0	0.667 95(801) (0.667 986)	2.393 47(84) (2.393 644)	4.696 1(76) (4.696 795)
1	0.489 0640(36)	2.200 9(28)	4.67(71)
2	0.494 64(5)	2.27(31)	4.9(5.7)
3	0.534 4(59)	2.35(79)	

we "shift" the perturbation up by an amount $\Delta = \beta g$, $g>0$, to guarantee that $W(x)>0$ for all $x \neq 0$. We thus consider the modified series expansion

$$G(\beta) = G^{(0)} = \beta g + \sum_{n=1}^{\infty} \overline{G}^{(n)} \beta^n, \tag{5.9}$$

where $\overline{G}^{(1)} = G^{(1)} + g$, $\overline{G}^{(k)} = G^{(k)}$ for $k \geqslant 2$. Typically, we have chosen $g = E^{(0)}$. This choice is somewhat arbitrary, but it ensures that the minimum of $W(x)$ is nonnegative. For the "linear" problems, where $p=0$, we can calculate the minimum value of g that guarantees positivity of $W(x)$.[36] In all cases, the CF representation is found to be an S fraction. The upper and lower bounds yielded by the convergents $w_{49}(1)$ and $w_{50}(1)$, respectively, are given in Table II. The eigenvalues $F_K^{(0,2)}$ have been calculated accurately by Bell *et al.*[39] and Reid.[40] These values are included in Table II for reference. (Because of a different scaling of the Hamiltonian, the eigenvalues reported in these papers must be divided by the factor $2^{2/3}$.) Recently, Handy[32] has independently obtained the approximate value $0.490\,45 \leqslant F_0^{(1,2)} \leqslant 0.490\,47$ using a method of moments.

The renormalization relation in (5.6) may also be employed to calculate the eigenvalues $E_K^{(p,q)}(\lambda)$ over the entire range $0 < \lambda < \infty$. We "invert" the scaling transformation $x \to \tau^{1/2} x$, $\tau^2 = 1 - \beta$, used to derive Eq. (5.6), to obtain

$$E(\lambda) = \tau^{-1} G(1 - \tau^2), \tag{5.10}$$

where τ is the root of the equation

$$\lambda \tau^{p+q+1} + \tau^2 - 1 = 0, \tag{5.11}$$

which satisfies $\tau = 1$ when $\lambda = 0$, $\tau \to 0$ as $\lambda \to \infty$. In fact, $\tau \sim \lambda^{-1/(p+q+1)}$ as $\lambda \to \infty$. In order to calculate $E_K^{(p,q)}(\lambda)$, we then (i) calculate the renormalized coefficients $G_K^{(p,q)(k)}$ by HVHF applied to Eq. (5.4); (ii) compute τ from Eq. (5.11) to a prescribed accuracy using the Newton–Raphson method; (iii) "sum" the β series by using Borel, Padé, or other summability techniques; and (iv) compute $E(\lambda)$ from (5.10).

In Table III are presented the estimates of $E_0^{(p,q)}(\lambda)$ for $q=2$, $p=0,1$, a range of λ values. The maximum error in these calculations is expected to be incurred in the high-field limit, i.e., $\beta = 1$. Thus, for a given value of p, the errors in the estimates of $E(\lambda)$ for $0 < \infty$ will be less than for those given in Table II. For $\lambda = 10\,000$, the values $\lambda^{-1/(p+2+1)} E_K^{(p,2)}(\lambda)$ approximate the eigenvalues $F_K^{(p,2)(0)}$ of Table II quite well. The relative behavior of the ground state eigenvalues $E_0^{(p,q)}(\lambda)$ for $0 \leqslant \lambda \leqslant 3$, $p = 0,...,4$, is shown in Fig. 1.

VI. CONCLUDING REMARKS

The model Hamiltonians studied here are very simplified versions of those used to model self-interacting systems. The large-order behavior of Rayleigh–Schrödinger eigenvalue expansions may be related to that of the associated linear eigenvalue problems. This is expected since the nonlinearity of these problems is quite "tame," manifesting itself as a modified coupling constant. The analysis was performed

TABLE III. Estimates of ground-state eigenvalues $E_0^{(p,2)}(\lambda)$ of one-dimensional quartic anharmonic oscillators in Eq. (3.11) for $p=0$ ("linear" Bender–Wu case), 1, 2, and 3. The entries represent the lower bound estimates afforded by the [24,25] (β) Padé approximant to the renormalized β series in Eq. (4.5), with $\beta = 1 - \tau^2$ and τ being the root of Eq. (4.10). Replacing the final n digits in each entry with the n digits in the accompanying parentheses gives the upper bound estimate yielded by the [24,24] (β) Padé.

	p			
λ	0	1	2	3
0.1	0.559 146 327 183519(21)	0.529 717 430 561 641	0.536 324 651 544 448(9)	0.556 718 751 80(6)
0.5	0.696 175 816(25)	0.597 374 410 315(6)	0.600 619 895 9(61)	0.627 289 6(905)
1.0	0.803 770 5(7)	0.649 058 485 4(5)	0.644 396 19(20)	0.670 209(17)
2.0	0.951 567(9)	0.718 266 089(90)	0.699 718 64(72)	0.721 69(73)
3.0	1.060 263(9)	0.768 002 287(92)	0.737 969 56(74)	0.756 09(16)
4.0	1.148 78(9)	0.807 851 82(3)	0.767 944 7(51)	0.782 5(6)
5.0	1.224 578(94)	0.841 539 59(61)	0.792 897 0(5)	0.804 2(4)
10.0	1.504 95(9)	0.963 031 00(6)	0.880 525(7)	0.878 8(91)
20.0	1.865 65(73)	1.113 125 1(3)	0.984 887(91)	0.964 9(55)
50.0	2.499 6(8)	1.363 503 3(6)	1.152 113(20)	1.099 5(95)
100.0	3.131 26(47)	1.599 659 0(6)	1.303 94(5)	1.216 4(78)
1 000.0	6.693 9(44)	2.780 180(82)	2.008 11(4)	1.734 6(77)
10 000.0	14.397(8)	4.907 511(5)	3.145 56(62)	2.511(7)

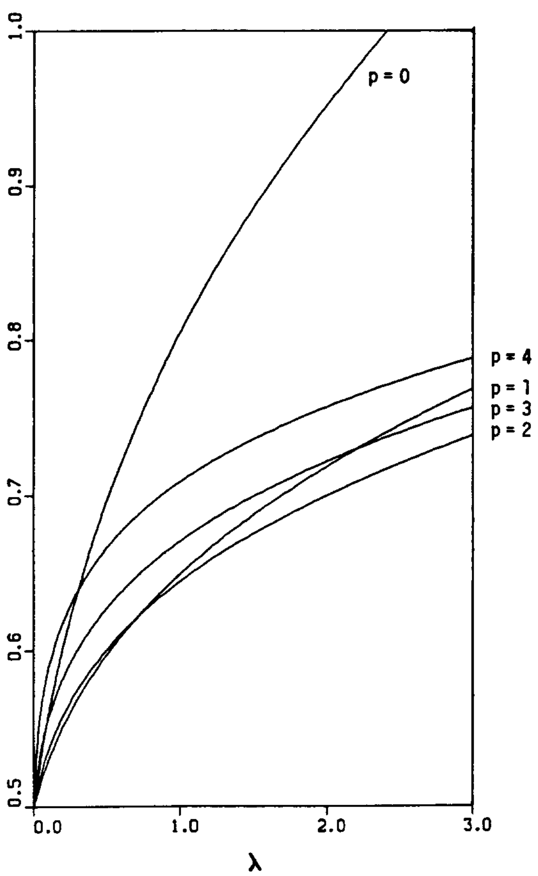


FIG. 1. Ground-state eigenvalues $E_0^{(p,2)}(\lambda)$ of the generalized one-dimensional quartic anharmonic oscillators defined in Eq. (3.11), $p = 0,...,4$.

only for a few specific cases. It is expected, however, that the pattern will appear in general.

The Rayleigh–Schrödinger eigenvalue expansions associated with these radial Hamiltonians are easily calculated via the HVHF method. No wave functions need be calculated, and the only input into the algorithm is the unperturbed energy of the eigenstate in question. For hydrogenic problems, the unperturbed continuum states present difficulties for conventional perturbative treatments, which include the method of Surjan and Angyan.[2] The HVHF method avoids these difficulties. Another method of bypassing these problems is the reformulation of hydrogenic eigenvalue equations by means of a specific realization of the so(4,2) Lie algebra.[41–43] This method is certainly feasible for the hydrogenic problems discussed in Sec. IV. However, the reformulated equation is equivalent to a perturbation problem defined over a nonorthogonal basis set. As such, it would introduce some additional complications into the Surjan–Angyan formulas.

Also, the methods employed here could be applied to more realistic and complicated situations, for example, nonradial perturbations where $\hat{A}$ and $\hat{B}$ in Eq. (1.3) represent the dipole moment operator. A so(4,2) Lie algebraic treatment of such problems would be quite straightforward, being similar to the Zeeman and Stark effects which have already been studied in this way.[42,44] A HVHF approach would involve a coupled system of equations, similar in basic form to that encountered in the Stark effect.[14]

ACKNOWLEDGMENTS

The author wishes to express his thanks to Professor J. Cizek and Professor C. R. Handy for fruitful discussions. He also wishes to thank Professor C. R. Handy and Dr. J. Cioslowski for sending preprints of research prior to their publication.

This research was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada, which is gratefully acknowledged.

[1]S. Yomosa, J. Phys. Soc. Jpn. **35**, 1738 (1972); **36**, 1655 (1974).
[2]P. R. Surjan and J. Angyan, Phys. Rev. A **28**, 45 (1983), and references therein.
[3]J. G. Kirkwood, J. Chem. Phys. **2**, 351 (1934); L. Ansager, J. Am. Chem. Soc. **58**, 1486 (1936).
[4]C. M. Bender and T. T. Wu, (a) Phys. Rev. **184**, 1231 (1969); (b) Phys. Rev. D **7**, 1620 (1973).
[5]C. M. Bender and T. T. Wu, Phys. Rev. Lett. **27**, 461 (1971).
[6]C. M. Bender, J. Math. Phys. **11**, 796 (1970).
[7]B. Simon, Ann. Phys. (NY) **58**, 76 (1970).
[8]J. J. Loeffel, A. Martin, B. Simon, and A. S. Wightman, Phys. Lett. B **30**, 656 (1969).
[9]S. Graffi, V. Grecchi, and B. Simon, Phys. Lett. B **32**, 631 (1970).
[10]R. J. Swenson and S. H. Danforth, J. Chem. Phys. **57**, 1734 (1972).
[11]J. Killingbeck, Phys. Lett. A **65**, 87 (1978).
[12]M. Grant and C. S. Lai, Phys. Rev. A **20**, 718 (1979).
[13]F. M. Fernandez and E. A. Castro, Int. J. Quantum Chem. **19**, 521, 533 (1981).
[14]E. J. Austin, Mol. Phys. **40**, 893 (1980).
[15]E. J. Austin, Mol. Phys. **42**, 1391 (1980).
[16]F. M. Fernandez and E. A. Castro, *Hypervirial Theorems, Lecture Notes in Chemistry*, Vol. 43 (Springer, New York, 1987).
[17]G. Marc and W. G. McMillan, Adv. Chem. Phys. **58**, 209 (1985).
[18]J. D. Louck, J. Mol. Spectrosc. **4**, 334 (1960); J. D. Louck and H. W. Galbraith, Rev. Mod. Phys. **48**, 69 (1976).
[19]J. O. Hirschfelder, J. Chem. Phys. **33**, 1762 (1960); J. O. Hirschfelder and C. A. Coulson, *ibid.* **36**, 941 (1962).
[20]L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968), 3rd ed.
[21]H. Hellmann, *Einfuhrung in die Quantenchemie* (Franz Deuticke, Leipzig, 1937); R. P. Feynman, Phys. Rev. **56**, 340 (1939).
[22]T. Kato, Prog. Theor. Phys. **4**, 514 (1949); **5**, 95, 207 (1950).
[23]A. Erdelyi, *Asymptotic Expansions* (Dover, New York, 1956).
[24]E. Borel, *Lecons sur les Series Divergentes* (Gauthier–Villars, Paris, 1923).
[25]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1978), Vol. 4.
[26]E. R. Vrscay and J. Cizek, J. Math. Phys. **27**, 185 (1986).
[27]W. B. Jones and W. J. Thron, *Continued Fractions, Analytic Theory and Applications* (Addison–Wesley, Reading, MA, 1980).
[28]P. Henrici, *Applied and Computational Complex Analysis* (Wiley, New York, 1977), Vol. 2.
[29]G. Baker, Jr., *Essentials of Padé Approximants* (Academic, New York, 1975).
[30]B. W. Char, K. O. Geddes, G. H. Gonnet, and S. M. Watt, *MAPLE User's Guide* (Watcom, Waterloo, 1985), 4th ed.
[31]J. Cioslowski, Phys. Rev. A **36**, 374 (1987).
[32]C. R. Handy, Phys. Lett. (to appear).
[33]H. S. Wall, *Analytic Theory of Continued Fractions* (Van Nostrand, Toronto, 1948).
[34]E. R. Vrscay, Phys. Rev. Lett. **53**, 2521 (1984).
[35]E. R. Vrscay, Phys. Rev. A **31**, 2054 (1985).
[36]E. R. Vrscay, Theor. Chim. Acta (to appear).
[37]W. Caswell, Ann. Phys. **123**, 153 (1979).
[38]A. Sokal, J. Math. Phys. **21**, 261 (1980).

[39]S. Bell, R. Davidson, and P. A. Warsop, J. Phys. B 3, 113 (1970).
[40]C. Reid, J. Mol. Spectrosc. 36, 183 (1970).
[41]M. Bednar, Ann. Phys. (NY) 75, 305 (1973).
[42]J. Cizek and E. R. Vrscay, Int. J. Quantum Chem. 21, 27 (1982).
[43]B. G. Adams, J. Cizek, and J. Paldus, Int. J. Quantum Chem. 21, 153 (1982).
[44]H. J. Silverstone, B. G. Adams, J. Cizek, and P. Otto, Phys. Rev. Lett. 43, 1498 (1979).

Comments on the paper "On the WKBJ approximation" [J. Math. Phys. 28, 556 (1987)]

Nanny Fröman and Per Olof Fröman
Institute of Theoretical Physics, University of Uppsala, Thunbergsvägen 3, S-752 38 Uppsala, Sweden

(Received 7 August 1987; accepted for publication 18 November 1987)

Attention is drawn to the fact that the "standard form for the generalized WKBJ approximation" of El Sawi [J. Math. Phys. **28**, 556 (1987)] already had been derived by N. Fröman [Ark. Fys. **32**, 541 (1966)].

The aim of this paper is to draw attention to the fact that the higher-order approximation, which is called "the standard form (SF) for the generalized WKBJ approximation" in El Sawi's paper "On the WKBJ approximation,"[1] was derived by N. Fröman[2] already in 1966. The improved form derived by El Sawi[1] is given by his Eq. (3.7) and corresponds to N. Fröman's[2] Eqs. (7a) and (7b). Considering the two lowest orders of approximation, El Sawi obtains his Eqs. (3.14), (3.13), and (2.8). Noting that the function T_3 in his Eq. (2.8) can be written in the simpler form

$$T_3 = \int_a^x \frac{1}{2} f^{-1/4} \frac{d^2}{dt^2} f^{-1/4}\, dt,$$

one can write his Eq. (3.13) as

$$\phi_2 = \int_a^x \left(1 + \frac{1}{2} f^{-3/4} \frac{d^2}{dt^2} f^{-1/4}\right) f^{1/2}\, dt.$$

With this expression for ϕ_2, El Sawi's Eq. (3.14) is seen to agree with the next lowest order of approximation in Ref. 2; see Eqs. (1), (7a)–(7c), (8c), (9a), and (9b) in that paper. The correction giving the next order of approximation is given by Eq. (9c) in Ref. 2. A recurrence formula for obtaining the approximation to any order directly is given by Eq. (8a) in Ref. 2, and on the basis of this formula Campbell[3] calculated explicit expressions for the ten lowest orders of approximation, the first three of which agreeing with those previously given by N. Fröman.[2] The great advantages of using the above approximation instead of the WKBJ approximation in higher orders was documented in a paper by Dammert and P. O. Fröman.[4]

A more general approximation, now called phase-integral approximation, generated from an *a priori* unspecified base function (and containing the approximation in Ref. 2 and hence also that in Ref. 1, as a special case) was derived by N. Fröman and P. O. Fröman.[5,6] Combined with the rigorous method for solving connection problems published in Ref. 7, this approximation forms part of the phase-integral method developed by the present authors.

[1] M. El Sawi, J. Math. Phys. **28**, 556 (1987).
[2] N. Fröman, Ark. Fys. **32**, 541 (1966).
[3] J. A. Campbell, J. Comput. Phys. **10**, 308 (1972).
[4] Ö. Dammert and P. O. Fröman, J. Math. Phys. **21**, 1683 (1980).
[5] N. Fröman and P. O. Fröman, Ann. Phys. (NY) **83**, 103 (1974).
[6] N. Fröman and P. O. Fröman, Nuovo Cimento B **20**, 121 (1974).
[7] N. Fröman and P. O. Fröman, *JWKB Approximation, Contributions to the Theory* (North-Holland, Amsterdam, 1965) (Russian translation: MIR, Moscow, 1967).

 0022-2488/88/040912-01$02.50

Canonical transformations in quantum mechanics: A canonically invariant path integral

A. Y. Shiekh
Blackett Laboratory, Imperial College, London SW7 2BZ, England and the International Centre for Theoretical Physics, Miramare, Trieste, Italy

(Received 3 April 1987; accepted for publication 16 September 1987)

A particular form of the path integral is presented that allows the implementation of general canonical transformations in quantum mechanics, as well as a consistent quantization procedure.

I. PATH INTEGRAL APPROACH TO QUANTUM MECHANICS

Classical mechanics is a description of nature using commuting variables that is well formulated by the Hamilton least action principle. But classical mechanics is not a complete model of nature, leaving several features unexplained. Some of the discrepancies in the classical theory have been overcome in the quantum generalization.

Dirac[1,2] proposed how a quantum theory might be induced from the classical theory. He postulated that the classical commuting variables become noncommuting operators and suggested how the quantum dynamics be obtained from the classical Hamilton equations of motion. Since the quantum theory is the more fundamental theory (the classical predictions following in the $\hbar \to 0$ limit), it might be argued that this should be the starting point. Dirac's scheme, however, has the advantage of starting from a well understood theory. In this scheme, to each classical system correspond many quantum generalizations (each yielding the same classical predictions for $\hbar \to 0$). Dirac's method, however, is ambiguous[3] in not generating one unique member of the quantum generalizations. Some further specification (such as normal ordering) is required to completely specify the quantum theory. This might not seem a disadvantage; but due to this ambiguity, classical techniques such as the use of canonical transformations cannot be used to directly induce a quantum counterpart. This results in the loss of powerful techniques such as the Hamilton–Jacobi approach so often employed in classical mechanics.[4]

This work is a preliminary investigation of the alternative path integral quantization technique of Feynman,[5,6] in an attempt to overcome some of these problems. The path integral technique, although equivalent, differs significantly from the usual operator formulation of quantum mechanics, in that it employs commuting (or badly called "classical") variables. Although these variables commute, the theory being equivalent to the operator formalism must contain "operator ordering" within its structure. Understanding just how this occurs is crucial in the use of the path integral and is reviewed below.

II. DERIVING THE PATH INTEGRAL[7]

Although no formalism of quantum mechanics is more fundamental than any other, each is supposed to be equivalent and so one should be derivable from another. Assuming a knowledge of traditional quantum mechanics one may deduce the path integral formalism. Starting from the position to position amplitude for Heisenberg eigenstates,

$$\langle q_b,t_b|q_a,t_a\rangle.$$

(This may be generalized to the transition between any two states.) Inserting position resolutions of unity:

$$1=\int_{-\infty}^{\infty} dq|q\rangle\langle q|$$

leads to

$$\langle q_b,t_b|q_a,t_a\rangle=\int_{-\infty}^{\infty}\cdots\prod_{j=1}^{N-1}dq(j)\times\prod_{k=1}^{N}\langle q(k),k|q(k-1),k-1\rangle.$$

Recall that for Heisenberg eigenstates

$$|q,t+\Delta t\rangle=\exp[(i/\hbar)\hat{\mathrm{H}}(\hat q,\hat p,t+\Delta t/2)\Delta t]|q,t\rangle,$$

$\hat{\mathrm{H}}$ being a Hermitian operator. The use of midpoint time is crucial for this evolution to be accurate to order Δt, as it must. So

$$\langle q_b,t_b|q_a,t_a\rangle=\lim_{N\to\infty}\int_{-\infty}^{\infty}\cdots\prod_{j=1}^{N-1}dq(j)\prod_{k=1}^{N}\langle q(k)|\times\exp\left[-\frac{i}{\hbar}\hat{\mathrm{H}}\left(\hat q,\hat p,k-\frac{1}{2}\right)\Delta t\right]|q(k-1)\rangle,$$

where

$$\Delta t\equiv(t_b-t_a)/N.$$

Proceeding by expanding the exponential leads to

$$\langle q_b,t_b|q_a,t_a\rangle=\lim_{N\to\infty}\int_{-\infty}^{\infty}\cdots\prod_{j=1}^{N-1}dq(j)\prod_{k=1}^{N}\langle q(k)|\times\sum_{m=0}^{\infty}\frac{(-(i/\hbar)\hat{\mathrm{H}}(\hat q,\hat p,k-\frac{1}{2})\Delta t)^m}{m!}|q(k-1)\rangle.$$

Further inserting momentum resolutions of unity

$$1=\int_{-\infty}^{\infty}dp|p\rangle\langle p|.$$

For the $m=1$ case this may be done in one of two ways, leading to alternative integrands, namely,

 0022-2488/88/040913-18$02.50

$$\langle q(k)|(-(i/\hbar)\hat{\mathbf{H}}(\hat{q},\hat{p},k-\tfrac{1}{2})\Delta t)|p(k)\rangle\langle p(k)|q(k-1)\rangle$$

or

$$\langle q(k)|p(k)\rangle\langle p(k)|(-(i/\hbar)\hat{\mathbf{H}}(\hat{q},\hat{p},k-\tfrac{1}{2})\Delta t)|q(k-1)\rangle$$

which equals

$$\langle q(k)|p(k)\rangle\langle p(k)|q(k-1)\rangle \times(-(i/\hbar)\mathbf{H}(q(k-1),p(k),k-\tfrac{1}{2})\Delta t),$$

where

$$\mathbf{H}(q,p,t)\equiv\langle p|\hat{\mathbf{H}}(\hat{q},\hat{p},t)|q\rangle/\langle p|q\rangle.$$

To evaluate this one should commute factors in the Hamiltonian operator (using $[\hat{q},\hat{p}]=i\hbar\hat{\mathbf{1}}$), such that $\hat{q}$ operators are shifted to the right and can be applied to the position eigenstates, while $\hat{p}$ operators (now on the left) apply to their eigenstates. This sifting induces additional terms that carry the operator ordering information for the path integral.

The $m=2$ case has three such alternatives; namely,

$$\tfrac{1}{2}\langle q(k)|(-(i/\hbar)\hat{\mathbf{H}}(\hat{q},\hat{p},k-\tfrac{1}{2})\Delta t)^2|p(k)\rangle \times\langle p(k)|q(k-1)\rangle$$

$$\times\tfrac{1}{2}\langle q(k)|(-(i/\hbar)\hat{\mathbf{H}}(\hat{q},\hat{p},k-\tfrac{1}{2})\Delta t)|p(k)\rangle \times\langle p(k)|(-(i/\hbar)\hat{\mathbf{H}}(\hat{q},\hat{p},k-\tfrac{1}{2})\Delta t)|q(k-1)\rangle$$

$$\times\tfrac{1}{2}\langle q(k)|p(k)\rangle \times\langle p(k)|(-(i/\hbar)\hat{\mathbf{H}}(\hat{q},\hat{p},k-\tfrac{1}{2})\Delta t)^2|q(k-1)\rangle.$$

This implies (using the $m=1$ results) that these are each, to order Δt, equivalent to

$$\tfrac{1}{2}\langle q(k)|p(k)\rangle\langle p(k)|q(k-1)\rangle \times(-(i/\hbar)\mathbf{H}(q(k-1),p(k),k-\tfrac{1}{2})\Delta t)^2.$$

This result generalizes, by induction, to finite m, leading to

$$\langle q_b,t_b|q_a,t_a\rangle =\lim_{N\to\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\cdots\prod_{j=1}^{N-1}dq(j)\prod_{i=1}^{N}dp(i) \times\prod_{k=1}^{N}\langle q(k)|p(k)\rangle\langle p(k)|q(k-1)\rangle \times\exp[-(i/\hbar)\mathbf{H}(q(k-1),p(k),k-\tfrac{1}{2})\Delta t],$$

where

$$\mathbf{H}(q,p,t)\equiv\langle p|\hat{\mathbf{H}}(\hat{q},\hat{p},t)|q\rangle/\langle p|q\rangle.$$

Now recall

$$\langle q|p\rangle=e^{iqp/\hbar}/2\pi\hbar.$$

Substituting these results leads to

$$\langle q_b,t_b|q_a,t_a\rangle =\lim_{N\to\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\cdots\prod_{j=1}^{N-1}dq(j)\prod_{i=1}^{N}\frac{dp(i)}{2\pi\hbar} \times\exp\left[\frac{i}{\hbar}\sum_{k=1}^{N}\{p(k)(q(k)-q(k-1)) -H(q(k-1),p(k),k-\tfrac{1}{2})\Delta t\}\right],$$

$$\Delta t=(t_b-t_a)/N,\quad q(0)=q_a,\quad q(N)=q_b$$

[the phase space (Hamilton) path integral], where

$$\mathbf{H}(q,p,t)=\langle p|\hat{\mathbf{H}}(\hat{q},\hat{p},t)|q\rangle/\langle p|q\rangle,$$

which may be *formally* written as

$$\langle q_b,t_b|q_a,t_a\rangle=\iint Dq\,Dp\,\exp\left[\frac{i}{\hbar}\mathbf{S}(q,p)\right],$$

$$q(t_a)=q_a,\quad q(t_b)=q_b,$$

$$\mathbf{S}(q,p)=\int_{t_a}^{t_b}dt\,(p\dot{q}-\mathbf{H}(q,p,t)).$$

This is formal due to the fact that this expression depends upon which finite difference scheme is adopted in the discretization. The reasoning behind this will become clear later.

Why this object is referred to as a "path integral" or "sum over histories" follows from considering the points $q(j)$, $p(j)$ connected by lines. Then we have a broken line path from q_a to q_b; the sum in the exponent being the action of classical mechanics. But unlike classical mechanics the least action path is not, *a priori*, preferred over any other. Each path is equally considered and carries a phase weighting. The amplitude contribution from each and every path is then summed to yield the total end amplitude. A heuristic but very appealing argument for the classical limit is that paths around the classical contribute amplitudes that are in phase (since the classical path is that with an extremum action); while those far from it contribute largely differing phases and so tend to cancel each other out. In this way contributions from around the classical path are favored and classical mechanics recovered in the $\hbar\to 0$ limit. These arguments can be made more precise.

In general, the result of performing the phase space integrals depends on the sequence in which they are performed. It is understood that if there is any ambiguity, the momentum integrals are to be performed first. How this comes about, and its cure, is discussed elsewhere.[8]

III. STOCHASTIC TERMS

Beginning from the Hamiltonian path integral,

$$\lim_{N\to\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\cdots\prod_{j=1}^{N-1}dq(j)\prod_{i=1}^{N}\frac{dp(i)}{2\pi\hbar}\exp\left[i\sum_{k=1}^{N}\frac{\{p(k)(q(k)-q(k-1))-\mathbf{H}(q(k-1),p(k),k-\tfrac{1}{2})\Delta t\}}{\hbar}\right],$$

$$\Delta t=T/N.$$

This consists of many integrals of the form

$$I \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dq\, dp \exp\left[i \frac{\{p\Delta q - \mathbf{H}(q,p,t)\Delta t\}}{\hbar} \right].$$

Taylor expand $\mathbf{H}$ about $q,p = 0$ to yield

$$I = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dq\, dp$$
$$\times \exp\left[i \frac{\{p\Delta q - p^2\Delta t/2m(q,t)\}}{\hbar} \right]$$
$$\times (C_{00}(t) + C_{10}(t)q + C_{01}(t)p + \cdots),$$

further, temporarily expand $\exp[ip\Delta q/\hbar]$ to yield

$$I = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dq\, dp$$
$$\times \exp\left[\frac{-ip^2\Delta t}{2\hbar m(q,t)} \right]$$
$$\times (D_{00}(t) + D_{20}(t)q^2 + D_{02}(t)p^2 + \cdots),$$

having held only p^2 in the exponential. The p integrals may then be performed using

$$\int_{-\infty}^{\infty} \exp[-\alpha s^2] ds = \sqrt{\frac{\pi}{\alpha}},$$
$$\int_{-\infty}^{\infty} s^2 \exp[-\alpha s^2] ds = \int_{-\infty}^{\infty} \frac{1}{2\alpha} \exp[-\alpha s^2] ds,$$
$$\int_{-\infty}^{\infty} s^4 \exp[-\alpha s^2] ds = \int_{-\infty}^{\infty} \frac{3}{4\alpha^2} \exp[-\alpha s^2] ds,$$

etc. (odd integrals disappearing),

which shows that each p contributes like $(\Delta t)^{-1/2}$, since each p^2 generates a $1/\alpha$, where $\alpha = i\Delta t/2m\hbar$.

In performing the p integrals in

$$I = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dq\, dp \exp\left[i \frac{\{p\Delta q - p^2\Delta t/2m(q,t)\}}{\hbar} \right]$$
$$\times (C_{00}(t) + C_{10}(t)q + C_{01}(t)p + \cdots)$$

and obtaining the Lagrange formalism; p becomes $m\Delta q/\Delta t$, so that in the Lagrange formalism $\Delta q \sim (\Delta t)^{1/2}$ [cf. $p \sim (\Delta t)^{-1/2}$ in the Hamiltonian formalism]. It is in this way that the contributing class of paths are seen to be stochastic (or Brownian) in nature. This behavior must be carefully taken into account when working to order Δt (as before), and is the manifestation of the path integrals sensitivity to the finite difference scheme adopted in discretization. Terms, such as $(\Delta q)^4/\Delta t$ in the Lagrangian, give (in the $\Delta t \to 0$ limit), for paths smooth in p and q, no contribution. This is because finite p implies $\Delta q \sim \Delta t$ and so $(\Delta q)^4/\Delta t \sim (\Delta t)^3$. But such terms are finitely contributing for the dominating unsmooth paths ($\Delta q \sim (\Delta t)^{1/2}$) and will be referred to as stochastic. This dependence on stochastic paths is where operator ordering is concealed. It is here that it is seen that a Hamiltonian can contribute like $(\Delta t)^{-1}$, since it behaves like $p^2/2m$ for small p; and that this is the strongest behavior that is not divergent when the $\Delta t \to 0$ limit is taken.

The midpoint rule (being accurate to second order) generates no stochastic terms and is the reason why it correctly reproduces the quantum mechanics. This is seen when looking at the *usual* definition of the differential,

$$\frac{df(t)}{dt} \equiv \lim_{\Delta t \to 0} \frac{f(t) - f(t - \Delta t)}{\Delta t}.$$

Higher order differentials follow from repeated application,

$$= \lim_{\Delta t \to 0} \left(\frac{df(t)}{dt} - \frac{d^2 f(t)}{dt^2} \frac{\Delta t}{2!} + \cdots \right)$$

[from Taylor expansion (f being assumed analytic).

The "strongest" terms in the action (such as that stemming from $p\dot{q}$) are of order $(\Delta t)^0$, stronger terms not occurring physically as they lead to infinities in the $\Delta t \to 0$ limit. Time derivatives appearing in such terms should be represented accurately to order Δt, since we must work to this order. In this way the formula given above for the derivative is seen to be an inconsistent definition if used in a formalism, such as path integration, that is sensitive to first order in Δt. Alternatively using a symmetric (midpoint) definition

$$\lim_{\Delta t \to 0} (f(t + \Delta t) - f(t - \Delta t))/2\Delta t$$
$$= \lim_{\Delta t \to 0} \left(\frac{df(t)}{dt} + \frac{d^3 f(t)}{dt^3} \frac{(\Delta t)^2}{3!} + \cdots \right).$$

This has all the error terms disappearing for the path integral in the $\Delta t \to 0$ limit, and so is a correct definition for use within the path integral. Other schemes might, in context, give no contribution, but the midpoint scheme is guaranteed not to. It might now be argued that the formal expression

$$\langle q_b,t_b | q_a,t_a \rangle = \int \int Dq\, Dp \exp\left[\frac{i S(q,p)}{\hbar} \right],$$
$$q(t_a) = q_a, \quad q(t_b) = q_b,$$
$$S(q,p) = \int_{t_a}^{t_b} dt (p\dot{q} - \mathbf{H}(q,p,t)),$$

can be made unambiguous if it is noted that all consistent discretization schemes lead to no stochastic terms and so the same answer in the $\Delta t \to 0$ limit. As was seen earlier, this use of midpoint expansion was not always necessary in Cartesian coordinates; but becomes essential when using a general coordinate system and is a matter approached in more detail later.

Before continuing, the knowledge gained so far may be used to perform the momentum integration in general. Starting from the Hamiltonian path integral,

$$\langle q_b,t_b | q_a,t_a \rangle$$
$$= \lim_{\Delta t \to 0} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \cdots \prod_{j=1}^{N-1} dq(j) \prod_{i=1}^{N} \frac{dp(i)}{2\pi\hbar}$$
$$\times \exp\left[\frac{i}{\hbar} \sum_{k=1}^{N} \{ p(k)(q(k) - q(k-1)) \right.$$
$$\left. - \mathbf{H}(q(k-1), p(k), k - \tfrac{1}{2})\Delta t \} \right]$$

and looking at one such integral,

$$I = \lim_{\Delta t \to 0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dq\, dp$$
$$\times \exp[(i/\hbar) \{ p\Delta q - \mathbf{H}(q,p,t)\Delta t \}].$$

Taylor expand $\mathbf{H}$ about $q,p = 0$ to yield

$$I = \lim_{\Delta t \to 0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dq\, dp$$
$$\times \exp[(i/\hbar)\{p\Delta q - (p^2/2m(q,t) + \alpha(q,t)p + V(q,t))\Delta t\}],$$

where

$$\frac{1}{m} = \frac{\partial^2 \mathbf{H}}{\partial p^2}\bigg|_{p=0}, \quad \alpha = \frac{\partial \mathbf{H}}{\partial p}\bigg|_{p=0}, \quad V = \mathbf{H}|_{p=0},$$

$p^2\Delta t$ being the strongest "physical term" allowed [recall $p \sim (\Delta t)^{-1/2}$] and $V(q,t)$ the weakest that contributes in the $\Delta t \to 0$ limit. Surprisingly the p integration has become Gaussian and may be evaluated using

$$\int_{-\infty}^{\infty} \exp[F(p)]dp = \sqrt{\frac{\pi}{\alpha}} \exp[F_0],$$

where

$$F(p) = -\alpha p^2 + \beta p + \chi$$

and F_0 is the minimum of $F(p)$ w.r.t. p, i.e.,

$$F_0 = \chi + \beta^2/4\alpha.$$

This means that part of the least action principle of classical mechanics, namely,

$$\frac{\delta S}{\delta p} = 0$$

continues to be valid quantum mechanically.

This leads to the general configuration space (Lagrange) path integral,

$$\langle q_b,t_b | q_a,t_a \rangle = \lim_{\Delta t \to 0} (2\pi\hbar i \Delta t)^{-N/2} \int_{-\infty}^{\infty} \cdots \prod_{j=1}^{N-1} dq(j)$$
$$\times \exp\left[\frac{i}{\hbar} \Delta t \left\{ \sum_{k=1}^{N} \frac{\ln(m(q(k-1),k-\frac{1}{2}))}{2i\Delta t} + m\left(q(k-1),k-\frac{1}{2}\right)\left(\frac{(q(k)-q(k-1))}{\Delta t} - \frac{1}{2}\alpha\left(q(k-1),k-\frac{1}{2}\right)\right)^2 - V\left(q(k-1),k-\frac{1}{2}\right)\right\}\right]$$

[the configuration space (Lagrange) path integral], where

$$\frac{1}{m} = \frac{\partial^2 \mathbf{H}}{\partial p^2}\bigg|_{p=0}, \quad \alpha = \frac{\partial \mathbf{H}}{\partial p}\bigg|_{p=0}, \quad V = \mathbf{H}|_{p=0}.$$

See Abers and Lee[9] for a discussion of the log term.

IV. CANONICAL TRANSFORMATIONS IN QUANTUM MECHANICS

In this preliminary work canonical transformations for the path integral[10] are identified to be a subset of those of classical mechanics, and a time discretization scheme is found that allows the transformation to a trivial Hamiltonian, as well as a consistent quantization of a classical system.

Traditional quantization through operators[2,3] does not generate a unique quantum theory. Equivalent classical systems (related by a canonical transformation) in general yield differing quantum systems.[11] Also, in the operator formalism, it is not clear how to implement a canonical transformation due to the use of noncommuting variables. The ambiguities in quantizing the classical theory means that one cannot fall back on the classical theory to perform the canonical transformation. The path integral description of quantum mechanics[5,6] offers an alternative method of quantization and poses a possible way out of this dilemma, since it uses commuting variables in its structure.

Starting from the classical action, one can form the path integral expression

$$\int\int \exp\left(i \int \frac{(p\dot{q} - \mathbf{H})dt}{\hbar}\right) Dq\, Dp.$$

This formal expression is deceptive in that it employs commuting variables, but is supposed to be equivalent to the traditional operator formalism. The above formal expression is in fact ill defined. In order to evaluate it one can discretize it in time; but the answer is in fact dependent upon the finite difference scheme adopted. Factor ordering is carried within the prescription.[12–14] In general, however, the prescription will change under a canonical transformation,[15] so a quantization scheme based on a particular prescription will in general generate inequivalent quantum systems from equivalent classical ones. These features have been dealt with in more detail in the Introduction. However, a particular discretization scheme has been found that is invariant under general canonical transformations, and opens the way to a consistent quantization scheme, as well as a quantum mechanical application of the Hamilton–Jacobi theory of classical mechanics.[4,12]

V. CANONICAL TRANSFORMATIONS IN THE PATH INTEGRAL

In classical mechanics a canonical transformation is one that preserves the least action principle.[4] For the path integral one might analogously require that there be a path integral representation in the new variables (Q,P,t), if one existed in the old ones (q,p,t). Such a transformation should be system independent, that is to say, the transformation should be canonical not only for some specific system, but for all problems with the same degrees of freedom. The amplitude may alter under such a transformation by at most a phase factor, i.e., *formally*, with end points (a,b) in phase space held fixed,

$$\int\int \exp\left(i \int \frac{(p\dot{q} - \mathbf{H})dt}{\hbar}\right) Dq\, Dp = \exp\left(\frac{i(F_b - F_a)}{\hbar}\right) \times \int\int \exp\left(i \int \frac{(P\dot{Q} - \mathbf{K})dt}{\hbar}\right) DQ\, DP, \quad \forall\, \mathbf{H}(q,p,t),$$

with F being an arbitrary smooth function. Assuming that a canonically invariant discretization prescription exists (just such a scheme being sought), that is to say this formal statement becomes true for that scheme. Any other expression should be manipulated into this form with the resulting $O(\hbar^2)$ term additions to the Hamiltonian. These terms may be replaced by "potential-like" terms of the same effect,[16] the technique for achieving this being illustrated later. It is being claimed that the quantum canonical transformation is a cleaner object when used with an invariant path integral scheme.

Since the above equation is to be true for all Hamiltonians, the integrands must be equal. This is perhaps most easily seen by choosing Hamiltonians that are highly localized in phase space. The integrands must then be equal at the "localization point." By choosing Hamiltonians localized at each point, it follows that the integrands must be equal everywhere. This implies that

$$p\dot{q} - \mathbf{H} = P\dot{Q} - \mathbf{K} + \frac{dF}{dt}$$

if the end points in phase space are fixed, i.e., we should work with a coherent state type path integral.[15] This is the same requirement as in classical mechanics,[4] as well as the condition that the Jacobian of the transformation be unity (which follows from above[4]).

Suppose $F = F(q,Q,t)$; then because

$$p\dot{q} - \mathbf{H} = P\dot{Q} - \mathbf{K} + \left(\frac{\partial F}{\partial q}\right)_{Qt}\dot{q} + \left(\frac{\partial F}{\partial Q}\right)_{qt}\dot{Q} + \left(\frac{\partial F}{\partial t}\right)_{qQ},$$

and by the independence of q and Q,

$$p = \left(\frac{\partial F}{\partial q}\right)_{Qt}, \quad P = -\left(\frac{\partial F}{\partial Q}\right)_{qt}, \quad \mathbf{K} = \mathbf{H} + \left(\frac{\partial F}{\partial t}\right)_{qQ},$$

with F being now seen to be the generating function of the canonical transformation. One concludes from this that the quantum canonical transformations are the same as those for classical mechanics, excepting that scaling transformations are excluded. It is possible that the momenta [defined by $p \equiv (\partial \mathbf{L}(q,\dot{q})/\partial\dot{q})_q$] are not all independent of the coordinates. Independence, and so a Hamiltonian description, can be achieved by employing the constraint analysis of Dirac, where the constraints are moved into the action using Lagrange multipliers.[8,17]

If one could perform general canonical transformations in quantum mechanics, then one might consider emulating the Hamilton–Jacobi philosophy of classical mechanics. In this approach,[4,12] rather than directly solve the equations of motion following from a given Hamiltonian $\mathbf{H}(q,p,t)$, a canonical transformation is implemented that renders the transformed Hamiltonian [Kamiltonian $\mathbf{K}(Q,P,t)$] equal to zero. The work then lies in finding this transformation, the generator of which is determined by the Hamilton–Jacobi equation. For this purpose it is convenient to work with an alternative generating function given by

$$\mathscr{F}(q,P,t) \equiv F(q,Q,t) + QP,$$

so

$$\begin{aligned} p\dot{q} - \mathbf{H} &= P\dot{Q} - \mathbf{K} + \frac{dF}{dt} \\ &= -Q\dot{P} - \mathbf{K} + \left(\frac{\partial \mathscr{F}}{\partial q}\right)_{Pt}\dot{q} \\ &\quad + \left(\frac{\partial \mathscr{F}}{\partial P}\right)_{qt}\dot{P} + \left(\frac{\partial \mathscr{F}}{\partial t}\right)_{qP}, \end{aligned}$$

from which follows, by the independence of q and P,

$$Q = \left(\frac{\partial \mathscr{F}}{\partial P}\right)_{qt}, \quad p = \left(\frac{\partial \mathscr{F}}{\partial q}\right)_{Pt}, \quad \mathbf{K} = \mathbf{H} + \left(\frac{\partial \mathscr{F}}{\partial t}\right)_{qP},$$

leading to the Hamilton–Jacobi equation

$$\mathbf{H}\left(q,\left(\frac{\partial \mathscr{F}}{\partial q}\right)_{Pt},t\right) + \left(\frac{\partial \mathscr{F}}{\partial t}\right)_{qP} = 0.$$

Classically one has transformed into a frame that "tracks" the system so that it then has trivial motion (constant phase space position). The transformation then carries the motion.

VI. THE SYMMETRIC PATH INTEGRAL

Due to the higher order sensitivities of the path integral,[12] in order to correctly transform the path integral one can start from some time discretized version. An especially convenient scheme for the propagator is the symmetric prescription given by

$$K(q_b,p_b,t_b|q_a,p_a,t_a) \equiv \lim_{N\to\infty} \int \cdots \int \prod_{j=1}^{N-1} \left(\frac{dq(j)dp_m(j)}{2\pi\hbar}\right) \times \exp\left[i\sum_{k=1}^{N} \frac{\{p_m(k)(q(k)-q(k-1)) - \mathbf{H}(q_m(k),p_m,k-\frac{1}{2})\Delta t}{\hbar}\right],$$

where

$$q_m(k) \equiv \tfrac{1}{2}(q(k) + q(k-1)),$$

$$p_m(k) \equiv \tfrac{1}{2}(p(k) + p(k-1)), \quad \Delta t \equiv (t_b - t_a)/N.$$

The end points not being integrated over, allowing them to be held constant. This is most closely related to the coherent state path integral[15] where end points are naturally not integrated over and there are an equal number of coordinate and conjugate momenta integrations. In general a given phase space path integral (with fixed end points) must be kneaded into this form with the corresponding $\hbar^2$ additions to the Hamiltonian, which is then highly quantum mechanical in nature. What has been achieved is an "exposure" of all stochastic terms. The use of the symmetric (midpoint) finite difference expressions was discussed and motivated in the introduction to path integrals. By considering p_m as a dummy variable, this midpoint ordering can be seen to correspond to "Weyl ordering" in the operator formalism,[18] but it must be borne in mind that this object, like the coherent state path integral, is not a physical amplitude. The properties of midpoint investigated earlier suggest that this proposed path integral might be the canonical invariant sought.

This manner of path integral has been investigated by Klauder[15] and would seem to be the starting point for making the path integral a well defined mathematical object.[19]

VII. TRANSFORMING THE PATH INTEGRAL

Consider a general canonical transformation of the p and q implemented by a generating function $F(q,Q,t)$,

$$p=\left(\frac{\partial F}{\partial q}\right)_{Qt}, \quad P=-\left(\frac{\partial F}{\partial Q}\right)_{qt}, \quad \mathbf{K}=\mathbf{H}+\left(\frac{\partial F}{\partial t}\right)_{qQ}$$

(we could use any other type of generating function[4]) or equivalently,

$$q=q(Q,P,t), \quad Q=Q(q,p,t),$$
$$p=p(Q,P,t), \quad P=P(q,p,t),$$

such that

$$\left(\frac{\partial Q}{\partial q}\right)_{pt}=\left(\frac{\partial p}{\partial P}\right)_{Qt}, \quad \left(\frac{\partial Q}{\partial p}\right)_{qt}=-\left(\frac{\partial q}{\partial P}\right)_{Qt},$$
$$\left(\frac{\partial P}{\partial q}\right)_{pt}=-\left(\frac{\partial p}{\partial Q}\right)_{Pt}, \quad \left(\frac{\partial P}{\partial p}\right)_{qt}=\left(\frac{\partial q}{\partial Q}\right)_{Pt}.$$

It is useful to first consider the *formal* canonical transformation of the path integral with no consideration of stochastic terms.

Formally, with all end points fixed,

$$\int\int \exp\left(i\frac{(p\dot{q}-\mathbf{H})dt}{\hbar}\right)Dq\,Dp$$
$$\Rightarrow \exp\left(i\frac{(F_b-F_a)}{\hbar}\right)$$
$$\times\int\int \exp\left(i\int\frac{(P\dot{Q}-\mathbf{K})dt}{\hbar}\right)DQ\,DP,$$

where

$$\mathbf{K}(Q,P,t)=\mathbf{H}(q,p,t)+\left(\frac{\partial F}{\partial t}\right)_{qQ},$$

so that using the symmetric expansion scheme this is interpreted as

$$e^{i(F_b-F_a)/\hbar}\lim_{N\to\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{j=1}^{N-1}\frac{dQ(j)dP_M(j)}{2\pi\hbar}$$
$$\times\exp\left[i\sum_{k=1}^{N}\frac{\{P_M(k)(Q(k)-Q(k-1))-\mathbf{K}(Q_M(k),P_M(k),k-\frac{1}{2})\Delta t\}}{\hbar}\right],$$

where

$$\mathbf{K}(Q_M(k),P_M(k),k-\tfrac{1}{2})$$
$$=\mathbf{H}\left(q_M(k),p_M(k),k-\frac{1}{2}\right)+\left(\frac{\partial F_M}{\partial t}\right)_{qQ},$$
$$Q_M(k)\equiv\tfrac{1}{2}(Q(k)+Q(k-1)),$$
$$P_M(k)\equiv\tfrac{1}{2}(P(k)+P(k-1)),$$
$$q_M(k)=q(Q_M(k),P_M(k),k-\tfrac{1}{2}),$$
$$p_M(k)=p(Q_M(k),P_M(k),k-\tfrac{1}{2}).$$

Under this notation an m subscript means at the q, p midpoint, while an M subscript means at the Q, P midpoint.

In general, the naively transformed expression is not equal to its parent, due to the generation of extra (stochastic) terms in the process. That is to say, a general discretization prescription is not canonically invariant.[15] The task ahead is to correctly perform a canonical transformation and determine the stochastic terms so generated in the hope that they sum to zero, as required of a canonically invariant scheme.

VIII. MIDPOINT EXPANSIONS

Of use are midpoint expansions, since in the symmetric prescription all functions are evaluated at the midpoint.

Adopting the following notation:

$$f_m(k)\equiv f(q_m(k),p_m(k),k-\tfrac{1}{2}),$$
$$f_M(k)\equiv f(q(Q_M(k),P_M(k),k-\tfrac{1}{2}),$$
$$p(Q_M(k),P_M(k),t-\tfrac{1}{2}),k-\tfrac{1}{2}),$$
$$f_{,Q}\equiv\left(\frac{\partial F}{\partial Q}\right)_{Pt}, \quad f^{,P}\equiv\left(\frac{\partial F}{\partial P}\right)_{Qt}, \quad f_{,t}\equiv\left(\frac{\partial F}{\partial t}\right)_{QP},$$
$$\Delta f(k)\equiv f(q(k),p(k),k)-f(q(k-1),p(k-1),k-1).$$

Note that this implies that

$$q_m(k)\equiv\tfrac{1}{2}(q(k)+q(k-1)),$$
$$Q_M(k)\equiv\tfrac{1}{2}(Q(k)+Q(k-1)),$$

whereas

$$q_M(k)\equiv q(Q_M(k),P_M(k),k-\tfrac{1}{2})$$

similarly for p, P.

Then by Taylor expansion

$$\begin{aligned} f_m=f_M &+\tfrac{1}{8}f_{M,QQ} && \Delta Q\Delta Q\\ &+\tfrac{1}{8}f_{M,}{}^{PP} && \Delta P\Delta P\\ &+\tfrac{1}{8}f_{M,tt} && \Delta t\Delta t\\ &+\tfrac{1}{4}f_{M,Q}{}^{P} && \Delta Q\Delta P\\ &+\tfrac{1}{4}f_{M,Qt} && \Delta Q\Delta t\\ &+\tfrac{1}{4}f_{M,}{}^{P}{}_{t} && \Delta P\Delta t\\ &+\cdots, \end{aligned}$$

$$
\begin{aligned}
\Delta f = \; & f_{M,Q} && \Delta Q \\
& + f_{,M}{}^{P} && \Delta P \\
& + f_{M,t} && \Delta t \\
& + \tfrac{1}{24} f_{M,QQQ} && \Delta Q \Delta Q \Delta Q \\
& + \tfrac{1}{24} f_{M,}{}^{PPP} && \Delta P \Delta P \Delta P \\
& + \tfrac{1}{24} f_{M,ttt} && \Delta t \Delta t \Delta t \\
& + \tfrac{1}{8} f_{M,QQ}{}^{P} && \Delta Q \Delta Q \Delta P \\
& + \tfrac{1}{8} f_{M,QQt} && \Delta Q \Delta Q \Delta t \\
& + \tfrac{1}{8} f_{M,}{}^{PP}{}_{,t} && \Delta P \Delta P \Delta t \\
& + \tfrac{1}{8} f_{M,Q}{}^{PP} && \Delta Q \Delta P \Delta P \\
& + \tfrac{1}{8} f_{M,Qtt} && \Delta Q \Delta t \Delta t \\
& + \tfrac{1}{4} f_{M,Q}{}^{P}{}_{,t} && \Delta Q \Delta P \Delta t \\
& + \cdots .
\end{aligned}
$$

By considering F as a function of Q, P, and t, as opposed to q, Q, and t, we may develop

$$
\begin{aligned}
\Delta F = \; & F_{M,Q} && \Delta Q \\
& + F_{M,}{}^{P} && \Delta P \\
& + F_{M,t} && \Delta t \\
& + \tfrac{1}{24} F_{M,QQQ} && \Delta Q \Delta Q \Delta Q \\
& + \tfrac{1}{24} F_{M,}{}^{PPP} && \Delta P \Delta P \Delta P \\
& + \tfrac{1}{24} F_{M,ttt} && \Delta t \Delta t \Delta t \\
& + \tfrac{1}{8} F_{M,QQ}{}^{P} && \Delta Q \Delta Q \Delta P \\
& + \tfrac{1}{8} F_{M,QQt} && \Delta Q \Delta Q \Delta t \\
& + \tfrac{1}{8} F_{M,}{}^{PP}{}_{,t} && \Delta P \Delta P \Delta t \\
& + \tfrac{1}{8} F_{M,Q}{}^{PP} && \Delta Q \Delta P \Delta P \\
& + \tfrac{1}{8} F_{M,Qtt} && \Delta Q \Delta t \Delta t \\
& + \tfrac{1}{4} F_{M,Q}{}^{P}{}_{,t} && \Delta Q \Delta P \Delta t \\
& + \cdots .
\end{aligned}
$$

The generating function derivatives may be converted to p,q derivatives by starting from

$$
\begin{aligned}
\delta F &= \left(\frac{\partial F}{\partial Q}\right)_{qt} \delta Q + \left(\frac{\partial F}{\partial q}\right)_{Qt} \delta q + \left(\frac{\partial F}{\partial t}\right)_{qQ} \delta t \\
&= -P \delta Q + p \delta q + \left(\frac{\partial F}{\partial t}\right)_{qQ} \delta t
\end{aligned}
$$

leading to

$$
\begin{aligned}
F_{,Q} &= -P + q_{,Q} p, \\
F_{,}{}^{P} &= q_{,}{}^{P} p, \\
F_{,t} &= q_{,t} p + \left(\frac{\partial F}{\partial t}\right)_{qQ},
\end{aligned}
$$

$$
\begin{aligned}
F_{,QQQ} &= q_{,QQQ} p + 2 q_{,QQ} p_{,Q} + q_{,Q} p_{,QQ}, \\
F_{,QQ}{}^{P} &= q_{,QQ}{}^{P} p + 2 q_{,Q}{}^{P} p_{,Q} + q_{,}{}^{P} p_{,QQ}, \\
F_{,Q}{}^{PP} &= q_{,Q}{}^{PP} p + 2 q_{,Q}{}^{P} p_{,}{}^{P} + q_{,Q} p_{,}{}^{PP}, \\
F_{,}{}^{PPP} &= q_{,}{}^{PPP} p + 2 q_{,}{}^{PP} p_{,}{}^{P} + q_{,}{}^{P} p_{,}{}^{PP}, \\
F_{,Qtt} &= q_{,Qtt} p + 2 q_{,Qt} p_{,t} + q_{,Q} p_{,tt}, \\
F_{,}{}^{P}{}_{tt} &= q_{,}{}^{P}{}_{tt} p + 2 q_{,}{}^{P} p_{,t} + q_{,}{}^{P} p_{,tt}.
\end{aligned}
$$

Having used the necessary and sufficient condition for a canonical transformation,[4]

$$
q_{,Q} p_{,}{}^{P} - q_{,}{}^{P} p_{,Q} = 1,
$$

which follows from the canonical conditions given earlier. Care should be taken of which variables are being held constant.

We may now convert the path integral in the search for additional contributions beyond the formal conversion. These additional stochastic terms will stem from both the Jacobian and action.

Look for stochastic contributions from the Jacobian. Since

$$
\int \cdots \int \prod_{j=1}^{N-1} \frac{dq(j)\,dp_m(j)}{2\pi\hbar} = \int \cdots \int \prod_{j=1}^{N-1} \frac{dq_m(j)\,dp_m(j)}{\pi\hbar}
$$

and

$$
\int \cdots \int \prod_{j=1}^{N-1} \frac{dQ(j)\,dP_M(j)}{2\pi\hbar} = \int \cdots \int \prod_{j=1}^{N-1} \frac{dQ_M(j)\,dP_M(j)}{\pi\hbar}
$$

the Jacobian of interest arises from

$$
dq_m(j)\,dp_m(j) \to dQ_M(j)\,dP_M(j).
$$

This is not unity because $q_m(j)$, $p_m(j)$ and $Q_M(j)$, $P_M(j)$ are based at different points in phase space (recalling that we are transforming q and p as opposed to q_m and p_m).

The Jacobian

$$
\mathscr{J} = \left(\frac{\partial q_m}{\partial Q_M}\right)_{Pt} \left(\frac{\partial p_m}{\partial P_M}\right)_{Qt} - \left(\frac{\partial q_m}{\partial P_M}\right)_{Qt} \left(\frac{\partial p_m}{\partial Q_M}\right)_{Pt}
$$

which leads to

$$
\begin{aligned}
\mathscr{J} = \; & q_{M,Q} p_{M,}{}^{P} - q_{M,}{}^{P} p_{M,Q} \\
& + \tfrac{1}{8}\left(q_{M,QQQ} p_{M,}{}^{P} + q_{M,Q} p_{M,QQ}{}^{P} - q_{M,}{}^{P} p_{M,QQQ} - q_{M,QQ}{}^{P} p_{M,Q}\right) && \Delta Q \Delta Q \\
& + \tfrac{1}{8}\left(q_{M,Q}{}^{PP} p_{M,}{}^{P} + q_{M,Q} p_{M,}{}^{PPP} - q_{M,}{}^{P} p_{M,Q}{}^{PP} - q_{M,}{}^{PPP} p_{M,Q}\right) && \Delta P \Delta P \\
& + \tfrac{1}{8}\left(q_{M,Qtt} p_{M,}{}^{P} + q_{M,Q} p_{M,}{}^{P}{}_{tt} - q_{M,}{}^{P} p_{M,Qtt} - q_{M,}{}^{P}{}_{tt} p_{M,Q}\right) && \Delta t \Delta t \\
& + \tfrac{1}{8}\left(q_{M,QQ}{}^{P} p_{M,}{}^{P} + q_{M,Q} p_{M,Q}{}^{PP} - q_{M,}{}^{P} p_{M,QQ}{}^{P} - q_{M,Q}{}^{PP} p_{M,Q}\right) && \Delta Q \Delta P \\
& + \tfrac{1}{8}\left(q_{M,QQt} p_{M,}{}^{P} + q_{M,Q} p_{M,Q}{}^{P}{}_{t} - q_{M,}{}^{P} p_{M,QQt} - q_{M,Q}{}^{P}{}_{t} p_{M,Q}\right) && \Delta Q \Delta t \\
& + \tfrac{1}{8}\left(q_{M,Q}{}^{P}{}_{t} p_{M,}{}^{P} + q_{M,Q} p_{M,}{}^{PP}{}_{t} - q_{M,}{}^{P} p_{M,Q}{}^{P}{}_{t} - q_{M,}{}^{PP}{}_{t} p_{M,Q}\right) && \Delta P \Delta t \\
& + \cdots ,
\end{aligned}
$$

but, for a canonical transformation,

$$
q_{,Q} p_{,}{}^{P} - q_{,}{}^{P} p_{,Q} = 1
$$

everywhere. By applying this at the **M** midpoint yields

$$
\begin{aligned}
\mathscr{J} = 1 &+ \tfrac{1}{4}(q_{M,Q}{}^{P} p_{M,QQ} - q_{M,QQ} p_{M,Q}{}^{P}) && \Delta Q \Delta Q \\
&+ \tfrac{1}{4}(q_{M,}{}^{PP} p_{M,Q}{}^{P} - q_{M,Q}{}^{P} p_{M,}{}^{PP}) && \Delta P \Delta P \\
&+ \tfrac{1}{4}(q_{M,}{}^{P}{}_{t} p_{M,Qt} - q_{M,Qt} p_{M,}{}^{P}{}_{t}) && \Delta t \Delta t \\
&+ \tfrac{1}{8}(q_{M,}{}^{PP} p_{M,QQ} - q_{M,QQ} p_{M,}{}^{PP}) && \Delta Q \Delta P \\
&+ \tfrac{1}{8}(q_{M,}{}^{P}{}_{t} p_{M,QQ} + q_{M,Q}{}^{P} p_{M,Qt} - q_{M,QQ} p_{M,}{}^{P}{}_{t} - q_{M,Qt} p_{M,Q}{}^{P}) && \Delta Q \Delta t \\
&+ \tfrac{1}{8}(q_{M,}{}^{P}{}_{t} p_{M,Q}{}^{P} + q_{M,}{}^{PP} p_{M,Qt} - q_{M,Q}{}^{P} p_{M,}{}^{P}{}_{t} - q_{M,Qt} p_{M,}{}^{PP}) && \Delta P \Delta t \\
&+ \cdots .
\end{aligned}
$$

Continue by looking at the action term $p_m(k)\Delta q(k)$ and comparing it to $P_M(k)\Delta Q(k) + \Delta F(k) - (\partial F_M(k)/\partial t)_{qQ}\Delta t$, its formal counterpart. This leads to

$$
\begin{aligned}
p_m \Delta q = \Delta F + P_M \Delta Q - \left(\frac{\partial F_M}{\partial t}\right)_{qQ} \Delta t \\
&+ \tfrac{1}{12}(q_{M,Q} p_{M,QQ} - q_{M,QQ} p_{M,Q}) && \Delta Q \Delta Q \Delta Q \\
&+ \tfrac{1}{12}(q_{M,}{}^{P} p_{M,}{}^{PP} - q_{M,}{}^{PP} p_{M,}{}^{P}) && \Delta P \Delta P \Delta P \\
&+ \tfrac{1}{24}(q_{M,ttt} p_M + 3 q_{M,t} p_{M,tt} - F_{M,ttt}) && \Delta t \Delta t \Delta t \\
&+ \tfrac{1}{4}(q_{M,Q} p_{M,Q}{}^{P} - q_{M,Q}{}^{P} p_{M,Q}) && \Delta Q \Delta Q \Delta P \\
&+ \tfrac{1}{8}(q_{M,QQt} p_M + q_{M,t} p_{M,QQ} + 2 q_{M,Q} p_{M,Qt} - F_{M,QQt}) && \Delta Q \Delta Q \Delta t \\
&+ \tfrac{1}{8}(q_{M,}{}^{PP}{}_{t} p_M + q_{M,t} p_{M,}{}^{PP} + 2 q_{M,}{}^{P} p_{M,}{}^{P}{}_{t} - F_{M,}{}^{PP}{}_{t}) && \Delta P \Delta P \Delta t \\
&+ \tfrac{1}{4}(q_{M,Q}{}^{P}{}_{t} p_M + q_{M,t} p_{M,Q}{}^{P} + q_{M,}{}^{P} p_{M,Qt} + q_{M,Q} p_{M,}{}^{P}{}_{t} - F_{M,Q}{}^{P}{}_{t}) && \Delta Q \Delta P \Delta t \\
&+ \tfrac{1}{4}(q_{M,}{}^{P} p_{M,Q}{}^{P} - q_{M,Q}{}^{P} p_{M,}{}^{P}) && \Delta Q \Delta P \Delta P \\
&+ \tfrac{1}{8}(q_{M,Q} p_{M,tt} + 2 q_{M,t} p_{M,Qt} + q_{M,Qtt} p_M - F_{M,Qtt}) && \Delta Q \Delta t \Delta t \\
&+ \tfrac{1}{8}(q_{M,}{}^{P} p_{M,tt} + 2 q_{M,t} p_{M,}{}^{P}{}_{t} + q_{M,}{}^{P}{}_{tt} p_M - F_{M,}{}^{P}{}_{tt}) && \Delta P \Delta t \Delta t \\
&+ \cdots .
\end{aligned}
$$

Finally, determining stochastic contributions from the Hamiltonian. By again applying the canonical transformation at the **M** midpoint

$$
\mathbf{K}_M = \mathbf{H}_M + \left(\frac{\partial F_M}{\partial t}\right)_{qQ}
$$

the original Hamiltonian becomes

$$
\begin{aligned}
\mathbf{H_m} = \mathbf{K_M} - \left(\frac{\partial F_M}{\partial t}\right)_{qQ} \\
&+ \tfrac{1}{8}\mathbf{H}_{M,QQ} \; \Delta Q \Delta Q \\
&+ \tfrac{1}{8}\mathbf{H}_{M,}{}^{PP} \; \Delta P \Delta P \\
&+ \tfrac{1}{8}\mathbf{H}_{M,tt} \; \Delta t \Delta t \\
&+ \tfrac{1}{4}\mathbf{H}_{M,Q}{}^{P} \; \Delta Q \Delta P \\
&+ \tfrac{1}{4}\mathbf{H}_{M,Qt} \; \Delta Q \Delta t \\
&+ \tfrac{1}{4}\mathbf{H}_{M,}{}^{P}{}_{t} \; \Delta P \Delta t \\
&+ \cdots .
\end{aligned}
$$

Collecting up the stochastic terms

$$
\left.
\begin{aligned}
&\tfrac{1}{4}(q_{M,Q}{}^{P} p_{M,QQ} - q_{M,QQ} p_{M,Q}{}^{P}) && \Delta Q \Delta Q \\
&+ \tfrac{1}{4}(q_{M,}{}^{PP} p_{M,Q}{}^{P} - q_{M,Q}{}^{P} p_{M,}{}^{PP}) && \Delta P \Delta P \\
&+ \tfrac{1}{4}(q_{M,}{}^{P}{}_{t} p_{M,Qt} - q_{M,Qt} p_{M,}{}^{P}{}_{t}) && \Delta t \Delta t \\
&+ \tfrac{1}{8}(q_{M,}{}^{PP} p_{M,QQ} - q_{M,QQ} p_{M,}{}^{PP}) && \Delta Q \Delta P \\
&+ \tfrac{1}{8}(q_{M,}{}^{P}{}_{t} p_{M,QQ} + q_{M,Q}{}^{P} p_{M,Qt} - q_{M,QQ} p_{M,}{}^{P}{}_{t} - q_{M,Qt} p_{M,Q}{}^{P}) && \Delta Q \Delta t \\
&+ \tfrac{1}{8}(q_{M,}{}^{P}{}_{t} p_{M,Q}{}^{P} + q_{M,}{}^{PP} p_{M,Qt} - q_{M,Q}{}^{P} p_{M,}{}^{P}{}_{t} - q_{M,Qt} p_{M,}{}^{PP}) && \Delta P \Delta t
\end{aligned}
\right\}
\quad \text{Jacobian contribution}
$$

$$+ \cdots ,$$

$$
\left.\begin{array}{ll}
i\{\tfrac{1}{12}(q_{M,Q}p_{M,QQ} - q_{M,QQ}p_{M,Q}) & \Delta Q\Delta Q\Delta Q \\
+\tfrac{1}{12}(q_{M,}{}^{P}p_{M,}{}^{PP} - q_{M,}{}^{PP}p_{M,}{}^{P}) & \Delta P\Delta P\Delta P \\
+\tfrac{1}{24}(q_{M,ttt}p_M + 3q_{M,t}p_{M,tt} - F_{M,ttt}) & \Delta t\Delta t\Delta t \\
+\tfrac{1}{4}(q_{M,Q}p_{M,Q}{}^{P} - q_{M,Q}{}^{P}p_{M,Q}) & \Delta Q\Delta Q\Delta P \\
+\tfrac{1}{8}(q_{M,QQt}p_M + q_{M,t}p_{M,QQ} + 2q_{M,Q}p_{M,Qt} - F_{M,QQt}) & \Delta Q\Delta Q\Delta t \\
+\tfrac{1}{8}(q_{M,}{}^{PP}{}_{t}p_M + q_{M,t}p_{M,}{}^{PP} + 2q_{M,}{}^{P}p_{M,}{}^{P}{}_{t} - F_{M,}{}^{PP}{}_{t}) & \Delta P\Delta P\Delta t \\
+\tfrac{1}{4}(q_{M,Q}{}^{P}{}_{t}p_M + q_{M,t}p_{M,Q}{}^{P} + q_{M,}{}^{P}p_{M,Qt} + q_{M,Q}p_{M,}{}^{P}{}_{t} - F_{M,Q}{}^{P}{}_{t}) & \Delta Q\Delta P\Delta t \\
+\tfrac{1}{4}(q_{M,}{}^{P}p_{M,Q}{}^{P} - q_{M,Q}{}^{P}p_{M,}{}^{P}) & \Delta Q\Delta P\Delta P \\
+\tfrac{1}{8}(q_{M,Q}p_{M,tt} + 2q_{M,t}p_{M,Qt} + q_{M,Qtt}p_M - F_{M,Qtt}) & \Delta Q\Delta t\Delta t \\
+\tfrac{1}{8}(q_{M,}{}^{P}p_{M,tt} + 2q_{M,t}p_{M,}{}^{P}{}_{t} + q_{M,}{}^{P}{}_{tt}p_M - F_{M,}{}^{P}{}_{tt}) & \Delta P\Delta t\Delta t
\end{array}\right\} \begin{array}{l} p\dot{q} \\ \text{contribution} \end{array}
$$

$$+\cdots$$

$$
\left.\begin{array}{ll}
-\tfrac{1}{8}\mathbf{H}_{M,QQ} & \Delta Q\Delta Q\Delta t \\
-\tfrac{1}{8}\mathbf{H}_{M,}{}^{PP} & \Delta P\Delta P\Delta t \\
-\tfrac{1}{8}\mathbf{H}_{M,tt} & \Delta t\Delta t\Delta t \\
-\tfrac{1}{4}\mathbf{H}_{M,Q}{}^{P} & \Delta Q\Delta P\Delta t \\
-\tfrac{1}{4}\mathbf{H}_{M,Qt} & \Delta Q\Delta t\Delta t \\
-\tfrac{1}{4}\mathbf{H}_{M,}{}^{P}{}_{t} & \Delta P\Delta t\Delta t
\end{array}\right\} \begin{array}{l} \text{Hamiltonian} \\ \text{contribution} \end{array}
$$

$$+\cdots\}/\hbar.$$

Working to order Δt, the correctly transformed path integral then has the form

$$
I \equiv e^{i(F_b - F_a)/\hbar} \lim_{N\to\infty} \int\cdots\int \prod_{j=1}^{N-1} \frac{dQ(j)dP_M(j)}{2\pi\hbar} \prod_{k=1}^{N} (1 + \text{Jacobian stochastic terms})
$$
$$
\times \exp\left[i\, \frac{\{P_M(k)(Q(k) - Q(k-1)) - \mathbf{K}(Q_M(k),P_M(k),k-\tfrac{1}{2})\Delta t + \text{action stochastic terms}\}}{\hbar} \right].
$$

Now work to replace the stochastic terms by a potential-like term[16]

$$\mathbf{V}(Q_M)\Delta t$$

of the same effect (to order Δt). This can only be achieved if the stochastic terms are of order Δt.

So further define

$$
J \equiv e^{i(F_b - F_a)/\hbar} \lim_{N\to\infty} \int\cdots\int \prod_{j=1}^{N-1} \frac{dQ(j)dP_M(j)}{2\pi\hbar}
$$
$$
\times \exp\left[i \sum_{k=1}^{N} \frac{\{P_M(k)(Q(k) - Q(k-1)) - \mathbf{K}(Q_M(k),P_M(k),k-\tfrac{1}{2})\Delta t - \mathbf{V}(Q_M(k))\Delta t\}}{\hbar} \right].
$$

Working to order Δt

$$
I = e^{i(F_b - F_a)/\hbar} \lim_{N\to\infty} \int\cdots\int \prod_{j=1}^{N-1} \frac{dQ(j)dP_M(j)}{2\pi\hbar}
$$
$$
\times \prod_{k=1}^{N} (1 + \text{Jacobian stochastic terms} + i \text{ action stochastic terms}/\hbar)
$$
$$
\times \exp\left[i\, \frac{\{P_M(k)(Q(k) - Q(k-1)) - \mathbf{K}(Q_M(k),P_M(k),k-\tfrac{1}{2})\Delta t\}}{\hbar} \right]
$$

and

$$
J = e^{i(F_b - F_a)/\hbar} \lim_{N\to\infty} \int\cdots\int \prod_{j=1}^{N-1} \frac{dQ(j)dP_M(j)}{2\pi\hbar} \prod_{k=1}^{N} \left(1 - \frac{i\mathbf{V}(Q_M(k))\Delta t}{\hbar}\right)
$$
$$
\times \exp\left[i\, \frac{\{P_M(k)(Q(k) - Q(k-1)) - \mathbf{K}(Q_M(k),P_M(k),k-\tfrac{1}{2})\Delta t\}}{\hbar} \right].
$$

Consider now the particular cases of (i) a Hamilton–Jacobi transformation, (ii) a point canonical transformation.

IX. HAMILTON–JACOBI TRANSFORMATIONS

The Hamilton–Jacobi transformation is the special case where the canonically transformed Hamiltonian (Kamiltonian) given by

$$
\mathbf{K} = \mathbf{H} + \left(\frac{\partial F}{\partial t}\right)_{qQ}
$$

is null, i.e.,

$$\mathbf{H}(q,p,t) + \left(\frac{\partial F}{\partial t}\right)_{qQ} = 0,$$

but recall

$$p = \left(\frac{\partial F}{\partial q}\right)_{Qt},$$

so that the generator (Γ) of the required transformation is given by

$$\mathbf{H}\left(q,\left(\frac{\partial \Gamma}{\partial q}\right)_{Qt},t\right) + \left(\frac{\partial \Gamma}{\partial t}\right)_{qQ} = 0,$$

the Hamilton–Jacobi equation. It should be recalled that this Hamiltonian will in general look "quantum mechanical" due to the adoption of $\hbar$ terms when Weyl ordering. In this special case of a Hamilton–Jacobi transformation

$$I = e^{i(\Gamma_b - \Gamma_a)/\hbar} \lim_{N\to\infty} \int \cdots \int \prod_{j=1}^{N-1} \frac{dQ(j)dP_M(j)}{2\pi\hbar} \prod_{k=1}^{N} (1 + \text{Jacobian stochastic terms}$$
$$+\, i \text{ action stochastic terms}/\hbar) \exp\left[i\,\frac{\{P_M(k)(Q(k) - Q(k-1))\}}{\hbar}\right]$$

to be compared with

$$J = e^{i(\Gamma_b - \Gamma_a)/\hbar} \lim_{N\to\infty} \int \cdots \int \prod_{j=1}^{N-1} \frac{dQ(j)dP_M(j)}{2\pi\hbar} \prod_{k=1}^{N} \left(1 - \frac{i\mathbf{V}(Q_M(k))\Delta t}{\hbar}\right)$$
$$\times \exp\left[i\,\frac{\{P_M(k)(Q(k) - Q(k-1))\}}{\hbar}\right].$$

These integrals are ill defined and may be made tractable by the inclusion of a "mass" term,

$$I = e^{i(\Gamma_b - \Gamma_a)/\hbar} \lim_{N,\Lambda\to\infty} \int \cdots \int \prod_{j=1}^{N-1} \frac{dQ(j)dP_M(j)}{2\pi\hbar} \prod_{k=1}^{N} (1 + \text{Jacobian stochastic terms}$$
$$+\, i \text{ action stochastic terms}/\hbar) \exp\left[i\,\frac{\{P_M(k)(Q(k) - Q(k-1)) - P_M{}^2(k)\Delta t/2\Lambda\}}{\hbar}\right],$$

$$J = e^{i(\Gamma_b - \Gamma_a)/\hbar} \lim_{N,\Lambda\to\infty} \int \cdots \int \prod_{j=1}^{N-1} \frac{dQ(j)dP_M(j)}{2\pi\hbar} \prod_{k=1}^{N} \left(1 - \frac{i\mathbf{V}(Q_M(k))\Delta t}{\hbar}\right)$$
$$\times \exp\left[i\,\frac{\{P_M(k)(Q(k) - Q(k-1)) - P_M{}^2(k)\Delta t/2\Lambda\}}{\hbar}\right],$$

where the stochastic terms are

$$\left.\begin{array}{ll}
\frac{1}{4}(q_{M,Q}{}^{P}p_{M,QQ} - q_{M,QQ}p_{M,Q}{}^{P}) & \Delta Q\Delta Q \\
+\frac{1}{4}(q_{M,}{}^{PP}p_{M,Q}{}^{P} - q_{M,Q}{}^{P}p_{M,}{}^{PP}) & \Delta P\Delta P \\
+\frac{1}{4}(q_{M,}{}^{P}{}_{t}p_{M,Qt} - q_{M,Qt}p_{M,}{}^{P}{}_{t}) & \Delta t\Delta t \\
+\frac{1}{8}(q_{M,}{}^{PP}p_{M,QQ} - q_{M,QQ}p_{M,}{}^{PP}) & \Delta Q\Delta P \\
+\frac{1}{8}(q_{M,}{}^{P}{}_{t}p_{M,QQ} + q_{M,Q}{}^{P}p_{M,Qt} - q_{M,QQ}p_{M,}{}^{P}{}_{t} - q_{M,Qt}p_{M,Q}{}^{P}) & \Delta Q\Delta t \\
+\frac{1}{8}(q_{M,}{}^{P}{}_{t}p_{M,Q}{}^{P} + q_{M,}{}^{PP}p_{M,Qt} - q_{M,Q}{}^{P}p_{M,}{}^{P}{}_{t} - q_{M,Qt}p_{M,}{}^{PP}) & \Delta P\Delta t \\
+\cdots, &
\end{array}\right\}\ \text{Jacobian contribution}$$

$$\left.\begin{array}{ll}
i\{\frac{1}{12}(q_{M,Q}p_{M,QQ} - q_{M,QQ}p_{M,Q}) & \Delta Q\Delta Q\Delta Q \\
+\frac{1}{12}(q_{M,}{}^{P}p_{M,}{}^{PP} - q_{M,}{}^{PP}p_{M,}{}^{P}) & \Delta P\Delta P\Delta P \\
+\frac{1}{24}(q_{M,ttt}p_M + 3q_{M,t}p_{M,tt} - \Gamma_{M,ttt}) & \Delta t\Delta t\Delta t \\
+\frac{1}{4}(q_{M,Q}p_{M,Q}{}^{P} - q_{M,Q}{}^{P}p_{M,Q}) & \Delta Q\Delta Q\Delta P \\
+\frac{1}{8}(q_{M,QQt}p_M + q_{M,t}p_{M,QQ} + 2q_{M,Q}p_{M,Qt} - \Gamma_{M,QQt}) & \Delta Q\Delta Q\Delta t \\
+\frac{1}{8}(q_{M,}{}^{PP}{}_{t}p_M + q_{M,t}p_{M,}{}^{PP} + 2q_{M,}{}^{P}p_{M,}{}^{P}{}_{t} - \Gamma_{M,}{}^{PP}{}_{t}) & \Delta P\Delta P\Delta t \\
+\frac{1}{4}(q_{M,Q}{}^{P}{}_{t}p_M + q_{M,t}p_{M,Q}{}^{P} + q_{M,}{}^{P}p_{M,Qt} + q_{M,Q}p_{M,}{}^{P}{}_{t} - \Gamma_{M,Q}{}^{P}{}_{t}) & \Delta Q\Delta P\Delta t \\
+\frac{1}{4}(q_{M,}{}^{P}p_{M,Q}{}^{P} - q_{M,Q}{}^{P}p_{M,}{}^{P}) & \Delta Q\Delta P\Delta P \\
+\frac{1}{8}(q_{M,Q}p_{M,tt} + 2q_{M,t}p_{M,Qt} + q_{M,Qtt}p_M - \Gamma_{M,Qtt}) & \Delta Q\Delta t\Delta t \\
+\frac{1}{8}(q_{M,}{}^{P}p_{M,tt} + 2q_{M,t}p_{M,}{}^{P}{}_{t} + q_{M,}{}^{P}{}_{tt}p_M - \Gamma_{M,}{}^{P}{}_{tt}) & \Delta P\Delta t\Delta t \\
+\cdots &
\end{array}\right\}\ p\dot{q}\ \text{contribution}$$

$$
\left.\begin{array}{ll}
-\tfrac{1}{8}\mathbf{H}_{M,QQ} & \Delta Q\Delta Q\Delta t \\
-\tfrac{1}{8}\mathbf{H}_{M,}{}^{PP} & \Delta P\Delta P\Delta t \\
-\tfrac{1}{8}\mathbf{H}_{M,tt} & \Delta t\Delta t\Delta t \\
-\tfrac{1}{4}\mathbf{H}_{M,Q}{}^{P} & \Delta Q\Delta P\Delta t \\
-\tfrac{1}{4}\mathbf{H}_{M,Qt} & \Delta Q\Delta t\Delta t \\
-\tfrac{1}{4}\mathbf{H}_{M,}{}^{P}{}_{t} & \Delta P\Delta t\Delta t
\end{array}\right\}\ \begin{array}{l}\text{Hamiltonian}\\ \text{contribution}\end{array}
$$
$$+\cdots\}/\hbar.$$

In this *special case* of a Hamilton–Jacobi transformation $\mathbf{K}=0$, so

$$\mathbf{H}=-\left(\frac{\partial\Gamma}{\partial t}\right)_{qQ}=-\Gamma_{,t}+pq_{,t}$$

as was shown previously.

This leads to the Hamilton stochastic terms canceling may of their $p\dot{q}$ counterparts, since the Hamilton stochastic contribution

$$
=i\left\{\begin{array}{ll}
-\tfrac{1}{8}(-\Gamma_{M,QQt}+q_{M,QQt}p_M+2q_{M,Qt}p_{M,Q}+q_{M,t}p_{M,QQ}) & \Delta Q\Delta Q\Delta t \\
-\tfrac{1}{8}(-\Gamma_{M,}{}^{PP}{}_{t}+q_{M,}{}^{PP}{}_{t}p_M+2q_{M,}{}^{P}{}_{t}p_{M,}{}^{P}+q_{M,t}p_{M,}{}^{PP}) & \Delta P\Delta P\Delta t \\
-\tfrac{1}{8}(-\Gamma_{M,ttt}+q_{M,ttt}p_M+2q_{M,tt}p_{M,t}+q_{M,t}p_{M,tt}) & \Delta t\Delta t\Delta t \\
-\tfrac{1}{4}(-\Gamma_{M,Q}{}^{P}{}_{t}+q_{M,Q}{}^{P}{}_{t}p_M+q_{M,Qt}p_{M,}{}^{P}+q_{M,}{}^{P}{}_{t}p_{M,Q}+q_{M,t}p_{M,Q}{}^{P}) & \Delta Q\Delta P\Delta t \\
-\tfrac{1}{4}(-\Gamma_{M,Qtt}+q_{M,Qtt}p_M+q_{M,Qt}p_{M,t}+q_{M,tt}p_{M,Q}+q_{M,t}p_{M,Qt}) & \Delta Q\Delta t\Delta t \\
-\tfrac{1}{4}(-\Gamma_{M,}{}^{P}{}_{tt}+q_{M,}{}^{P}{}_{tt}p_M+q_{M,}{}^{P}{}_{t}p_{M,t}+q_{M,tt}p_{M,}{}^{P}+q_{M,t}p_{M,}{}^{P}{}_{t}) & \Delta P\Delta t\Delta t \\
+\cdots\}/\hbar, &
\end{array}\right.
$$

yielding the set of stochastic terms:

$$
\left.\begin{array}{ll}
\tfrac{1}{4}(q_{M,Q}{}^{P}p_{M,QQ}-q_{M,QQ}p_{M,Q}{}^{P}) & \Delta Q\Delta Q \\
+\tfrac{1}{4}(q_{M,}{}^{PP}p_{M,Q}{}^{P}-q_{M,Q}{}^{P}p_{M,}{}^{PP}) & \Delta P\Delta P \\
+\tfrac{1}{4}(q_{M,}{}^{P}{}_{t}p_{M,Qt}-q_{M,Qt}p_{M,}{}^{P}{}_{t}) & \Delta t\Delta t \\
+\tfrac{1}{8}(q_{M,}{}^{PP}p_{M,QQ}-q_{M,QQ}p_{M,}{}^{PP}) & \Delta Q\Delta P \\
+\tfrac{1}{8}(q_{M,}{}^{P}{}_{t}p_{M,QQ}+q_{M,Q}{}^{P}p_{M,Qt}-q_{M,QQ}p_{M,}{}^{P}{}_{t}-q_{M,Qt}p_{M,Q}{}^{P}) & \Delta Q\Delta t \\
+\tfrac{1}{8}(q_{M,}{}^{P}{}_{t}p_{M,Q}{}^{P}+q_{M,}{}^{PP}p_{M,Qt}-q_{M,Q}{}^{P}p_{M,}{}^{P}{}_{t}-q_{M,Qt}p_{M,}{}^{PP}) & \Delta P\Delta t
\end{array}\right\}\ \begin{array}{l}\text{Jacobian}\\ \text{contribution}\end{array}
$$
$$+\cdots,$$

$$
\left.\begin{array}{ll}
i\{\tfrac{1}{12}(q_{M,Q}p_{M,QQ}-q_{M,QQ}p_{M,Q}) & \Delta Q\Delta Q\Delta Q \\
+\tfrac{1}{12}(q_{M,}{}^{P}p_{M,}{}^{PP}-q_{M,}{}^{PP}p_{M,}{}^{P}) & \Delta P\Delta P\Delta P \\
+\tfrac{1}{12}(\Gamma_{M,ttt}-q_{M,ttt}p_M-3q_{M,tt}p_{M,t}) & \Delta t\Delta t\Delta t \\
+\tfrac{1}{4}(q_{M,Q}p_{M,Q}{}^{P}-q_{M,Q}{}^{P}p_{M,Q}) & \Delta Q\Delta Q\Delta P \\
+\tfrac{1}{4}(q_{M,Q}p_{M,Qt}-q_{M,Qt}p_{M,Q}) & \Delta Q\Delta Q\Delta t \\
+\tfrac{1}{4}(q_{M,}{}^{P}p_{M,}{}^{P}{}_{t}-q_{M,}{}^{P}{}_{t}p_{M,}{}^{P}) & \Delta P\Delta P\Delta t \\
+\tfrac{1}{4}(q_{M,}{}^{P}p_{M,Qt}+q_{M,Q}p_{M,}{}^{P}{}_{t}-q_{M,Qt}p_{M,}{}^{P}-q_{M,}{}^{P}{}_{t}p_{M,Q}) & \Delta Q\Delta P\Delta t \\
+\tfrac{1}{4}(q_{M,}{}^{P}p_{M,Q}{}^{P}-q_{M,Q}{}^{P}p_{M,}{}^{P}) & \Delta Q\Delta P\Delta P \\
+\tfrac{1}{4}(q_{M,Q}p_{M,tt}-q_{M,tt}p_{M,Q}) & \Delta Q\Delta t\Delta t \\
+\tfrac{1}{4}(q_{M,}{}^{P}p_{M,tt}-q_{M,tt}p_{M,}{}^{P}) & \Delta P\Delta t\Delta t
\end{array}\right\}\ \begin{array}{l}\text{Action}\\ \text{contribution}\end{array}
$$
$$+\cdots\}/\hbar$$

having used

$$\Gamma_{M,Qtt}=q_{M,Qtt}p_M+2q_{M,Qt}p_{M,t}+q_{M,Q}p_{M,tt},\quad \Gamma_{M,}{}^{P}{}_{tt}=q_{M,}{}^{P}{}_{tt}p_M+2q_{M,}{}^{P}{}_{t}p_{M,t}+q_{M,}{}^{P}p_{M,tt},$$

which were developed earlier.

From the Appendix it follows that only the highlighted terms contribute, namely,

$$
\begin{array}{ll}
\tfrac{1}{4}(q_{M,Q}{}^{P}p_{M,QQ}-q_{M,QQ}p_{M,Q}{}^{P}) & \Delta Q\Delta Q \\
\tfrac{1}{8}(q_{M,}{}^{P}{}_{t}p_{M,QQ}+q_{M,Q}{}^{P}p_{M,Qt}-q_{M,QQ}p_{M,}{}^{P}{}_{t}-q_{M,Qt}p_{M,Q}{}^{P}) & \Delta Q\Delta t \\
\tfrac{1}{12}(i/\hbar)(q_{M,Q}p_{M,QQ}-q_{M,QQ}p_{M,Q}) & \Delta Q\Delta Q\Delta Q \\
\tfrac{1}{4}(i/\hbar)(q_{M,Q}p_{M,Qt}-q_{M,Qt}p_{M,Q}) & \Delta Q\Delta Q\Delta t.
\end{array}
$$

Their nature can be discovered from recalling that (in the limit $\Delta t \to 0$) $p \sim (\Delta t)^{-1/2}$ as does P, which informs us that the leading stochastic terms contribute, respectively, like Δt, $(\Delta t)^{3/2}$, Δt, $(\Delta t)^{3/2}$. This is because differentiating w.r.t. P induces a $(\Delta t)^{1/2}$. Considering the two remaining terms, namely,

$$\tfrac{1}{4}(q_{M,Q}{}^{P}p_{M,QQ} - q_{M,QQ}p_{M,Q}{}^{P}) \qquad \Delta Q\Delta Q$$

$$\tfrac{1}{12}(i/\hbar)(q_{M,Q}p_{M,QQ} - q_{M,QQ}p_{M,Q}) \quad \Delta Q\Delta Q\Delta Q,$$

which have leading behavior that must then be of the form

$$\tfrac{1}{4}E(Q,t)\Delta Q\Delta Q, \qquad \tfrac{1}{12}(i/\hbar)G(Q,t)P\Delta Q\Delta Q\Delta Q,$$

to be or order Δt.

Performing these integrals (see Appendix) leads to leading contributions,

$$-\tfrac{1}{4}(\hbar/i)E(Q,t)\Delta t/\Lambda, \qquad \tfrac{1}{4}(\hbar/i)G(Q,t)\Delta t/\Lambda,$$

where $1/2\Lambda$ is the coefficient preceding the p^2 term in the Hamiltonian. A distinction is then made between the case when this term is present, and as here when not (tackled as $\lim \Lambda \to \infty$). In the latter case the terms are lost trivially; while the former is contained in the example of a point canonical transformation, considered next. In this case the stochastic terms mutually cancel.

In actual fact we have allowed rather too many canonical transformations; as those generating terms containing p^n ($n > 2$) are unphysical and lead to infinite contributions.

X. POINT CANONICAL TRANSFORMATIONS: AN EXPLICIT EXAMPLE

It is possible, in the special case of a point canonical transformation,[18] to explicitly calculate the stochastic contributions. According to the previous results, these should sum to zero.

In the case of the point canonical transformation, only the $p_m{}^2\Delta t/2m(q_m)$ "term" of the Hamiltonian need be considered, since this is the strongest term permissible [order $(\Delta t)^0$] and the only one that is active in the calculation of the order Δt effective potential $\mathbf{V}$.[16]

So look at

$$\begin{aligned}-p_m{}^2\Delta t/2m(q_m) = -\Delta t\,[p_M{}^2 &+ \tfrac{1}{4}p_M p_{M,QQ} && \Delta Q\Delta Q\\ &+ \tfrac{1}{2}p_M p_{M,Q}{}^{P} && \Delta Q\Delta P\\ &+ \tfrac{1}{4}p_M p_{M,}{}^{PP} && \Delta P\Delta P\\ &+ \cdots]/2m(q_M).\end{aligned}$$

Collecting all stochastic terms to be converted,

$$\left.\begin{aligned}&\tfrac{1}{4}(q_{M,Q}{}^{P}p_{M,QQ} - q_{M,QQ}p_{M,Q}{}^{P}) && \Delta Q\Delta Q\\ &+\tfrac{1}{4}(q_{M,}{}^{PP}p_{M,QQ} - q_{M,QQ}p_{M,}{}^{PP}) && \Delta Q\Delta P\\ &+\tfrac{1}{4}(q_{M,}{}^{PP}p_{M,Q}{}^{P} - q_{M,Q}{}^{P}p_{M,}{}^{PP}) && \Delta P\Delta P\\ &+\cdots,\end{aligned}\right\}\ \begin{array}{l}\text{Jacobian}\\ \text{contribution}\end{array}$$

$$\left.\begin{aligned}i\{\ &\tfrac{1}{12}(q_{M,Q}p_{M,QQ} - q_{M,QQ}p_{M,Q}) && \Delta Q\Delta Q\Delta Q\\ &+\tfrac{1}{4}(q_{M,Q}p_{M,Q}{}^{P} - q_{M,Q}{}^{P}p_{M,Q}) && \Delta Q\Delta Q\Delta P\\ &+\tfrac{1}{4}(q_{M,}{}^{P}p_{M,Q}{}^{P} - q_{M,Q}{}^{P}p_{M,}{}^{P}) && \Delta Q\Delta P\Delta P\\ &+\tfrac{1}{12}(q_{M,}{}^{P}p_{M,}{}^{PP} - q_{M,}{}^{PP}p_{M,}{}^{P}) && \Delta P\Delta P\Delta P\\ &+\cdots,\end{aligned}\right\}\ \begin{array}{l}p\dot{q}\\ \text{contribution}\end{array}$$

$$\left.\begin{aligned}-\Delta t\,[\ &\tfrac{1}{4}p_M p_{M,QQ} && \Delta Q\Delta Q\\ &+\tfrac{1}{2}p_M p_{M,Q}{}^{P} && \Delta Q\Delta P\\ &+\tfrac{1}{4}p_M p_{M,}{}^{PP} && \Delta P\Delta P\\ &+\cdots]/2m(q_M)\}/\hbar.\end{aligned}\right\}\ \begin{array}{l}\text{Hamiltonian}\\ \text{contribution}\end{array}$$

For the case of a point canonical transformation, consider the generating function $\mathscr{F}(q,P,t) \equiv F(q,Q,t) + QP$. So as shown earlier,

$$Q = \left(\frac{\partial \mathscr{F}}{\partial P}\right)_{qt}, \quad p = \left(\frac{\partial \mathscr{F}}{\partial q}\right)_{Pt}, \quad \mathbf{K} = \mathbf{H} + \left(\frac{\partial \mathscr{F}}{\partial t}\right)_{qP},$$

leading to a general point canonical transformation generated by

$$\mathscr{F}(q,P,t) = Q(q,t)P + G(q,t),$$

i.e.,

$$q = q(Q,t), \quad p = q_{,Q}{}^{-1}(Q,t)P + g(Q,t),$$

g being the inverse of G. Then

$$q_{M,Q}{}^{P}p_{M,QQ} = 0, \quad q_{M,QQ}p_{M,Q}{}^{P} = -q_{M,Q}{}^{-2}q_{M,QQ}{}^{2},$$

$$q_{M,Q}p_{M,QQ} = [2q_{M,Q}{}^{-2}q_{M,QQ}{}^{2} - q_{M,Q}{}^{-1}q_{M,QQQ}]P_M + q_{M,Q}g_{M,QQ},$$

$$q_{M,QQ}p_{M,Q} = -q_{M,Q}{}^{-2}q_{M,QQ}{}^{2}P_M + q_{M,QQ}g_{M,Q},$$

$$p_{M,}p_{M,QQ} = (q_{M,Q}{}^{-1}P_M + g_M)[(2q_{M,Q}{}^{-3}q_{M,QQ}{}^{2} - q_{M,Q}{}^{-2}q_{M,QQQ})P_M + g_{M,QQ}],$$

$$p_M{}^2/2m(q_M) \to (2q_{M,Q}{}^{-1}P_M + g_M)^2/2m(q_M).$$

Carrying only contributing (order Δt) terms $\Rightarrow$

$$
\begin{aligned}
-i\mathbf{V}(Q_M)\Delta t/\hbar = {} & (\tfrac{1}{4})q_{M,Q}{}^{-2}q_{M,QQ}{}^2 && \Delta Q\Delta Q \\
& + (i/12\hbar)[3q_{M,Q}{}^{-2}q_{M,QQ}{}^2 - q_{M,Q}{}^{-1}q_{M,QQQ}]P_M && \Delta Q\Delta Q\Delta Q \\
& - (i\Delta t/8m(q_M)\hbar)[2q_{M,Q}{}^{-4}q_{M,QQ}{}^2 - q_{M,Q}{}^{-3}q_{M,QQQ}]P_M{}^2 && \Delta Q\Delta Q.
\end{aligned}
$$

In this case,

$$
\begin{aligned}
\mathbf{I} = {} & e^{i(F_b - F_a)/\hbar} \lim_{N\to\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j=1}^{N-1} \frac{dQ(j)dP_M(j)}{2\pi\hbar} \prod_{k=1}^{N} (1 + \text{Jacobian stochastic terms} \\
& + i\,\text{action stochastic terms}/\hbar)\exp\left[i\frac{\{P_M(k)(Q(k) - Q(k-1)) + (q_{M,Q}{}^{-2}(k)P_M{}^2(k)/2m(q_M) + \cdots)\Delta t\}}{\hbar} \right]
\end{aligned}
$$

to be compared against

$$
\begin{aligned}
\mathbf{J} = {} & e^{i(F_b - F_a)/\hbar} \lim_{N\to\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j=1}^{N-1} \frac{dQ(j)dP_M(j)}{2\pi\hbar} \prod_{k=1}^{N} \left(1 - \frac{i\mathbf{V}(Q_M(k))\Delta t}{\hbar}\right) \\
& \times\exp\left[i\frac{\{P_M(k)(Q(k) - Q(k-1)) + (q_{M,Q}{}^{-2}(k)P_M{}^2(k)/2m(q_M) + \cdots)\Delta t\}}{\hbar} \right],
\end{aligned}
$$

evaluating the integrals to order Δt (see Appendix, with $\alpha = m(q_M)q_{M,Q}{}^2$, a function of Q,t only) yields

$$
\begin{aligned}
& i\hbar\Delta t\{q_{M,Q}{}^{-4}q_{M,QQ}{}^2 - [3q_{M,Q}{}^{-4}q_{M,QQ}{}^2 - q_{M,Q}{}^{-3}q_{M,QQQ}] \\
& \quad + [2q_{M,Q}{}^{-4}q_{M,QQ}{}^2 - q_{M,Q}{}^{-3}q_{M,QQQ}]\}/4m(q_M) = 0.
\end{aligned}
$$

One might argue, in fact, that this constitutes a proof of invariance of the symmetric path integral under a *general* canonical transformation. This is because, by the group property, a general canonical transformation may be decomposed into successive q and p transformations. The general p transformation is obtained from the general q (point) transformation by following it with an exchange transformation ($q\to P$, $p\to -Q$) which generates no additional stochastic terms.

XI. GENERAL CANONICAL TRANSFORMATION

Since no extra potential term is generated for the symmetric scheme in the transformation between the Hamiltonian and trivial Kamiltonian; using the fact that canonical transformations form a group, one can make a general canonical transformation via the Hamilton–Jacobi transformation. In this way we see indirectly that there is no contribution generated by a general canonical transformation.

The argument leading to this point assumed the use of a physical Hamiltonian [one for which $\mathbf{H}\Delta t$ is of order $(\Delta t)^p$, ($p \geqslant 0$) in the path integral], and here it is understood that only canonical transformations that generate such Hamiltonians are considered.

Having achieved a Hamilton–Jacobi transformation one is led to the path integral

$$
K(q_b,p_b,t_b|q_a,p_a,t_a) = e^{i(\Gamma_b - \Gamma_a)/\hbar}K(Q_b,P_b,t_b|Q_a,P_a,t_a),
$$

where

$$
\begin{aligned}
& K(Q_b,P_b,t_b|Q_a,P_a,t_a) \\
& \quad = \int \cdots \int \prod_{j=1}^{N-1} \frac{dQ(j)dP_M(j)}{2\pi\hbar} \\
& \quad \times \prod_{k=1}^{N} \exp\left[i\frac{\{P_M(k)(Q(k) - Q(k-1))\}}{\hbar} \right] \\
& \quad = 2\pi\hbar\,\delta(Q_b - Q_a)\delta(P_b - P_a)
\end{aligned}
$$

so that the kernel becomes simply

$$
\begin{aligned}
& K(q_b,p_b,t_b|q_a,p_a,t_a) \\
& \quad = 2\pi\hbar\, e^{i(\Gamma_b - \Gamma_a)/\hbar}\delta(Q_b - Q_a)\delta(P_b - P_a),
\end{aligned}
$$

where Γ is a generating function of the Hamilton–Jacobi transformation, and the new coordinates (Q,P) are constant (determined from the end conditions of q,p). This is not a physical quantity (cf. coherent states) and has still to be converted to an amplitude between physical states.

XII. SUMMARY

It would seem that within the symmetric scheme, no additional terms are generated by a general canonical transformation. This was explicitly demonstrated in the case of a point canonical transformation.

This suggests that the symmetric scheme is a canonically invariant prescription (see Fig. 1).

This allows a consistent quantization of a classical system regardless of which canonical variables are used in its description, as well as the use of the trivializing Hamilton–Jacobi transformation in quantum mechanics.

It should perhaps be remarked that the midpoint rule is favored (not compelled) over others, in that it generates no additional terms during the canonical transformation. Note, however, that to get into and from the midpoint scheme, stochastic $\hbar^2$ terms appear. The virture of no terms occurring during the transformation is that the Hamiltonian is then trivalized by the Hamilton–Jacobi transformation. The

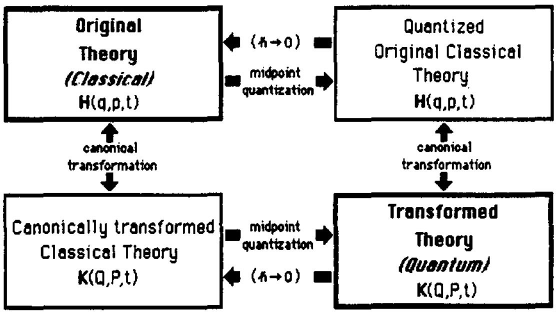


FIG. 1. Consistent quantization.

appearance of stochastic terms during the transformation would spoil this attempt (see Fig. 2).

It is also not clear what a canonical transformation is outside of the midpoint scheme. It still remains to relate this invariant path integral to a physical amplitude.

That this scheme is so well behaved and would seem to be the manner of starting point for making the path integral a well defined mathematical object,[19] tends to indicate that it is a useful portrayal of the path integral.

XIII. THE COHERENT STATE PATH INTEGRAL

Beginning from the coherent state defined by

$$|p,q\rangle \equiv \exp[-(i/\hbar)q\hat{p}]\exp[(i/\hbar)p\hat{q}]|0\rangle$$

(Ref. 15), where $|0\rangle$ denotes the normalized ground state for a harmonic oscillator with unit angular frequency (this definition differs by a phase factor from the common choice). Then

$$\langle p_b,q_b,t_b|p_a,q_a,t_a\rangle$$

becomes, on inserting resolutions of unity,

$$1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|p,q\rangle\langle p,q|\frac{dp\,dq}{2\pi\hbar},$$

$$\langle p_b,q_b,t_b|p_a,q_a,t_a\rangle = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\cdots\prod_{j=1}^{N-1}\frac{dp(j)dq(j)}{2\pi\hbar} \times\prod_{k=1}^{N}\langle p(k),q(k),k|p(k-1),q(k-1),k-1\rangle,$$

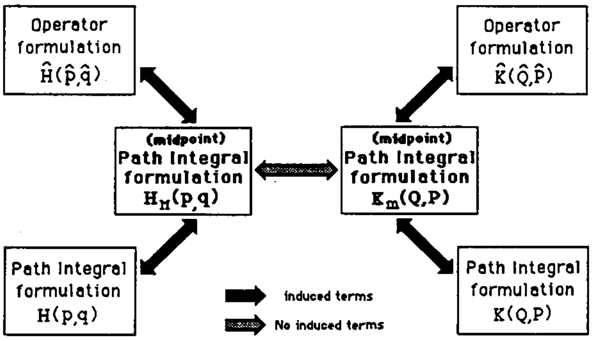


FIG. 2. The midpoint crossing.

where the end points in phase space are not integrated over. Recall that for Heisenberg eigenstates

$$|p,q,t+\Delta t\rangle = \exp[(i/\hbar)\hat{\mathbf{H}}(\hat{q},\hat{p},t+\Delta t/2)\Delta t]|p,q,t\rangle$$

(accurate to order Δt), $\hat{\mathbf{H}}$ being a Hermitian operator, so

$$\langle p_b,q_b,t_b|p_a,q_a,t_a\rangle = \lim_{N\to\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\cdots\prod_{j=1}^{N-1}\frac{dp(j)dq(j)}{2\pi\hbar} \times\prod_{k=1}^{N}\langle p(k),q(k)|\exp\left[-\frac{i}{\hbar}\hat{\mathbf{H}}\left(\hat{q},\hat{p},k-\frac{1}{2}\right)\Delta t\right] \times|p(k-1),q(k-1)\rangle$$

where the end points in phase space are not integrated over.

Proceeding by expanding the exponential, leads to

$$\langle p_b,q_b,t_b|p_a,q_a,t_a\rangle = \lim_{N\to\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\cdots\prod_{j=1}^{N-1}\frac{dp(j)dq(j)}{2\pi\hbar} \times\prod_{k=1}^{N}\langle p(k),q(k)|\sum_{m=0}^{\infty}\frac{(-(i/\hbar)\hat{\mathbf{H}}(\hat{q},\hat{p},k-\frac{1}{2})\Delta t)^m}{m!} \times|p(k-1),q(k-1)\rangle.$$

Taking advantage of the resolution of unity, the operator Hamiltonians may then be freely transferred back and forth across a

$$|p,q\rangle\langle p,q|.$$

For the $m=1$ term this leads to alternative integrands, each therefore equivalent to order Δt, namely,

$$\langle p(k),q(k)|p(k-1),q(k-1)\rangle\langle p(k-1),q(k-1)| \times(-(i/\hbar)\hat{\mathbf{H}}(\hat{q},\hat{p},k-\tfrac{1}{2})\Delta t)|p(k-2),q(k-2)\rangle$$

or

$$\langle p(k),q(k)|(-(i/\hbar)\hat{\mathbf{H}}(\hat{q},\hat{p},k-\tfrac{1}{2})\Delta t)|p(k-1),q(k-1)\rangle \times\langle p(k-1),q(k-1)|p(k-2),q(k-2)\rangle,$$

which equals

$$\langle p(k),q(k)|p(k-1),q(k-1)\rangle \times\langle p(k-1),q(k-1)|p(k-2),q(k-2)\rangle \times(-(i/\hbar)\mathbf{H}(p(k),q(k);\; p(k-1),q(k-1);k-\tfrac{1}{2})\Delta t),$$

where

$$\mathbf{H}(p,q;\rho,g;t) = \langle p,q|\hat{\mathbf{H}}(\hat{q},\hat{p},t)|\rho,g\rangle/\langle p,q|\rho,g\rangle.$$

To evaluate this object one should commute factors in the Hamiltonian operator (using $[\hat{q},\hat{p}]=i\hbar\hat{1}$ which implies $[\hat{a},\hat{a}^\dagger]=\hat{1}$), such that $\hat{a}$ operators are shifted to the right and can be applied to the coherent state, while the $\hat{a}^\dagger$ operators (now on the left) are also suitably applied. This sifting induces additional terms that carry the operator ordering information for the path integral.

The $m=2$ case has three such equivalent alternatives; namely,

$$\tfrac{1}{2}\langle p(k),q(k)|(-(i/\hbar)\hat{\mathbf{H}}(\hat{q},\hat{p},k-\tfrac{1}{2})\Delta t)^2|p(k-1),q(k-1)\rangle\langle p(k-1),q(k-1)|p(k-2),q(k-2)\rangle$$
$$\times\tfrac{1}{2}\langle p(k),q(k)|(-(i/\hbar)\hat{\mathbf{H}}(\hat{q},\hat{p},k-\tfrac{1}{2})\Delta t)|p(k-1),q(k-1)\rangle$$
$$\times\langle p(k-1),q(k-1)|(-(i/\hbar)\hat{\mathbf{H}}(\hat{q},\hat{p},k-\tfrac{1}{2})\Delta t)|p(k-2),q(k-2)\rangle$$
$$\times\tfrac{1}{2}\langle p(k),q(k)|p(k-1),q(k-1)\rangle\langle p(k-1),q(k-1)|(-(i/\hbar)\hat{\mathbf{H}}(\hat{q},\hat{p},k-\tfrac{1}{2})\Delta t)^2|p(k-2),q(k-2)\rangle .$$

This implies (using the $m=1$ results) that these are each, to order Δt, equivalent to

$$\tfrac{1}{2}\langle p(k),q(k)|p(k-1),q(k-1)\rangle\langle p(k-1),q(k-1)|p(k-2),q(k-2)\rangle$$
$$\times(-(i/\hbar)\mathbf{H}(p(k),q(k);p(k-1),q(k-1);k-\tfrac{1}{2})\Delta t)^2.$$

This result then generalizes, by induction, to finite m, leading to

$$\langle p_b,q_b,t_b|p_a,q_a,t_a\rangle=\lim_{N\to\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\cdots\prod_{j=1}^{N-1}\frac{dp(j)dq(j)}{2\pi\hbar}$$

$$\oint\!\Sigma \qquad \times\prod_{k=1}^{N}\langle p(k),q(k)|p(k-1),q(k-1)\rangle\exp\left[-\frac{i}{\hbar}\mathbf{H}\left(p(k),q(k);p(k-1),q(k-1);k-\frac{1}{2}\right)\Delta t\right],$$

where

$$\mathbf{H}(p,q;\rho,g;t)=\langle p,q|\hat{\mathbf{H}}(\hat{q},\hat{p},t)|\rho,g\rangle/\langle p,q|\rho,g\rangle.$$

From the definition of the coherent state

$$|p,q\rangle\equiv\exp[-(i/\hbar)q\hat{p}]\exp[(i/\hbar)p\hat{q}]|0\rangle$$

and the fact

$$[\hat{q},\hat{p}]=i\hbar\hat{1},$$

as well as

$$e^{A+B}=e^Ae^Be^{-[A,B]}$$

if $[A,B]$ commutes with A and B, it then follows that

$$\langle p,q|\rho,g\rangle=\exp\{[(i/2)(p+\rho)(q-g)-\tfrac{1}{4}[(p-\rho)^2+(q-g)^2]]/\hbar\}$$

from which we deduce

$$\langle p_b,q_b,t_b|p_a,q_a,t_a\rangle=\lim_{N\to\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{j=1}^{N-1}\frac{dp(j)dq(j)}{2\pi\hbar}\exp\Big[i\sum_{k=1}^{N}\{\tfrac{1}{2}(p(k)+p(k-1))$$
$$\times(q(k)-q(k-1))+(i/4)[(p(k)-p(k-1))^2+(q(k)-q(k-1))^2]$$
$$-H(p(k),q(k);p(k-1),q(k-1);k-\tfrac{1}{2})\Delta t\}/\hbar\Big],$$

where $\Delta t\equiv(t_b-t_a)/N$.

Now, as shown in the Appendix, in the $\Delta t\to 0$ limit, Δp does not contribute while Δq is of order $(\Delta t)^{1/2}$. Only the term $p^2\Delta t/2m(q,t)$ of the Hamiltonian need be considered, since this is the strongest permissible [order $(\Delta t)^0$] and the only one that is active in the calculation to order Δt. More specifically the term $-(\Delta q)^2/4\hbar$ contributes like a potential of strength $\hbar/4m(q,t)$, where $\tfrac{1}{2}m(q,t)$ is the coefficient of the strongest (i.e., p^2) term of the Taylor expanded Hamiltonian. This leaves conversion of the Hamiltonian to midpoint variables (if not already in this form) to yield the representation

$$\langle p_b,q_b,t_b|p_a,q_a,t_a\rangle=\lim_{N\to\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{j=1}^{N-1}\frac{dp_m(j)dq(j)}{2\pi\hbar}$$
$$\times\exp\left[i\sum_{k=1}^{N}\frac{\{p_m(k)(q(k)-q(k-1))-H(p_m(k),q_m(k),k-\tfrac{1}{2})\Delta t\}}{\hbar}\right],$$

where $\Delta t=(t_b-t_a)/N$. [Displaying the Hamiltonian in midpoint variables has the virtue that it then has the same functional form in the induced differential (Schrödinger) equation for the kernel.[12]]

This object is now identified to be the canonical invariant investigated earlier. To then determine, for example, the point to point amplitude

$$\langle x_b,t_b|x_a,t_a\rangle,$$

insert coherent state resolutions of unity, to yield

$$\langle x_b,t_b|x_a,t_a\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dp_b\,dq_b}{2\pi\hbar}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dp_a\,dq_a}{2\pi\hbar} \times\langle x_b,t_b|p_b,q_b,t_b\rangle\langle p_b,q_b,t_b|p_a,q_a,t_a\rangle \times\langle p_a,q_a,t_a|x_a,t_a\rangle,$$

but from the definition for the coherent state

$$\langle x,t|p,q,t\rangle = (\pi\hbar)^{-1/4} \times\exp[-\{(x-(q+ip))^2+p^2\}/2\hbar],$$

which may be derived by developing and solving for $\langle x|p,q\rangle$ partial differential equations in p and q and using the fact that

$$\langle x|0\rangle = (\pi\hbar)^{-1/4}\exp[-x^2/2\hbar]$$

(Ref. 20), leading to

$$\langle x_b,t_b|x_a,t_a\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dp_b\,dq_b}{2\pi\hbar}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dp_a\,dq_a}{2\pi\hbar} \times(\pi\hbar)^{-1/4}\exp\left[-\frac{\{(x_b-(q_b+ip_b))^2+p_b{}^2\}}{2\hbar}\right] \times\langle p_b,q_b,t_b|p_a,q_a,t_a\rangle \times(\pi\hbar)^{-1/4}\exp\left[-\frac{\{(x_a-(q_a-ip_a))^2+p_a{}^2\}}{2\hbar}\right],$$

but for the special case of Hamilton–Jacobi

$$\langle p_b,q_b,t_b|p_a,q_a,t_a\rangle = 2\pi\hbar\exp[i(\Gamma_b-\Gamma_a)/\hbar]\delta(Q_b-Q_a)\delta(P_b-P_a),$$

where the generator Γ is determined from the Hamilton–Jacobi equation,

$$\mathbf{H}\left(q,\left(\frac{\partial\mathscr{F}}{\partial q}\right)_{Pt},t\right)+\left(\frac{\partial\mathscr{F}}{\partial t}\right)_{qP}=0,$$

where $\Gamma=\mathscr{F}-QP$.

Then

$$\langle x_b,t_b|x_a,t_a\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dP\,dQ}{2(\pi\hbar)^{3/2}} \times\exp\left[-\frac{\{(x_b-(q_b+ip_b))^2+p_b{}^2\}}{2\hbar}\right] \times\exp\left[\frac{i(\Gamma_b-\Gamma_a)}{\hbar}\right] \times\exp\left[-\frac{\{(x_a-(q_a-ip_a))^2+p_a{}^2\}}{2\hbar}\right],$$

where the Hamilton–Jacobi analysis yields

$$q_a(P,Q,t_a),\quad p_a(P,Q,t_a),\quad \Gamma_a(P,Q,t_a),$$
$$q_b(P,Q,t_b),\quad p_b(P,Q,t_b),\quad \Gamma_b(P,Q,t_b);$$

this is illustrated for the case of the simple harmonic oscillator.

Here again, the result of performing the phase space integrals, in general, depends on the sequence in which they are performed. It is understood that if there is any ambiguity, the momentum integrals are to be performed first. How this comes about, and its cure, is discussed elsewhere.[8]

XIV. AN EXPLICIT EXAMPLE: THE SIMPLE HARMONIC OSCILLATOR

$$\mathbf{H}=p^2/2+q^2/2+\hbar/4\quad(m=k=1).$$

The p,q symmetry resulting from this choice of units avoids the ambiguities in the sequence of performing the phase space integrals.[8]

The Hamilton–Jacobi equation for $\mathscr{F}(P,q,t)$, where

$$\Gamma(Q,q,t)=\mathscr{F}-QP$$

and

$$p=\left(\frac{\partial\mathscr{F}}{\partial q}\right)_{Pt},\quad Q=\left(\frac{\partial\mathscr{F}}{\partial P}\right)_{qt}$$

reads[4]

$$\frac{1}{2}\left(\frac{\partial\mathscr{F}}{\partial q}\right)^2_{Pt}+\frac{q^2}{2}+\frac{\hbar}{4}+\left(\frac{\partial\mathscr{F}}{\partial t}\right)_{qP}=0.$$

A solution can be found of the form

$$\mathscr{F}(q,f,t)=W(q,f)-(f+\hbar/4)t,$$

where f is an arbitrary function of P; and W is determined from

$$\frac{1}{2}\left(\frac{\partial W}{\partial q}\right)^2_{Pt}+\frac{q^2}{2}=f,\quad W=\int dq\sqrt{(2f-q^2)},$$

$$\mathscr{F}=\int dq\sqrt{(2f-q^2)}-\left(f+\frac{\hbar}{4}\right)t$$
$$=\sqrt{(2f-q^2)}q/2+f\arcsin(q/\sqrt{(2f)})-\left(f+\frac{\hbar}{4}\right)t$$

[Ref. 21, p. 86, Eq. 2.271(3)], which generates

$$p=\left(\frac{\partial\mathscr{F}}{\partial q}\right)_{Pt}=\sqrt{(2f-q^2)},$$
$$Q=\left(\frac{\partial\mathscr{F}}{\partial P}\right)_{qt}=[\arcsin(q/\sqrt{(2f)})-t]\frac{df}{dP},$$

then unravel

$$q=\sqrt{(2f)}\sin\left(Q\frac{dP}{df}+t\right),$$
$$p=\sqrt{(2f)}\cos\left(Q\frac{dP}{df}+t\right),$$
$$\Gamma=\mathscr{F}-QP=f\sin\left(Q\frac{dP}{df}+t\right)\cos\left(Q\frac{dP}{df}+t\right)+fQ\frac{dP}{df}-QP-\frac{\hbar t}{4},$$

a nonlinear, time dependent canonical transformation. Recall

$$\langle x_b,t_b|x_a,t_a\rangle = \int\int \frac{dP\,dQ}{2\pi\hbar}(\pi\hbar)^{-1/4} \times\exp\left[-\frac{\{(x_b-(q_b+ip_b))^2+p_b{}^2\}}{2\hbar}\right] \times\exp\left[\frac{i(\Gamma_b-\Gamma_a)}{\hbar}\right](\pi\hbar)^{-1/4} \times\exp\left[-\frac{\{(x_a-(q_a-ip_a))^2+p_a{}^2\}}{2\hbar}\right],$$

where the Hamilton–Jacobi analysis yields

$$q_a(P,Q,t_a),\quad p_a(P,Q,t_a),\quad \Gamma_a(P,Q,t_a),$$
$$q_b(P,Q,t_b),\quad p_b(P,Q,t_b),\quad \Gamma_b(P,Q,t_b).$$

For the simple harmonic oscillator, Hamilton–Jacobi leads to

$$\langle x_b,t_b|x_a,t_a\rangle = \int\int\frac{df\,d\theta}{2(\pi\hbar)^{3/2}} \times\exp[-\alpha f+\beta\sqrt{f}+\gamma],$$

where

$$\alpha=(1-\tfrac{1}{2}\exp[-2i(\theta+t_b)]-\tfrac{1}{2}\exp[2i(\theta+t_a)])/\hbar,$$
$$\beta=i\sqrt{2}(x_b\exp[-i(\theta+t_b)]-x_a\exp[i(\theta+t_a)])/\hbar,$$
$$\gamma=-(x_b{}^2+x_a{}^2)/2\hbar-i(t_b-t_a)/4,$$

and $\theta\equiv Q\,dP/df$; the Jacobian of the transformation $P,Q\to f,\ \theta$ being unity. This demonstrates that, in general, the amplitude is independent of the arbitrary function f (now a dummy variable). Translating $\theta\to\theta-(t_b+t_a)/2$ leads to the simplification

$$\alpha=(1-e^{-iT}\cos(2\theta))/\hbar,$$
$$\beta=i\sqrt{2}(x_be^{-i\theta}-x_ae^{i\theta})e^{-iT/2}/\hbar,$$
$$\gamma=-(x_a{}^2+x_b{}^2)/2\hbar-iT/4,$$

where $T\equiv t_b-t_a$.

Let $r=\sqrt{f}$, then

$$\langle x_b,t_b|x_a,t_a\rangle = e^{-iT/4}\int\int\frac{r\,dr\,d\theta}{(\pi\hbar)^{3/2}} \times\exp\left\{\frac{[-(e^{-iT}+1)r^2+2e^{-iT}r^2\cos^2\theta+i\sqrt{2}(x_bre^{-i\theta}-x_are^{i\theta})e^{-iT/2}-\tfrac{1}{2}(x_a{}^2+x_b{}^2)]}{\hbar}\right\},$$

now transform these "cylindrical polar" coordinates to Cartesian,

$$\langle x_b,t_b|x_a,t_a\rangle = e^{-iT/4}(\pi\hbar)^{-3/2}\exp\{[-\tfrac{1}{2}(x_a{}^2+x_b{}^2)]/\hbar\} \times\int_{-\infty}^{\infty}du\,\exp\left\{\frac{[-(1-e^{iT})u^2-i\sqrt{2}(x_a-x_b)e^{-iT/2}u]}{\hbar}\right\} \times\int_{-\infty}^{\infty}dv\,\exp\left\{\frac{[-(1+e^{-iT})v^2+\sqrt{2}(x_a+x_b)e^{-iT/2}v]}{\hbar}\right\},$$

but

$$\int_{-\infty}^{\infty}\exp[-\alpha x^2+\beta x]dx=\sqrt{\left(\frac{\pi}{\alpha}\right)}\exp\left[\frac{\beta^2}{4\alpha}\right].$$

Hence

$$\langle x_b,t_b|x_a,t_a\rangle = e^{iT/4}(2i\pi\hbar\sin T)^{-1/2}\exp[i\{(x_a{}^2+x_b{}^2)\cos T-2x_ax_b\}/(2\hbar\sin T)]$$

where $T\equiv t_b-t_a$ (Refs. 6 and 12), or more precisely

$$\langle x_b,t_b|x_a,t_a\rangle = e^{iT/4}\exp[-i(\pi/4+\pi/2\,\mathrm{Int}(T/\pi))] \times[\sqrt{(2\pi\hbar|\sin T|)}]^{-1}\exp[i\{(x_a{}^2+x_b{}^2)\cos T-2x_ax_b\}/(2\hbar\sin T)]$$

to avoid the root ambiguity, where $\mathrm{Int}(T/\pi)$ is the integer part of T/π.[22,23]

ACKNOWLEDGMENTS

The author is heavily indebted to C. Isham for advice and support throughout most of this toil. Thanks also to J. Dowker and R. Rivers for examination of this work; and to all the many others, who helped shape and make this labor bear fruit.

After "completion" came searching comments from D. Haws, T. Kibble, G. Dunne, and P. Horváthy; helping to clarify much. Without them things would have been different.

For kindly sending a copy of his lecture notes[24] reaching similar conclusions I should like to express my gratitude to J. Klauder.

APPENDIX: WORKING OUT THE GENERAL GAUSSIAN INTEGRALS

By differentiating [w.r.t $\nu(l)$ or $\mu(l)$] the Gaussian integral,

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{j=1}^{N-1}\frac{du(j)dv(j)}{2\pi\hbar}\exp\left[i\sum_{k=1}^{N-1}\frac{\{v(k)u(k)-v^2(k)\Delta t/2\alpha+\nu(k)u(k)+\mu(k)v(k)\}}{\hbar}\right]$$

$$=\exp\left[-i\sum_{k=1}^{N-1}\frac{\{\nu^2(k)\Delta t/2\alpha+\nu(k)\mu(k)\}}{\hbar}\right],$$

one may develop:

Integral

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{j=1}^{N-1}\frac{du(j)dv(j)}{2\pi\hbar},$$

Integrand

$$\times\exp\left[i\sum_{k=1}^{N-1}\frac{\{v(k)u(k)-v^2(k)\Delta t/2\alpha\}}{\hbar}\right].$$

Integrand	Integral (Equivalent integrand)
1	1
$v^n(l)\quad (n\geqslant 1)$	0
$u(l)$	0
$v(l)\,u(l)$	$-(\hbar/i)$
$v^n(l)\,u(l)\quad (n\geqslant 2)$	0
$u^2(l)$	$-(\hbar/i)\Delta t/\alpha$
$v(l)\,u^2(l)$	0
$v^2(l)u^2(l)$	$2(\hbar/i)^2$
$v^n(l)\,u^2(l)\quad (n\geqslant 3)$	0
$u^3(l)$	0
$v(l)\,u^3(l)$	$3(\hbar/i)^2\Delta t/\alpha$
$v^2(l)\,u^3(l)$	0
$v^3(l)\,u^3(l)$	$-6(\hbar/i)^3$
$v^n(l)\,u^3(l)\quad (n\geqslant 4)$	0
	etc.

These expressions continue to be valid to order Δt if α is a function of u,t; since to this order they may be held constant at the midpoint value. Further, $(\partial/\partial\mu(l))_\nu - (\partial/\partial\mu(l-1))_\nu = \Delta t(\partial^2/\partial t\partial\mu)_\nu$ induces a $\Delta v(l)$, and since all currents are held constant (at zero), so the contribution of Δv containing terms is null. It is integrals such as these that tell us that u (ΔQ) contributes with strength order $(\Delta t)^{1/2}$ and $v(P)$ at order $(\Delta t)^{-1/2}$, excepting that for the presence of v's (P's) only [no u's (ΔQ's)] the contribution is zero when the currents are turned off.

From these arguments one concludes that the stochastic terms previously carried are all those that contribute to order Δt.

[1] P. Dirac, "The fundamental equations of quantum mechanics," Proc. R. Soc. London Ser. A **109**, 642 (1925).

[2] P. Dirac, *The Principles of Quantum Mechanics* (Oxford U.P., Oxford, 1958), 4th ed.

[3] P. Chernoff, "Mathematical obstructions to quantization," Hadronic J. **4**, 879 (1981).

[4] H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, MA, 1980), 2nd ed.

[5] R. Feynman, "Space-time approach to quantum mechanics," Rev. Mod. Phys. **20**, 267 (1948) [reprinted in *Quantum Electrodynamics*, edited by J. Schwinger (Dover, New York, 1958)].

[6] R. Feynman and A. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[7] A. Shiekh, "Deriving the path integral from the operator formalism," I.C.T.P. preprint IC/87/47, 1987.

[8] A. Shiekh, "The trivialization of constraints in the quantum theory," I.C.T.P. preprint IC/87/314, 1987.

[9] E. Abers and B. Lee, "Gauge theories," Phys. Rep. C **9**, 5 (1973).

[10] R. Fanelli, "Canonical transformations and phase space path integrals," J. Math. Phys. **17**, 490 (1976).

[11] A. Kapoor, "Quantization in nonlinear coordinates via Hamiltonian path integrals," Phys. Rev. D **29**, 2339 (1984).

[12] L. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).

[13] I. Mayes and J. Dowker, "Canonical functional integrals in general coordinates," Proc. R. Soc. London Ser. A **327**, 131 (1972).

[14] I. Mayes and J. Dowker, "Hamiltonian orderings and functional integrals," J. Math. Phys. **14**, 434 (1973).

[15] J. Klauder, "Path Integrals, Proceedings of the XIX Internationale Universitätswochen für Kernphysik," Schladming, Austria, Acta Phys. Austriaca, Suppl. **XXII**, 3 (1980).

[16] D. McLaughlin and L.Schulman, "Path integrals in curved spaces," J. Math. Phys. **12**, 2520 (1976).

[17] P. Dirac, "Lectures on Quantum Mechanics," Belfer Graduate School of Science, 1964.

[18] J. Gervais and A. Jevicki, "Point canonical transformations in the path integral," Nucl. Phys. B **10**, 93 (1980).

[19] I. Daubechies and J. Klauder, "True measures for real time path integrals," in *Path Integrals from meV to MeV*, edited by M. Gutzwiller, A. Inomata, J. Klauder, and L. Streit (World Scientific, Singapore, 1985), p. 425.

[20] E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1970), 2nd ed.

[21] I. Gradshteyn and I. Ryzhik. *Table of Integrals, Series, and Products* (Academic, New York, 1980), 4th ed.

[22] B. Felsager, *Geometry, Particles and Fields* (Odense U. P., Odense, 1983), 2nd ed.

[23] P. Horváthy, "Extended Feynman formula for harmonic oscillator," Int. J. Theor. Phys. **18**, 245 (1979).

[24] J. Klauder, "Wiener measure regularization for quantum mechanical path integrals" in *Recent Developments in Mathematical Physics*, proceedings of XXVI Internationale Universitätswocher für Kernphysik Schladming, Austria, 1987, edited by H. Mitter and L. Pittner (Springer, Berlin, 1987).

The energy levels and the corresponding normalized wave functions for a model of a compressed atom. II

Staffan Yngve
Institute of Theoretical Physics, University of Uppsala, Thunbergsvägen 3, S-752 38 Uppsala, Sweden

(Received 23 January 1987; accepted for publication 25 November 1987)

Recently Fröman *et al.* [J. Math. Phys. **28**, 1813 (1987)] obtained a simple formula for the energy shift due to the compression of an atom. In the model considered the usual condition (associated with the corresponding uncompressed atom) that the wave function must vanish at $r=\infty$ is replaced by the boundary condition that the wave function must have a node at the finite distance $r=a$. The treatment of the problem of obtaining the energy shift due to the compression is based on the phase-integral method developed by Fröman and Fröman. In the present work the formula for the energy shift is generalized to a possible additional choice of the base function in the phase-integral treatment. The normalized wave function close to the origin is also obtained. As an application, the resulting general formula for the energy shift is particularized to the case of a hydrogenic atom and a numerical illustration of the accuracy of the formula for the two alternative choices of the base function is given.

I. INTRODUCTION

In a previous paper,[1] which forms part I of the present work, and which is henceforth denoted by I, Fröman *et al.* studied a model of a compressed atom, which was first introduced in 1937 for the case of the hydrogen atom by Michels *et al.*[2] Problems concerning confined atoms have been studied by many authors since then. For a comprehensive discussion of representative works up to 1983 see I, and as an example of the continuing interest in the problem of a compressed atom see Ref. 3. The aim of I and the present paper is to derive simple, accurate analytical results for a system consisting of a nonrelativistic quantal particle bound in an unspecified, smooth, spherically symmetric, single-well potential, which is enclosed in a large, impenetrable sphere with the aid of the phase-integral method developed by Fröman and Fröman. For the general background of that method we refer to I and the references given therein. It is important to note that the phase-integral approximations used in that method are more general than the JWKB approximations since they contain an unspecified function $Q(r,E)$, the base function, which can be chosen in a way appropriate for the problem under consideration.

Our model problem involves the boundary condition that the wave function must have a node at the finite distance $r=a$. An analogous model situation occurs for odd parity states in a symmetric double oscillator, a problem which has been treated rigorously by Fröman[4] with the aid of the phase-integral method. In Sec. II, for our model problem, Fröman's solution of the double-oscillator problem is exploited to obtain the energy shift due to compression. The resulting formula agrees with a formula obtained in I in a more rigorous way. In I the base function $Q(r,E)$ is chosen such that the phase-integral approximations are valid also when r tends to zero. In the present investigation the base function $Q(r,E)$ may either be chosen as in I or according to condition (8a) in Ref. 5.

The introduction of an alternative choice of the base function $Q(r,E)$ in the present paper enables us in Sec. III to generalize the formula for the normalized wave function close to the origin obtained by Fröman *et al.*[5] to the present boundary value problem.

In Sec. IV the phase-integral formula for the energy shift due to compression is applied to the particular case of a compressed hydrogenic atom and the accuracy of the formula thus obtained is demonstrated numerically. For either choice of the base function mentioned above, a considerable improvement of the accuracy is obtained by taking the simple third-order correction into account.

Consider a nonrelativistic particle moving in a spherically symmetric potential. The state considered corresponds to the angular momentum quantum number l and the energy E. The radial Schrödinger equation is written as

$$\frac{d^2\psi}{dr^2}+R(r,E)\psi=0, \tag{1}$$

where, with obvious notations,

$$R(r,E)=(2m/\hbar^2)(E-V(r)-\hbar^2 l(l+1)/2mr^2). \tag{2}$$

To obtain a useful approximate expression for the radial wave function, we shall use the phase-integral method, involving phase-integral approximations of arbitrary order, devised by Fröman and Fröman (see I and the references given therein). In the context of the present paper there are two distinct situations and two main alternatives, which we denote cases A and B, for choosing the base function $Q(r,E)$ in these phase-integral approximations. As our first alternative (case A) we choose the last-mentioned function such that the phase-integral approximations are valid also when r tends to zero, which is achieved if we impose the condition[6,7]

$$\lim_{r\to 0} r^2(Q^2(r,E)-R(r,E))=-\tfrac{1}{4},\quad \text{case A}. \tag{3}$$

If the potential behaves as $-Z\hbar^2/(ma_0 r)$ close to $r=0$, where a_0 is equal to the Bohr radius and where the atomic number Z is assumed to be positive, we may as an alternative choose $Q(r,E)$ as in Ref. 5 such that

$$\lim_{r\to 0} r^2(Q^2(r,E)-R(r,E))=l(l+1),\quad \text{case B}. \tag{4}$$

 0022-2488/88/040931-06\$02.50

A possible choice in each case is

$$Q^2(r,E) = R(r,E) - (1/4r^2)$$
$$= \frac{2m}{\hbar^2}(E - V(r)) - \frac{(l+\frac{1}{2})^2}{r^2}, \quad \text{case A}, \tag{5a}$$

and

$$Q^2(r,E) = R(r,E) + l(l+1)/r^2$$
$$= (2m/\hbar^2)(E - V(r)), \quad \text{case B}. \tag{5b}$$

For the phase-integral approximations of the order $2N+1$, we have

$$q(r,E) = Q(r,E) \sum_{\nu=0}^{N} Y_{2\nu}, \tag{6}$$

with the first few quantities $Y_{2\nu}$ being

$$Y_0 = 1, \tag{7a}$$

$$Y_2 = \tfrac{1}{2}\epsilon_0, \tag{7b}$$

$$Y_4 = -\frac{1}{8}\epsilon_0^2 - \frac{1}{8}\left(\frac{1}{Q(r,E)}\frac{d}{dr}\right)^2 \epsilon_0, \tag{7c}$$

with

$$\epsilon_0 = \frac{R - Q^2}{Q^2} + \frac{1}{16Q^6}\left[5\left(\frac{dQ^2}{dr}\right)^2 - 4Q^2\frac{d^2Q^2}{dr^2}\right]. \tag{8}$$

Explicit expressions for $Y_{2\nu}$ up to Y_8 are given in Refs. 8 and 9 and for up to Y_{20} in Ref. 10.

II. ENERGY LEVELS FOR THE UNCOMPRESSED AND COMPRESSED ATOMS

In the treatment of the radial Schrödinger equation (1) with (2) for an atomic electron in a free, uncompressed atom, the boundary conditions are $u(0) = u(\infty) = 0$. If, however, the atom is enclosed in a sphere of radius a, the boundary conditions are instead $u(0) = u(a) = 0$. This change of boundary conditions causes a shift upward of energy level. This effect has been studied theoretically during several decades by many authors: see I and the references given therein.

We introduce the definitions

$$L(E) = \frac{1}{2}\int_\Gamma q(r,E)dr, \quad \text{cases A and B}, \tag{9}$$

and

$$K(r_2,E) = \frac{1}{2}i\int_{\Gamma(r_2)} q(r,E)dr, \quad \text{cases A and B}, \tag{10}$$

where in (9) Γ is the closed contour of integration pertinent to the case one is treating, i.e., Γ stands for Γ_A or Γ_B depicted in Fig. 1(a) and 1(b), respectively, together with $\Gamma(r_2)$, and where $q(r,E)$ is given by (6) with (7) and (8). It should be noted that r_2 is a point in the classically forbidden region.

For the uncompressed atom, i.e., for infinite values of a, the approximate quantization conditions pertaining to cases A and B are given by Eq. (19) with (20a) and (20b), respectively, in Ref. 11, which in the notation of I read as

$$L(E_s^\infty) = (s+\tfrac{1}{2})\pi, \quad \text{case A}, \tag{11a}$$

$$L(E_s^\infty) = (s+l+1)\pi, \quad \text{case B}, \tag{11b}$$

and where $s = 0,1,2,...$ is the radial quantum number. In case B the function $Q^2(r,E)$ has a first-order pole at $r = 0$ and a first-order zero on the positive real r axis and the quantization condition (11b) differs from the generalized Bohr–Sommerfeld quantization condition (11a), which applies to a smooth function $Q^2(r,E)$ with two zeros.

In our model the compression of the atom corresponds to moving the node of the wave function in the classically forbidden region from $r = \infty$ to $r = a$. The effect on each energy level of such a change in boundary condition can be understood by means of the following analogy. Consider a one-dimensional symmetric double oscillator symmetric around $r = a$. For convenience we denote the variable by r, although it may be extended to $-\infty$. The square of the base function with reversed sign is illustrated in Fig. 1 of Ref. 12, where, however, the center of symmetry should be moved from the origin to $r = a$ (cf. Fig. 2 of Ref. 12) and where the

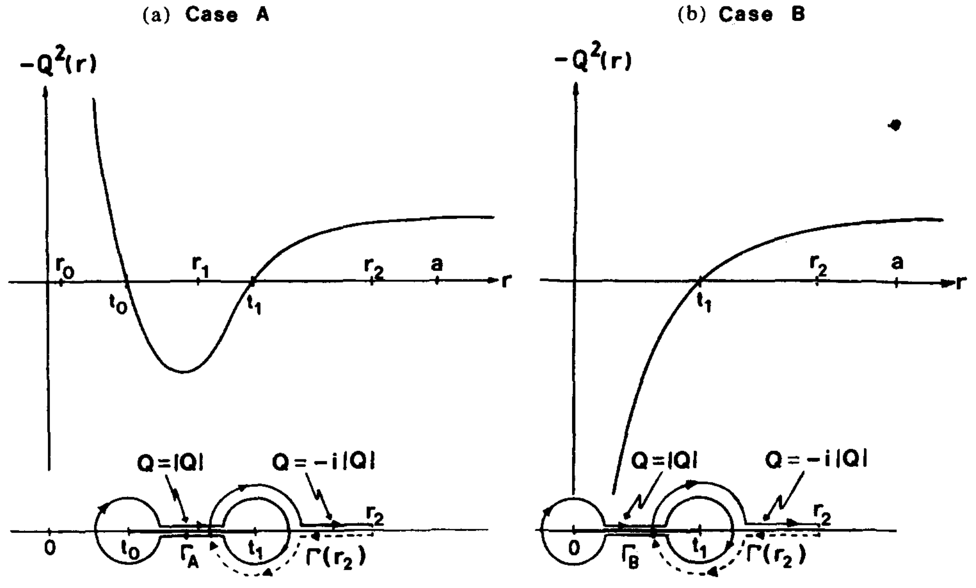


FIG. 1. (a) and (b) show for real values of r the qualitative behavior of $-Q^2(r)$ in cases A and B, respectively. The cuts in the complex r plane, introduced in order to make $Q(r)$ single-valued, are indicated by bold lines; the contours of integration occurring in the formulas are also depicted. The contour Γ_A is a closed loop encircling the two classical turning points t_0 and t_1. The contour Γ_B is a closed loop encircling the origin and the classical turning point t_1. The contour $\Gamma(r_2)$ is a nonclosed contour which goes from a real value r_2 on the lower Riemann sheet, passes around the classical turning point t_1, and returns to the corresponding real value r_2 on the upper Riemann sheet. The phase of $Q(r)$, which is also indicated, is chosen such that $Q(r) = |Q(r)|$ on the upper lip of the cut in the classically allowed region, which implies that the contour integrals over Γ_A and Γ_B are positive.

turning points from left to right are denoted by t_0, t_1, $2a - t_1$, and $2a - t_0$ in our notation. For this symmetric double oscillator an odd-parity bound state solution of the Schrödinger equation (1) has a node at $r = a$ and according to Eq. (51) in Ref. 3 (rewritten to suite our notation), the energy levels for odd-parity states are obtained from the approximate quantization condition

$$L(E_s^{\text{odd}}) - \tfrac{1}{2}\exp\{-2K(a,E_s^{\text{odd}})\} = (s + \tfrac{1}{2})\pi,$$
$$s = 0,1,2,\ldots . \quad (12)$$

In (12) L and K are defined by (9) and (10), respectively, with the contours of integration $\Gamma(=\Gamma_A)$ and $\Gamma(r_2)$ illustrated in the lower part of Fig. 1(a) where $r = a$ is taken as the point of symmetry for the double oscillator.

In Ref. 3 the square of the base function is chosen equal to $R(r,E)$ and formula (12) is derived under the assumption that $Q^2(r,E)$ is smooth and may be continued analytically into the complex plane. We shall now assume that (12) is valid also when there is a possible small discontinuity of the derivative of $Q^2(r,E)$ at the center of symmetry $r = a$. If we construct a symmetric double oscillator by using a simple oscillator from $-\infty$ to a and joining this simple oscillator by the function (defined from a to $+\infty$) obtained by reflection in the line $r = a$ of the original potential function, a discontinuity of the derivative at $r = a$ will, in general, occur (cf. Fig. 2 in Ref. 12).

We conclude that (12) is an approximate quantization condition for a confined simple oscillator for which the wave function has a node at $r = a$. One can easily generalize (12) to the case of a radial problem with two turning points t_0 and t_1. Denoting by E_s^a the eigenvalue of the boundary value problem (1) with boundary conditions $u(0) = u(a) = 0$, for which obviously

$$\lim_{a\to\infty} E_s^a = E_s^\infty, \quad (13)$$

we hence in the two-turning-point case obtain E_s^a from the approximate quantization condition (12) with E_s^{odd} replaced by E_s^a. Subtracting the resulting quantization condition from (11a), we obtain the approximate formula

$$\begin{aligned} L(E_s^a) - L(E_s^\infty) &= \tfrac{1}{2}\exp\{-2K(a,E_s^a)\} \\ &= \tfrac{1}{2}\exp\{-2K(a,E_s^\infty)\}, \quad \text{cases A and B.} \end{aligned} \quad (14)$$

By approximating E_s^a by E_s^∞ in the second member of (14) we introduce only a very small error. It should be noted that E_s^a and E_s^∞ in general depend on l.

Introducing into (14) the approximate formula

$$L(E_s^a) - L(E_s^\infty) = (E_s^a - E_s^\infty)\frac{dL(E_s^\infty)}{dE_s^\infty} \quad (15)$$

and replacing in the resulting equation in accordance with the quantization condition (11a) for case A and (11b) for case B $dL(E_s^\infty)/dE_s^\infty$ by $\pi\, ds/dE_s^\infty$, i.e., by $\pi/(dE_s^\infty/ds)$, we obtain the approximate formula for the energy shift due to compression:

$$E_s^a - E_s^\infty = \frac{1}{2\pi}\frac{dE_s^\infty}{ds}\exp\{-2K(a,E_s^\infty)\}, \quad (16)$$

where dE_s^∞/ds can be obtained, e.g., from spectroscopic data for the uncompressed atom.

For a rigorous derivation (in case A) of (16) see I.

III. GENERALIZED FERMI–SEGRÈ FORMULA

The normalized radial wave function of the electron in the model of a compressed atom can be written as

$$\psi_{\text{norm}}(r,E_s^a) = \psi(r,E_s^a)\left(\int_0^a \psi^2(r,E_s^a)dr\right)^{-1/2}, \quad (17)$$

where $\psi(r,E_s^a)$ is the real, unnormalized bound-state solution of (1) with (2) corresponding to the eigenvalue E_s^a, i.e., $\psi(r,E_s^a)$ vanishes for $r = 0$ and $r = a$. The integral appearing in (17) can be expressed in a convenient form by means of a method devised by Furry[13] and simplified by the present author.[14] The general formula given in Ref. 14 and adapted to the present situation is, according to (4.6) with (4.5) in I,

$$\int_0^a [\psi(r,E_s^a)]^2 dr = \frac{\hbar^2}{2m}\left\{\frac{\partial}{\partial E}\left(\psi\frac{d\bar\psi}{dr} - \bar\psi\frac{d\psi}{dr}\right)\right\}_{E=E_s^a}. \quad (18)$$

Formula (18) is an exact relation provided that ψ and $\bar\psi$ are exact solutions of (1) which, for any value of E, have the properties that $\psi\to 0$ as $r\to 0$ and $\bar\psi\to 0$ as $r\to a$, and that, furthermore, $\psi(r,E_s^a) = \bar\psi(r,E_s^a)$. It should be noted that the expression on the right-hand side of (18) is independent of r since the Wronskian $\psi\, d\bar\psi/dr - \bar\psi\, d\psi/dr$ is independent of r.

To obtain useful approximate expressions for the radial wave functions ψ and $\bar\psi$, which appear in (17) and (18), we shall use the phase-integral approximations mentioned previously. In the first half of the present section we shall regard case B and as in Ref. 5 we shall assume that $R(r,E)$ close to $r = 0$ is represented by the expression

$$R(r,E) = -\frac{l(l+1)}{r^2} + \frac{2Z}{a_0 r} + a_0^{-2}\sum_{n=0}^\infty b_n\left(\frac{r}{a_0}\right)^n, \quad (19)$$

where the quantities b_n are dimensionless. In the derivations preceding formula (26) we further assume that the square of the base function Q^2 has precisely one zero t_1 on the positive real r axis, that this zero is well separated from the pole at $r = 0$, and that no other zeros of Q^2 lie on or close to the positive real r axis. According to the results obtained in Ref. 15 (pp. 74–79) for the first-order phase-integral approximation the particular solution $\psi(r,E)$, for which $\psi/r^{l+1}\to 1$ when $r\to 0$ for any value of E, is given by the approximate formula (7.28) in Ref. 15, which reads as

$$\psi(r,E) = (\pi c_l)^{-1/2}Q^{-1/2}(r,E) \times\cos\left(\int_0^r Q(r,E)dr - \left(l + \frac{3}{4}\right)\pi\right), \quad (20)$$

where

$$c_l = \operatorname*{Res}_{r=0}[\psi(r,E)]^{-2},$$
when $2l + 1$ is a non-negative integer, (21)

and where r is a point lying in the interior (i.e., sufficiently far away from both end points) of the interval $0<r<t_1$, with t_1 the zero of $Q^2(r,E)$. For higher-order phase-integral approximations see (12a) in Ref. 5.

In the limit as $a\to\infty$ the solution $\bar\psi(r,E)$ of the radial Schrödinger equation is in the first-order approximation given by formula (8) in Ref. 16. For $a<\infty$ there will be a phase correction and the solution $\bar\psi(r,E)$ of the radial Schrödinger equation (1) which vanishes at $r=a$ (for any value of E) is approximately

$$\bar\psi(r,E)=C_lQ^{-1/2}(r,E)\times\cos\left[\int_r^{t_1}Q(r,E)dr-\frac{1}{4}\pi-\Phi_l\right] \quad (22)$$

when $0<r<t_1$ (first-order approximation). Here Φ_l is a phase correction (independent of r) which tends to zero in the limit as $a\to\infty$. The requirement that the functions $\psi(r,E)$ and $\bar\psi(r,E)$ shall be equal for $E=E_s^a$ gives, by means of (20) and (22), the quantization condition

$$\int_0^{t_1}Q(r,E_s^a)dr=(s+l+1)\pi+\Phi_l \quad (23)$$

and the condition that

$$C_l=(-1)^s/(\pi c_l)^{1/2},\quad \text{when } E=E_s^a. \quad (24)$$

The approximate solutions ψ and $\bar\psi$ given by (20)–(22) satisfy the conditions required for the approximate validity of (18) if the conditions (23) and (24) are fulfilled; hence we can obtain an approximate expression for the normalization integral. In fact, from (20)–(22) and (24) it follows that

$$\psi\frac{d\bar\psi}{dr}-\bar\psi\frac{d\psi}{dr}=\frac{(-1)^s}{\pi c_l}\sin\left(\int_0^{t_1}Q(r,E)dr-(l+1)\pi-\Phi_l\right) \quad (25)$$

and when this expression is inserted into (18) and use is made of (23), we obtain

$$\int_0^a\psi^2(r,E_s^a)dr=\frac{\hbar^2}{2\pi mc_l}\left(\frac{\partial}{\partial E}\left[\int_0^{t_1}Q(r,E)dr-\Phi_l\right]\right)_{E=E_s^a}. \quad (26)$$

Expression (26) can, with the aid of the quantization condition (23), be rewritten as

$$\int_0^a\psi^2(r,E_s^a)dr=\frac{\hbar^2}{2m\,c_l}\left(\frac{dE_s^a}{ds}\right)^{-1}. \quad (27)$$

The same final formula (27) would have been obtained if we had used phase-integral approximations of higher order.

Inserting (27) into (17) and recalling that

$$\psi(r,E_s^a)/r^{l+1}\to 1 \quad \text{as } r\to 0,$$

we arrive at the formula

$$\lim_{r\to 0}\left(\frac{\psi_{\text{norm}}(r,E_s^a)}{r^{l+1}}\right)=\left[\frac{2m\,c_l}{\hbar^2}\frac{dE_s^a}{ds}\right]^{1/2}, \quad (28)$$

where, according to Ref. 5 for integral values of l,

$$c_l=\frac{(2Z/a_0)^{2l+1}}{[(2l+1)!]^2}\left\{\prod_{j=0}^{l}\left[1+\left(\frac{b_0}{Z^2}-\frac{l(l+1)}{2}\frac{b_1}{Z^3}\right)j^2+\frac{3}{2}\frac{b_1}{Z^3}j^4\right]+R_l\right\}, \quad (29)$$

wherein j takes all non-negative integral values up to l and where

$$R_l=0,\quad \text{for } l=0 \text{ and } 1, \quad (30a)$$

$$R_l=18\left[\frac{b_2}{Z^4}+\frac{b_3}{Z^5}+\left(\frac{b_1}{Z^3}\right)^2\right],\quad \text{for } l=2, \quad (30b)$$

$$R_l=45\left[6\frac{b_2}{Z^4}+22\frac{b_3}{Z^5}+13\left(\frac{b_1}{Z^3}\right)^2\right]+675\left[4\frac{b_4}{Z^6}+\frac{b_1}{Z^3}\frac{b_2}{Z^4}+6\frac{b_5}{Z^7}\right]+45\frac{b_0}{Z^2}\left[14\frac{b_2}{Z^4}+18\frac{b_3}{Z^5}+27\left(\frac{b_1}{Z^3}\right)^2\right],\quad \text{for } l=3. \quad (30c)$$

It should be noted that formula (29) with (30) is exact. The structure of formula (28) as compared, e.g., to the more complicated formula for the uncompressed atom obtained by Durand and Durand[17] should be noted. The quantity c_l is determined entirely from the local properties of the function $R(r,E)$ close to $r=0$, whereas dE_s^a/ds is a global quantity that can be obtained, e.g., by means of spectroscopic data.

It should be noted that formula (28) is obtained formally from the generalized Fermi–Segrè formula (18) in Ref. 5 by replacing the derivative of the energy eigenvalue with respect to the quantum number by dE_s^a/ds and replacing the energy eigenvalue by E_s^a in the expression for c_l. For $l=0$ formula (28) with (29) and (30) is recognized as the usual Fermi–Segrè formula[16] with dE_n/dn replaced by dE_n^a/dn. Since c_0 does not depend on the energy, for s states the change of the normalized wave function due to compression depends entirely on the difference between $(dE_s^a/ds)^{1/2}$ and $(dE_s^\infty/ds)^{1/2}$.

If dE_s^a/ds is unknown, it can be calculated from (14) and (11) as

$$\frac{dE_s^a}{ds}=\frac{\pi}{((\partial/\partial E)[L(E)-\frac{1}{2}\exp\{-2K(a,E)\}])_{E=E_s^a}},\quad \text{cases A and B}, \quad (31)$$

where L and K are obtained from (9) and (10). It should be noted, however, that the introduction of (31) into (28) may, in case B, deteriorate the accuracy of (28) for higher values of l. By calculating the energy eigenvalue (and dE_s^a/ds) for the optimal alternative A or B, (28) with (29)–(31) constitutes a formula for the normalized wave function at the origin, which should be sufficiently accurate for many purposes.

IV. APPLICATION TO THE CASE OF A HYDROGENIC ATOM

In the application to the case of a hydrogenic atom the potential $V(r)$ in (2) is the attractive Coulomb potential

$$V(r)=-Z\hbar^2/ma_0r. \quad (32)$$

We shall choose $Q^2(r,E)$ according to (5a) with (32) in case A and according to (5b) with (32) in case B. Inserting the respective $Q(r,E)$ with the phase shown in Fig. 1 into (9) with (6)–(8), we obtain

$$L(E) = [([Z^2\hbar^2/(ma_0^2)]/-2E)^{1/2} - (l+\tfrac{1}{2})]\pi, \quad \text{case A}, \tag{33a}$$

$$L(E) = ([Z^2\hbar^2/(ma_0^2)]/-2E)^{1/2}\pi, \quad \text{case B}. \tag{33b}$$

We realize immediately that for the free hydrogenic atom (11a) with (34a), as well as (11b) with (34b), yield the well-known exact formula

$$E_s^\infty = -[Z^2\hbar^2/(ma_0^2)]/2n^2, \tag{34}$$

where

$$n = s + l + 1. \tag{35}$$

Introducing the notation

$$E_s^a = -[Z^2\hbar^2/(ma_0^2)]/2n_a^2, \tag{36}$$

where n_a is the "effective principal quantum number" for the compressed hydrogenlike atom, we obtain from (14) and (33)–(36) the formula

$$n_a - n = (1/2\pi)\exp\{-2K(a,E_s^\infty)\}, \quad \text{cases A and B}. \tag{37}$$

We shall evaluate the quantity $\exp\{-2K(a,E_s^\infty)\}$, with $K(a,E_s^\infty)$ defined by (10) and (6)–(8) up to the third-order approximation and with E_s^∞ given by (34). As in I we put

$$\tilde{r} = Zr/a_0 \tag{38}$$

and introduce expression (34) for E_s^∞ into the alternative expressions (5a) with (32) and (5b) with (32) for $Q^2(r,E)$, obtaining in cases A and B, respectively,

$$Q(r,E_s^\infty)dr = \left[-\frac{1}{n^2} + \frac{2}{\tilde{r}} - \frac{(l+\frac{1}{2})^2}{\tilde{r}^2}\right]^{1/2} d\tilde{r}, \quad E = E_s^\infty, \quad \text{case A}, \tag{39a}$$

$$Q(r,E_s^\infty)dr = \left[-\frac{1}{n^2} + \frac{2}{\tilde{r}}\right]^{1/2} d\tilde{r}, \quad E = E_s^\infty, \quad \text{case B}. \tag{39b}$$

It can easily be shown with the aid of (39a), (39b), (2), (32), and (6)–(8) that the functions $Y_{2\nu}$ for the energy E_s^∞ can be expressed in terms of the variable $\tilde{r}$ and the parameters n and l. Hence $K(a,E_s^\infty)$ defined by (10) can be expressed in terms of the parameters n, l, and

$$\tilde{a} = Za/a_0. \tag{40}$$

For the quantity $a|Q(a,E_s^\infty)|$ appearing in our final formula

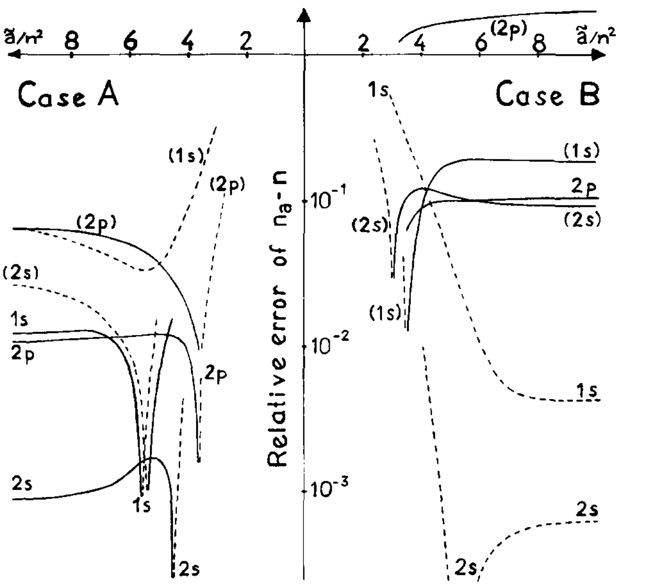


FIG. 2. The relative error of $n_a - n$ is plotted against $\tilde{a}/n^2 = Za/(a_0 n^2)$ for the states $1s$ ($n=1$, $l=0$), $2s$ ($n=2$, $l=0$), and $2p$ ($n=2$, $l=1$). The left-hand figure is based on the choice

$$a_0^2 Q^2(r,E_s^a) = -n_a^{-2} + 2Za_0/r - (l+\tfrac{1}{2})^2 a_0^2/r^2$$

and the right-hand figure is based on the choice

$$a_0^2 Q^2(r,E_s^a) = -n_a^{-2} + 2Za_0/r,$$

where in both cases we have $n_a = n$ for the uncompressed hydrogenlike atom. The solid lines correspond to a positive error and the broken lines correspond to a negative error. It should be mentioned that the cusps in the figure actually correspond to a relative error equal to zero, although for practical reasons this is not seen. Here (1s), (2s), and (2p) on the respective curves indicate the first-order phase-integral approximation for the respective states and 1s, 2s, and 2p on the respective curves indicate the third-order phase-integral approximations.

(43) for $K(a,E_s^\infty)$ we easily obtain the following expressions for cases A and B, respectively:

$$a|Q(a,E_s^\infty)| = [\tilde{a}^2/n^2 - 2\tilde{a} + (l+\tfrac{1}{2})^2]^{1/2}, \quad a > t_1, \quad \text{case A}, \tag{41}$$

$$a|Q(a,E_s^\infty)| = [\tilde{a}^2/n^2 - 2\tilde{a}]^{1/2}, \quad a > t_1, \quad \text{case B}. \tag{42}$$

With the aid of (6)–(8) and (39)–(42) and with due regard to the contours of integration shown in Fig. 1, we obtain from (10) with $r_2 = a > t_1$, after the resulting integrals have been evaluated in the first and third order,

$$\exp\{-2K(a,E_s^\infty)\} = \exp\{-2(K^{(1)} + K^{(3)} + \cdots)\}, \quad \text{cases A and B}, \tag{43}$$

where for case A we have

$$\exp\{-2K^{(1)}\} = \left(\frac{\tilde{a} - n^2 + na|Q(a)|}{\tilde{a} - n^2 - na|Q(a)|}\right)^n \left(\frac{\tilde{a} - (l+\frac{1}{2})^2 - (l+\frac{1}{2})a|Q(a)|}{\tilde{a} - (l+\frac{1}{2})^2 + (l+\frac{1}{2})a|Q(a)|}\right)^{l+1/2} \exp\{-2a|Q(a)|\}, \quad E = E_s^\infty, \quad a > t_1, \tag{44a}$$

$$K^{(3)} = [n^2 - (l+\tfrac{1}{2})^2]^{-1}[a|Q(a)|]^{-3}\tfrac{1}{24}[-\tilde{a}^3/n^2 + 6\tilde{a}^2 - 3(l+\tfrac{1}{2})^2\tilde{a}^2/n^2 - 3n^2\tilde{a} - n^2(l+\tfrac{1}{2})^2 + 2(l+\tfrac{1}{2})^4], \tag{44b}$$

and for case B we have

$$\exp\{-2K^{(1)}\} = [(\tilde{a} - n^2 + na|Q(a)|)/(\tilde{a} - n^2 - na|Q(a)|)]^n \exp\{-2a|Q(a)|\}, \quad E = E_s^\infty, \quad a > t_1, \tag{45a}$$

$$K^{(3)} = [a|Q(a)|]^{-3}\tfrac{1}{24}[2\tilde{a}^3/n^4 - 6\tilde{a}^2/n^2 + 9\tilde{a}] + \tfrac{1}{2}l(l+1)a|Q(a)|/\tilde{a}. \tag{45b}$$

In Fig. 2 the accuracy of formula (37) with (43) is illustrated for both cases A and B, i.e., with the use of (44) and (45), respectively. The approximations deteriorate and eventually break down as a approaches the zero t_1 of $Q^2(r,E)$ on the positive real r axis. We see that the choice of $Q^2(r,E)$ as in case B (corresponding to the omission of the centrifugal barrier in the first order of approximation) gives within its region of validity a good accuracy when the third-order correction is included. It should be noted, however, that this modification cannot in general be used for high values of l. For the choice of $Q^2(r,E)$ as in case A, which in the first-order approximation corresponds to the replacement of $l(l+1)$ by $(l+\frac{1}{2})^2$, we obtain a considerable numerical improvement by taking the simple third-order correction into account.

ACKNOWLEDGMENTS

The author is very grateful to Professor N. Fröman and Professor P. O. Fröman for suggesting the investigation and would like to thank Anders Hökback and Ulf Sundström for helpful assistance with the numerical calculations.

[1]P. O. Fröman, S. Yngve, and N. Fröman, J. Math. Phys. **28**, 1813 (1987).
[2]A. Michels, J. De Boer, and A. Bijl, Physica **IV**, 981 (1937).
[3]M. Friedman, A. Rabinovitch, and R. Thieberger, Z. Phys. A**316**, 1 (1984).
[4]N. Fröman, Ark. Fys. **32**, 79 (1966).
[5]S. Yngve, Phys. Rev. A **33**, 96 (1986).
[6]N. Fröman and P. O. Fröman, J. Phys. (Paris) **42**, 1491 (1981).
[7]S. Yngve, Phys. Rev. D **31**, 2554 (1985).
[8]N. Fröman and P. O. Fröman, Nucl. Phys. A **147**, 606 (1970).
[9]S. Yngve, J. Math. Phys. **12**, 114 (1971).
[10]J. A. Campbell, J. Comput. Phys. **10**, 308 (1972).
[11]N. Fröman, Phys. Rev. A **17**, 493 (1978).
[12]S. Yngve, Am. J. Phys. **54**, 1103 (1986).
[13]W. H. Furry, Phys. Rev. **71**, 360 (1947).
[14]S. Yngve, J. Math. Phys. **13**, 324 (1972).
[15]N. Fröman and P. O. Fröman, *JWKB Approximation, Contributions to the Theory* (North-Holland, Amsterdam, 1965) [Russian translation (MIR, Moscow, 1967)].
[16]N. Fröman and P. O. Fröman, Phys. Rev. A **6**, 2064 (1972).
[17]B. Durand and L. Durand, Phys. Rev. A **33**, 2899 (1986).

Petrov types D and II perfect-fluid solutions in generalized Kerr–Schild form

F. Martín-Pascual and J. M. M. Senovilla[a)]
Departamento de Física Teórica, Universidad de Salamanca, 37008, Salamanca, Spain

(Received 24 April 1986; accepted for publication 2 December 1987)

Petrov types D and II perfect-fluid solutions are obtained starting from conformally flat perfect-fluid metrics and by using a generalized Kerr–Schild ansatz. Most of the Petrov type D metrics obtained have the property that the velocity of the fluid does not lie in the two-space defined by the principal null directions of the Weyl tensor. The properties of the perfect-fluid sources are studied. Finally, a detailed analysis of a new class of spherically symmetric static perfect-fluid metrics is given.

I. INTRODUCTION

In a previous paper,[1] the first results concerning perfect-fluid solutions of Einstein's equations in generalized Kerr–Schild form were given. The generalized Kerr–Schild metrics have the following form:

$$\tilde{g}_{\alpha\beta} = g_{\alpha\beta} + 2Hl_\alpha l_\beta, \tag{1}$$

where $g_{\alpha\beta}$ is the metric of any space-time, H is a scalar field, and l_α is a null geodesic vector field for both metrics g and $\tilde{g}$.

As is well known, the Kerr–Schild ansatz has been a powerful tool in searching for solutions of Einstein's equations in either the vacuum case or the Einstein–Maxwell case.[2–6] In contrast, excepting a metric due to Vaidya,[7] no perfect-fluid solutions in generalized Kerr–Schild form were known up to the appearance of Ref. 1. In our opinion, there were two reasons for that lack. First, it was usual to start from metrics g which were solutions of the vacuum Einstein equations. But, if $T_{\alpha\beta}$ and $\widetilde{T}_{\alpha\beta}$ are the energy-momentum tensors for the metrics g and $\tilde{g}$, respectively, it was shown in Ref. 1 that the following interesting relation holds[8]:

$$l^\alpha \widetilde{T}_{\alpha\beta} = l^\alpha T_{\alpha\beta} + f l_\beta, \tag{2}$$

where f is a scalar field. Therefore, if $T_{\alpha\beta}$ vanishes then l_α is a null eigenvector of $\widetilde{T}_{\alpha\beta}$, so that $\widetilde{T}_{\alpha\beta}$ cannot be the energy-momentum tensor of a perfect fluid. The same happens when l_α is an eigenvector of $T_{\alpha\beta}$. Thus in order to obtain perfect-fluid metrics $\tilde{g}$ we must consider only the case in which l_α is not a null eigenvector of $T_{\alpha\beta}$. Of course, the most interesting case arises when $T_{\alpha\beta}$ itself is an energy-momentum tensor for a perfect fluid.

The second reason emerges from the fact that it is necessary to allow great freedom in choosing the vector field l^α. For example, in the classical Kerr–Schild metrics (Ref. 2), the great variety of shear-free null geodesic vector fields in flat space-time was used. For any metric g, there will be, in general, a great number of null geodesic vector fields.[9] But, in order to solve the Einstein equations, it is also very useful to know an explicit expression of the general solution for vector fields of this kind. In the classical Kerr–Schild metrics it was very useful that the Kerr theorem[10] provides the general solution for the shear-free geodesic null vector fields in flat space-time explicitly. This is not the case for an arbitrary metric g. However, it is known that the geodesic (shear-free) null vector fields in a conformally flat space-time are the geodesic (shear-free) null vector fields in flat space-time, and conversely. Thus if we start from a conformally flat space-time then we can use the Kerr theorem. Moreover, the conformally flat perfect-fluid metrics have an additional advantage: all metrics of this kind are known. They are either generalized interior Schwarzschild solutions or generalized Friedmann solutions.[11]

Therefore, we shall start from conformally flat perfect-fluid metrics g. We devote Sec. II to writing down the Einstein equations for this case. The case when the vector field l^α is shearing was studied in Ref. 1. On the other hand, the case when l^α is shear-free was solved only in a very particular subcase. In this paper, we try to solve the shear-free case in general. There are two outstanding subcases which are studied in Sec. III.

All the solutions we obtain are Petrov types D and II. We also point out that most of the type-D metrics are new since the velocity of the fluid does not lie in the two-space spanned by the two multiple null eigenvectors of the Weyl tensor. Apart from the results obtained in Ref. 1, only two metrics (Wahlquist,[12] Kramer[13]) with this property were known previously.

In Sec. IV we give some explicit examples. In Ref. 1 two explicit examples of how the method works and their respective metrics were given. Although the method always works in the same way, in this paper we present some explicit solutions again. In particular, a class of spherically symmetric static perfect-fluid metrics is obtained. The properties of the perfect-fluid sources themselves are discussed in Sec. V. Finally, Sec. VI is devoted to the study of the new class of spherically symmetric static perfect-fluid space-times.

II. THE EINSTEIN EQUATIONS

Hereafter, we choose the metric g and the null vector field l of (1) with the following properties. First, g is a solution of Einstein's equations for a perfect-fluid energy-momentum tensor,

$$R_{\alpha\beta} = \chi(T_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}T), \tag{3}$$

$$T_{\alpha\beta} = (q+p)u_\alpha u_\beta + pg_{\alpha\beta}, \quad g^{\alpha\beta}u_\alpha u_\beta = -1. \tag{4}$$

Second, g is conformally flat, that is to say

$$C_{\alpha\beta\lambda\mu} = 0, \tag{5}$$

[a)] Present address: Theoretical Astronomy Unit, School of Mathematical Sciences, Queen Mary College, Mile End Road, London E1 4NS, England.

 0022-2488/88/040937-08$02.50

where $C_{\alpha\beta\lambda\mu}$ is the Weyl tensor for the metric g. Or, equivalently, g can be transformed into

$$g_{\alpha\beta}\, dx^\alpha dx^\beta = 2\phi^2(-du\, dv + dz\, d\bar{z}), \tag{6}$$

where ϕ^2 is a positive function of the coordinates (the conformal factor) and where suitable coordinates $\{u,v,z,\bar{z}\}$ have been chosen for the flat metric. Finally, we choose the null geodesic vector field l_α to be shear-free. The general solution for vector fields of this kind in the metric (6) is known and is given by[10,11] (Kerr theorem)

$$l_\alpha\, dx^\alpha = du + \bar{Y} dz + Y d\bar{z} + Y\bar{Y} dv, \tag{7}$$

where Y is a complex function of the coordinates defined implicitly by

$$F(Y, \bar{z}Y + u, vY + z) = 0 \tag{8}$$

and where F is an arbitrary analytic function of three complex variables.

Now, we choose a null tetrad $\{l,k,m,\bar{m}\}$ associated with l as follows:

$$m = m_0 + \bar{Z}l, \quad k = k_0 + Zm_0 = \bar{Z}\bar{m}_0 + Z\bar{Z}l, \tag{9}$$

where

$$m_0 = \phi(dz + Y dv), \quad k_0 = \phi^2\, dv \tag{10}$$

and

$$\bar{Z} = -m_0 \cdot u / l \cdot u. \tag{11}$$

Therefore, in this null tetrad we have

$$m^\alpha u_\alpha = 0 \tag{12}$$

or, equivalently,[14]

$$\phi_{01} = \phi_{02} = \phi_{12} = 0, \quad \phi_{00} = (\chi/2)(q+p)(l^\alpha u_\alpha)^2,$$
$$\phi_{11} = (\chi/8)(q+p), \quad \phi_{22} = (\chi/8)(q+p)(l^\alpha u_\alpha)^{-2},$$
$$\Lambda = (\chi/24)(q-3p). \tag{13}$$

The spin coefficients[14] for the null tetrad (9) are given in Refs. 1 and 15 explicitly. Here, we only remark that

$$\varkappa = \epsilon = \sigma = \lambda = \beta = 0. \tag{14}$$

If we are looking for solutions $\tilde{g}$ for a perfect-fluid energy-momentum tensor

$$\tilde{T}_{\alpha\beta} = (\tilde{q} + \tilde{p})\tilde{u}_\alpha \tilde{u}_\beta + \tilde{p}\tilde{g}_{\alpha\beta}, \quad \tilde{g}^{\alpha\beta}\tilde{u}_\alpha\tilde{u}_\beta = -1, \tag{15}$$

and if we define

$$U \equiv \delta H - 2H\bar{\alpha}, \quad V \equiv DH - 2H\rho, \tag{16}$$

the Einstein equations become finally[1] (χ is the gravitational constant)

$$\chi\tilde{p} = \chi p - DV - V(3\rho + \bar{\rho}) - 3H\rho^2 - 4H\rho\bar{\rho}$$
$$+ H(\bar{\rho}^2 + 4\phi_{00}), \tag{17}$$

$$\chi\tilde{q} = \chi q + V(\bar{\rho} - \rho) - DV + 3H(\rho^2 + \bar{\rho}^2), \tag{18}$$

$$(l^\alpha \tilde{u}_\alpha)^2 = 2\phi_{00}\{\chi(q+p) + 4H\phi_{00}$$
$$-2[DV + 2V\rho + 2H\bar{\rho}(\rho - \bar{\rho})]\}^{-1}, \tag{19a}$$

$$m^\alpha \tilde{u}_\alpha = 0, \tag{19b}$$

$$\delta V + (\rho + \bar{\rho})U + (\tau - \bar{\alpha})V$$
$$+ H[\bar{\rho}\bar{\alpha} + \delta\bar{\rho} + 2\tau(\bar{\rho} - \rho)] = 0, \tag{20}$$

$$\rho\Delta H = -H\Delta(\rho + \bar{\rho}) - \delta\bar{U} + \bar{\alpha}\bar{U} - \bar{\tau}U - \tau\bar{U} - \mu V$$
$$- 4H\phi_{11} + H(\gamma + \bar{\gamma})(\rho - \bar{\rho}) - W, \tag{21}$$

where we have put

$$W \equiv (1/4\phi_{00})[DV + 2V\rho + 2H\bar{\rho}(\rho - \bar{\rho})]$$
$$\times [DV + 2V\rho + 2H\bar{\rho}(\rho - \bar{\rho}) - 4H\phi_{00} - 8\phi_{11}]. \tag{22}$$

The expressions (17)–(19) are not equations, but they define $\tilde{q}$, $\tilde{p}$, and $\tilde{u}^\alpha$ as functions of q, p, u^α, and the unknown H. Thus we need only solve Eqs. (20) and (21), which are differential equations for H. However, in addition to Eqs. (20) and (21) we have the compatibility conditions for U and V, that is,[1,15]

$$DU + (2\bar{\rho} - \rho)U + (\tau + \bar{\pi})V$$
$$+ H[\bar{\rho}\bar{\alpha} + \delta\bar{\rho} + 2\pi(\rho - \bar{\rho})] = 0, \tag{23}$$

$$\bar{\delta}U + \bar{\rho}\Delta H + \bar{\mu}\bar{V} - U\alpha + 2H\rho(\gamma + \bar{\gamma}) = \text{c.c.} \tag{24}$$

Also, from the definition (16) of V it follows that

$$V - \bar{V} = 2H(\bar{\rho} - \rho). \tag{25}$$

The Weyl tensor for the metric $\tilde{g}$ is given by[16]

$$\tilde{\psi}_0 = \tilde{\psi}_1 = 0, \quad -6\tilde{\psi}_2 = [D - 2(\rho - \bar{\rho})]V,$$
$$\tilde{\psi}_3 = (2\rho - \bar{\rho})\bar{U} + H[\rho\alpha + \bar{\delta}\rho + \bar{\tau}(\rho - \bar{\rho})], \tag{26}$$
$$\tilde{\psi}_4 = -[\bar{\delta} - (3\alpha - 2\bar{\tau})]\bar{U}.$$

Therefore, all the metrics $\tilde{g}$ are algebraically special and the vector field l^α is a multiple null eigenvector of the Weyl tensor.

When $V = 0$, from (23), (25), and the Newman–Penrose equations, it follows that $U = 0$ and $\rho = \bar{\rho}$. Then, from (26) it is evident that $\tilde{g}$ is conformally flat. Thus we shall only consider the case $V \neq 0$.

It has become clear to the authors that cases which have a function W nonlinear in H do not have solutions in general. This is because W is always of greater order in H than V and U. And then, Eqs. (16) and (21) are not compatible in general.[1,15] Perhaps, this fact has something to do with a theorem due to Xanthopoulos (see Refs. 5 and 6). Moreover, it is convenient to assume that

$$\rho = \bar{\rho}. \tag{27}$$

This assumption simplifies the calculations substantially. Therefore, we shall treat the following two cases:

$$\text{case A,} \quad DV + 2V\rho = 0, \tag{28}$$

$$\text{case B,} \quad DV + 2V\rho - 4H\phi_{00} = 0. \tag{29}$$

Of course, there are more cases in which W becomes linear in H, for example, when $DV + 2V\rho - 8\phi_{11} = 0$ and $DV + 2V\rho - 4H\phi_{00} - 8\phi_{11} = 0$. These two cases would provide different resulting metrics, which is a proof of the variety of possibilities in our Kerr–Schild transformation. However, all the cases are formally equivalent to either the case A or the case B, and then the calculations are just a pure repetition in other cases. We shall study both cases in the following section.

Because of (13) and (27) we can use the Bianchi identities for the metric g as given in the Appendix of Ref. 1. These identities and the Newman–Penrose equations for the metric

g (when they are conveniently restricted to each case) will be used repeatedly (but not explicitly) in Sec. III. Anyway, we shall omit the details.

III. THE SOLUTIONS

First, we assume the condition (28). From this assumption and (27), Eqs. (20), (21), (23), (24), and (25) become, respectively,

$$\delta V + 2\rho U + (\tau - \bar{\alpha}) V = 0, \tag{30}$$

$$\rho \Delta H = -2H\Delta\rho - \delta\bar{U} + \bar{\alpha}\bar{U} - \bar{\tau}U - \mu V - 4H\phi_{11}, \tag{31}$$

$$DU + \rho U + (\tau + \bar{\pi}) V = 0, \tag{32}$$

$$\bar{\delta} U - U\alpha = \delta\bar{U} - \bar{U}\bar{\alpha}, \tag{33}$$

$$V = \bar{V}. \tag{34}$$

The compatibility of (30) with (28), making use of (32), leads us to

$$(\rho^2 - \phi_{00}) U = \tau\rho V. \tag{35}$$

We can consider two cases.

(A1) $\rho^2 - \phi_{00} \neq 0$. Then Eq. (35) tells us that

$$U = \tau\rho V/(\rho^2 - \phi_{00}). \tag{35$'$}$$

(A2) $\rho^2 - \phi_{00} = 0$. Then Eq. (35) implies that we must have

$$\tau = 0. \tag{35$''$}$$

In case (A1), from (28) and (32) and by applying the operator D to (35$'$) we obtain

$$\pi = 0, \tag{36}$$

$$\tau[\rho D\phi_{00} + 2\phi_{00}(\rho^2 + \phi_{00})] = 0. \tag{37}$$

Therefore, we can consider two subcases again.

(A1a) $\rho D\phi_{00} = -2\phi_{00}(\rho^2 + \phi_{00})$. From the Newman–Penrose equations and the Bianchi identities it is easily shown that this condition is equivalent to

$$\phi_{00} = C\rho^2, \tag{38}$$

where C is an arbitrary positive real constant. In order to distinguish this case from the case (A2) we must assume that

$$C \neq 1. \tag{39}$$

(A1b) $\tau = 0$. In this case (37) is automatically satisfied. Furthermore, this case is different from the case (A2) since now we have $\rho^2 - \phi_{00} \neq 0$.

Next, we are going to solve the three subcases (A1a), (A1b), and (A2) separately.

The subcase (A1a): From the above considerations, Eqs. (30) and (28) are already compatible and Eq. (32) is satisfied. Also, Eq. (31) now becomes

$$\Delta \mathrm{H} = 2H(\mu + \gamma + \bar{\gamma}) - \mu V/\rho + 2V(\mu\rho + \phi_{11} + \Lambda)/\rho^2(1-C) + 2V\tau\bar{\tau}/\rho^2(1-C)^2. \tag{40}$$

The compatibilities of this equation with V and U give us, respectively,

$$\Delta V = \mu V(2-C) - \frac{2V\Lambda}{\rho} + (\gamma + \bar{\gamma}) V - \frac{4\phi_{11}V}{\rho} - \frac{2\tau\bar{\tau}V(3-C)}{\rho(1-C)^2} - \frac{4V(\mu\rho + \phi_{11} + \Lambda)}{\rho(1-C)}, \tag{41}$$

$$\Delta\tau = \mu\tau(2+C) - (\gamma - \bar{\gamma})\tau + \frac{2\tau\Lambda}{\rho} - \frac{2\tau(\mu\rho + \phi_{11} + \Lambda)}{\rho(1-C)} - \frac{2\tau\phi_{11}(1-2C)}{\rho C} - \frac{2\bar{\tau}\tau^2}{\rho(1-C)}. \tag{42}$$

Equation (41) is compatible with (28) and (30). Finally, a new integrability condition arises from (42) and the Newman–Penrose equations,

$$(1+C)\mu\rho + 2\Lambda + 2(2C-1)\phi_{11}/C = 0. \tag{43}$$

It is easily shown that condition (43) is possible.

The subcase (A1b): Now, we assume

$$\tau = 0, \quad \rho^2 \neq \phi_{00} \tag{44}$$

so that (35$'$) tells us that

$$U = 0. \tag{45}$$

Thus Eqs. (28) and (30) are compatible and Eqs. (32) and (33) are automatically satisfied. Moreover, Eq. (31) becomes

$$\rho\Delta H = 2H\rho(\gamma + \bar{\gamma} - \mu) - 4H(\phi_{11} + \Lambda) - \mu V, \tag{46}$$

which is compatible with (45). The compatibility of (46) with V leads us to

$$\Delta V = V(2\mu + \gamma + \bar{\gamma}) - 2V(2\phi_{11} + \Lambda)/\rho - \mu\phi_{00}V/\rho^2 - 4H\phi_{00}(\mu\rho + \phi_{11} + \Lambda)/\rho^2. \tag{47}$$

Also, from (44) and (45) Eq. (30) becomes

$$\delta V = \bar{\alpha} V. \tag{48}$$

Equations (47) and (48) are compatible and the integrability of (47) and (28) gives us the following condition:

$$V\left\{2\phi_{11} + \frac{\phi_{00}(7\mu\rho + 8\phi_{11} + 4\Lambda)}{\rho^2} + \frac{2\mu\phi_{00}{}^2}{\rho^3} + \frac{1}{\rho}\left[2D\phi_{11} + \frac{\mu}{\rho}D\phi_{00}\right]\right\} + 4\frac{H}{\rho^2}(\mu\rho + \phi_{11} + \Lambda)\left(D\phi_{00} + 4\rho\phi_{00} + \frac{2\phi_{00}{}^2}{\rho}\right) = 0. \tag{49}$$

It is very difficult to know if the expression (49) is possible in general. In fact, Eq. (49) should be interpreted as an equation from which V is obtained as a function of H. Then we should put this V in Eqs. (28), (47), and (48) and we should obtain an (or more!) expression for H which is not, in general, a solution of Eqs. (16) and (46). This procedure is useless in general. However, we can assume

$$\mu\rho + \phi_{11} + \Lambda = 0 \tag{50}$$

so that Eq. (49) becomes

$$2\phi_{11} + \frac{\phi_{00}(\phi_{11} - 3\Lambda)}{\rho^2} + \frac{2\mu\phi_{00}{}^2}{\rho^3} + \frac{1}{\rho}\left[2D\phi_{11} + \frac{\mu}{\rho}D\phi_{00}\right] = 0. \tag{51}$$

This is a condition on only g and therefore we only have to check it. In fact, it may be shown that (51) is possible.

The subcase (A2): This case is defined by the assumptions

$$\tau=0, \quad \phi_{00}=\rho^2 \tag{52}$$

so that Eqs. (28) and (30) are compatible. Furthermore, Eq. (31) may be written

$$\rho\Delta H=2H[\rho(\gamma+\bar{\gamma})-\mu\rho-2(\phi_{11}+\Lambda)] -\mu V-\bar{\delta}U+\alpha U. \tag{53}$$

The compatibility of this equation with V gives us

$$\Delta V=2(\bar{\delta}U-\alpha U)-4H(\mu\rho+\phi_{11}+\Lambda)+(V/\rho) \times[\rho(\mu+\gamma+\bar{\gamma})-2(2\phi_{11}+\Lambda)]. \tag{54}$$

In order to make this expression compatible with (28) we must have

$$\mu\rho+\phi_{11}+\Lambda=0 \tag{55}$$

and then, Eqs. (53) and (54) become, respectively,

$$\rho\Delta H=2H\rho(\mu+\gamma+\bar{\gamma})-\mu V-\bar{\delta}U+\alpha U, \tag{56}$$

$$\Delta V=2(\bar{\delta}U-\alpha U)+(V/\rho) \times[\rho(\mu+\gamma+\bar{\gamma})-2(2\phi_{11}+\Lambda)]. \tag{57}$$

The compatibility of (56) with U is

$$\rho\Delta U+\bar{\delta}\delta U-3\bar{\alpha}\bar{\delta}U + U[3\alpha\bar{\alpha}-\rho(5\mu+3\bar{\gamma}+\gamma)]=0. \tag{58}$$

Keeping this equation in mind, Eqs. (57) and (30) are compatible. Now, the Weyl tensor is given by

$$3\tilde{\psi}_2=V\rho, \quad \tilde{\psi}_3=\bar{U}\rho, \quad \tilde{\psi}_4=-(\bar{\delta}-3\alpha)\bar{U}. \tag{59}$$

Since V does not vanish, we only can obtain solutions of Petrov types D and II. For Petrov type-D solutions we must have

$$\delta U=U(3\bar{\alpha}-2\rho U/V). \tag{60}$$

Otherwise, the solutions are Petrov type II. By using the condition (33), and after a little computation, it is easily shown that Eqs. (60), (58), and (32) are compatible.

Now, we are going to solve case B. Therefore, we assume conditions (29) and (27) so that Eqs. (20), (23), (24), and (25) become, respectively, Eqs. (30), (32), (33), and (34). Furthermore, Eq. (21) now may be written

$$\rho\Delta H=-2H\Delta\rho-\delta\bar{U}+\bar{\alpha}\bar{U} -\bar{\tau}U-\tau\bar{U}-\mu V+4H\phi_{11}. \tag{61}$$

The compatibility of (30) with (29), making use of (32), leads us to (36) and (37). Also, we must have

$$(\rho^2+\phi_{00})U=\tau(V\rho-2H\phi_{00}). \tag{62}$$

Consequently, as in case A, we could consider two subcases again but it turns out that the only interesting case arises when

$$\tau=0 \tag{63}$$

and then, from (62) we have

$$U=0. \tag{64}$$

Thus, Eqs. (29) and (30) are compatible and also Eq. (32) is satisfied. On the other hand, Eq. (61) becomes

$$\rho\Delta H=2H\rho(\gamma+\bar{\gamma}-\mu)+4H(\phi_{11}-\Lambda)-\mu V, \tag{65}$$

which is compatible with (64). The integrability of (65) and V leads us to

$$\Delta V=V(2\mu+\gamma+\bar{\gamma})-2\Lambda(V/\rho)-\mu V\phi_{00}/\rho^2 +4\phi_{11}V/\rho+4(H/\rho)[2D\phi_{11}+2\rho\phi_{11} +(\phi_{00}/\rho)(\phi_{11}-\Lambda-2\mu\rho)]. \tag{66}$$

This equation is compatible with (30). Moreover, it may be shown that under the conditions

$$\mu\phi_{00}=2\rho\phi_{11}, \tag{67}$$

$$D\phi_{11}+\phi_{00}(\phi_{11}-\Lambda)/2\rho=0, \tag{68}$$

Eq. (66) is compatible with (29). In fact, condition (67) is not very much restrictive because it is satisfied for the metrics g which have a constant energy density. In other words, all the generalized interior Schwarzschild metrics satisfy the above-mentioned condition (67).[11]

The results obtained in this section can be summarized as follows.

Let us choose the conformally flat perfect-fluid metric g and the shear-free geodesic null vector field l^α such that they verify the possible conditions given in the first row of Table I. Then, let us define U and V by (16) and let us solve the integrable system of equations for U and V which appear in the second row of the table. Once this has been done, the solutions H of the compatible system of equations given by (16) and the third row of the table provide us generalized Kerr–Schild metrics $\tilde{g}$. These metrics are solutions of the Einstein equations for a perfect-fluid energy-momentum tensor (15), where the energy density $\tilde{q}$, the pressure $\tilde{p}$, and the velocity $\tilde{u}^\alpha$ are given, for each case, in the fourth row of the table. The Weyl tensor of the resulting metrics as well as their Petrov types are also shown in Table I.

As we can see in the table, the Petrov type-D metrics of cases A1a and A2 satisfy $\tilde{\psi}_3\neq0$ and $\tilde{\psi}_4\neq0$. Therefore, for these metrics, $\tilde{k}^\alpha$ is not a multiple null eigenvector of the Weyl tensor, and then the form of $\tilde{u}^\alpha$ tells us that the velocity of the fluid does not lie in the preferred two-space defined by the multiple null eigenvectors of the Weyl tensor. Thus these solutions are new. On the other hand, the metrics of cases A1b and B have u^α lying in that preferred two-space so that they may be already known.[11]

The particular case $V=CH\rho$ (where C is a constant $\neq-2$) belongs to the more general case A1a and it had been solved previously by one of us.[17] Likewise, the case $V=-2H\rho$ solved in Ref. 1 (when $\sigma=0$) belongs to the general case A2.

IV. EXPLICIT EXAMPLES

In this section we give some examples of how the equations may be solved for each particular case.

(1) The most simple metric g we can choose is the "flat" Robertson–Walker metric,

$$ds^2=2R^2(-du\,dv+dz\,d\bar{z}), \quad R=R(t), \quad q=q(t),$$

$$p=p(t), \quad t\equiv\frac{u+v}{2^{1/2}}, \quad \dot{q}=-\frac{3(q+p)\dot{R}}{R},$$

$$\dot{R}^2=\frac{\chi}{3}qR^4, \quad u_\alpha\,dx^\alpha=-\frac{R(du+dv)}{2^{1/2}},$$

$$l_\alpha\,dx^\alpha=du+\bar{Y}dz+Y\,d\bar{z}+Y\bar{Y}dv, \quad \dot{X}\equiv\frac{dX}{dt}.$$

TABLE I. Integrability conditions, compatible systems of equations, and properties for Kerr–Schild metrics.

Case	A1a	A1b	A2	B
Condition on g and l	(27),(36),(38),(42)	(27),(44),(50),(51)	(27),(52),(55)	(27),(63),(67),(68)
Equations for U and V	(35′),(28),(30),(41)	(45),(28),(47),(48)	(32),(33),(58),(28),(30),(57)	(64),(29),(66),(30)
Equation for ΔH	(40)	(46)	(56)	(65)
$\tilde{q}$, $\tilde{p}$, and $\tilde{u}^\alpha$		(17)–(19) with (27),(28)		(17)–(19) with (27),(29)
Weyl tensor	$3\tilde{\psi}_2 = V\rho$ $\tilde{\psi}_3 = \bar{\tau}V/(1-C)$ $\tilde{\psi}_4 = 2\bar{\tau}^2 V/\rho(1-C)^2$	$3\tilde{\psi}_2 = V\rho$ $\tilde{\psi}_3 = 0$ $\tilde{\psi}_4 = 0$	(59)	$3\tilde{\psi}_2 = V\rho - 2H\phi_{00}$ $\tilde{\psi}_3 = 0$ $\tilde{\psi}_4 = 0$
Petrov	D	D	II or D	D

We try to solve the equations for the case A2. Thus we must restrict the metric g and the vector field l^α such that conditions (27), (52), and (55) are satisfied. It is easily shown that these conditions are verified if and only if[18]

$$Y=0, \quad q+3p=0.$$

Then, the Robertson–Walker metric must be restricted such that

$$q=(A/R)^2, \quad R=Be^{\pm Ct},$$

$$C\equiv(\chi/3)^{1/2}A, \quad A, B=\text{const}.$$

Equations (32), (33), and (58) for U and Eqs. (28), (30), and (57) for V leads us to

$$U=f(z,\bar{z},u)/e^{\pm Ct}, \quad V=G(z,\bar{z},u)/e^{\pm 2Ct},$$

where the functions f and G verify the following equations:

$$\frac{\partial f}{\partial z}=\text{c.c.}, \quad \frac{\partial G}{\partial u}B=-2\frac{\partial f}{\partial z},$$
$$\frac{\partial G}{\partial \bar{z}}B=\pm 2^{1/2}Cf, \quad -2^{1/2}\frac{\partial^2 f}{\partial z\,\partial\bar{z}}=\pm C\frac{\partial f}{\partial u}. \tag{69}$$

A particular solution of these equations is given by

$$BG=\pm 2^{1/2}C[M(\bar{z})+\bar{M}(z)], \quad f=\frac{\partial M}{\partial \bar{z}},$$

where M is an arbitrary complex function of the variable $\bar{z}$. Now, we know that the system of equations (56) and (16) for H is compatible. The integration of this system is standard and we obtain for H

$$H=B[M(\bar{z})+\bar{M}(z)]+Ee^{\pm 2Ct},$$

where E is an arbitrary constant. These metrics belong to the class of generalized Robinson–Trautman solutions.[11,19] Unless we have $M=\text{const}$, the resulting metrics $\tilde{g}$ are Petrov type II.

Another particular solution of Eqs. (69) is given by

$$GB=-2a\exp\{\mp C(u+z+\bar{z})/2^{1/2}\}=-2f,$$
$$a=\text{const},$$

and then, the solution of Eqs. (56) and (16) for H is

$$CH=\mp 2^{1/2}aB\exp\{\mp C(u+z+\bar{z})/2^{1/2}\}+Ee^{\pm 2Ct}.$$

In this case, Eq. (60) is also satisfied and therefore the resulting metrics $\tilde{g}$ are Petrov type D. Since $\tilde{\psi}_3$ and $\tilde{\psi}_4$ do not vanish they are new.

(2) In this example we take the interior Schwarzschild metric in canonical coordinates, that is to say[11]

$$ds^2=-(Ar)^2\,dt^2+N^{-2}\,dr^2+r^2(d\theta^2+\sin^2\theta\,d\phi^2), \tag{70}$$

$$Ar=a-bN, \quad N^2=1-(r/R)^2, \quad a, b, R=\text{const}, \tag{71}$$

$$\chi q=3/R^2, \quad \chi p=(3bN-a)/R^2Ar, \tag{72}$$

$$u_\alpha\,dx^\alpha=-Ar\,dt, \tag{73}$$

and we try solve the equations for the case B. It may be shown that the only shear-free geodesic null vector field which satisfies $\tau=0$, (27) and (68) is given by

$$l_\alpha\,dx^\alpha=M(-dt+dr/ArN), \tag{74}$$

where

$$M\equiv 2^{1/2}A(1+\cos\theta)(A+B\sin\omega t+\omega\cos\omega t)$$
$$\times[\sin\theta(A\sin\omega t+B)]^{-2}, \tag{75}$$

$$\mathrm{B}r\equiv aN-b, \quad \omega^2=A^2-B^2=(a^2-b^2)/R^2. \tag{76}$$

A null tetrad associated with l_α and such that $m^\alpha u_\alpha=0$ is given by[20]

$$8^{1/2}k_\alpha\,dx^\alpha=r(1-\cos\theta)(\omega\cos\omega t-A-B\sin\omega t)$$
$$\times(Ar\,dt+dr/N), \tag{77}$$

$$2^{1/2}m_\alpha\,dx^\alpha=re^{i\phi}(-d\theta+i\sin\theta\,d\phi), \tag{78}$$

so that the only non-null spin coefficients are

$$\rho=MN/Ar^2, \quad \mu=AN/2M,$$
$$\alpha=-(1+\cos\theta)e^{-i\phi}/2^{1/2}r\sin\theta,$$
$$\gamma=(\tfrac{1}{8})^{1/2}(1-\cos\theta)$$
$$\times[N\omega\cos\omega t+(r/R^2)(a\sin\omega t-b)]. \tag{79}$$

Solving the system of equations defined by (29), (30), (66), (16), and (65) for V and H we obtain

$$V=\frac{A}{M}\left[\frac{N}{2}rf'-\frac{B}{A}f\right], \quad 2H=\left(\frac{Ar}{M}\right)^2 f, \tag{80}$$

where $f(r)$ is a solution of the following differential equation:

$$Ar^3N^2f''+(r^3/R^2)(4bN-a)f'-2Arf=0. \tag{81}$$

The resulting metric $\tilde{g}$ is a Petrov type-D static spherically symmetric perfect-fluid solution. In the following sections, we are going to discuss the properties of the solutions we have obtained.

V. PROPERTIES OF THE SOLUTIONS

It is evident that the properties of the generalized Kerr–Schild metrics $\tilde{g}$ depend, in general, on the properties of the initial metrics g themselves. However, some considerations may be made without loss of generality and then the specific properties of the explicit solutions can be deduced.

Thus, for example, in Ref. 1 it was shown that Petrov type-N metrics cannot be obtained by means of the Kerr–Schild transformation as defined by us. Also, the Petrov type of the resulting metrics has been always given in Sec. III. It is convenient to remark that this has been possible because we knew the Petrov type of the initial metrics (they are conformally flat).

With regard to the symmetries of the solutions, one of us[15] has shown the following result: "S is a Killing vector field of the Kerr–Schild metric $\tilde{g}$ if and only if

$$\pounds(\mathbf{S})\mathbf{l} = f\mathbf{l}, \quad \pounds(\mathbf{S})g_{\alpha\beta} = -2[\pounds(\mathbf{S})H + 2Hf]l_\alpha l_\beta, \tag{82}$$

where f is a function of the coordinates and we use standard notation for Lie derivatives." This result provides us a method to find all the Killing vector fields of the explicitly known Kerr–Schild metrics. For the first solution of the previous section, the former conditions (82) lead us to

$$\mathbf{S} = [(A_1+\bar{A}_1)u + A_3]\frac{\partial}{\partial u} - \left[(A_1+\bar{A}_1)u + A_3 \pm \frac{A_1+\bar{A}_1}{2^{1/2}C}\right]\frac{\partial}{\partial v} + (A_1\bar{z}+A_2)\frac{\partial}{\partial \bar{z}} + (\bar{A}_1 z + \bar{A}_2)\frac{\partial}{\partial z},$$

$$(A_1+\bar{A}_1)(E+B^2) = 0,$$

$$2(A_1+\bar{A}_1)(\mathrm{M}+\bar{M}) + (A_1\bar{z}+A_2)\frac{dM}{d\bar{z}} + (\bar{A}_1 z+\bar{A}_2)\frac{d\bar{M}}{dz} = 0, \tag{83}$$

where A_1, A_2, and A_3 are arbitrary constants (A_3 real). From these expressions it is evident that, in general, the only Killing vector is given by $A_1 = A_2 = 0$, that is to say

$$\mathbf{S} = \frac{\partial}{\partial u} - \frac{\partial}{\partial v}. \tag{84}$$

But also, there are some particular cases depending on the form of the function $\mathrm{M}(\bar{z})$. These are the following:

(a) If $\mathrm{M}(\bar{z}) = c\bar{z} + d$, $c,d = \text{const}$, $c \neq 0$,

then

$$\mathbf{S} = i\bar{c}\frac{\partial}{\partial \bar{z}} - ic\frac{\partial}{\partial z}$$

is another Killing vector.

(b) If $\mathrm{M}(\bar{z}) = d$, then

$$\frac{\partial}{\partial z} + \frac{\partial}{\partial \bar{z}}, \quad i\left(\frac{\partial}{\partial \bar{z}} - \frac{\partial}{\partial z}\right), \quad i\left[\bar{z}\frac{\partial}{\partial z} - z\frac{\partial}{\partial \bar{z}}\right]$$

are Killing vectors as well.

(c) If $\mathrm{M}(\bar{z}) = (a/b)\log(ib\bar{z} + A_2)$, $a,b = \text{real const}$, then

$$(ib\bar{z}+A_2)\frac{\partial}{\partial \bar{z}} + (\bar{A}_2 - ibz)\frac{\partial}{\partial z}$$

is another Killing vector.

(d) If $M(\bar{z}) = c(A_1\bar{z}+A_2)^{-2(1+\bar{A}_1/A_1)} - ia/(A_1 + \bar{A}_1)$ and $E = -B^2$ then (83) with $A_3 = 0$ is a Killing vector. This is the only new Killing vector which is not a Killing vector of the initial metric g.

Similarly, the symmetries of the second solution of the previous section may be obtained. The result is that there are the following two Killing vectors:

$$i\left(\frac{\partial}{\partial \bar{z}} - \frac{\partial}{\partial z}\right), \quad \frac{\partial}{\partial u} - \frac{\partial}{\partial v} + \frac{1}{2}\left(\frac{\partial}{\partial \bar{z}} + \frac{\partial}{\partial z}\right).$$

Both of them were Killing vectors for the initial Robertson–Walker metric g.

In relation to the properties of the density and pressure of the solutions, first of all we must obtain the explicit expressions for these quantities, which are given by

$$\chi\tilde{q} = I\exp(\mp 2Ct) + (M+\bar{M})(A^2/3B^3)\exp(\mp 4Ct), \tag{85}$$

$$\chi\tilde{p} = -(I/3)\exp(\mp 2Ct) + (M+\bar{M})(A^2/3B^3)\times\exp(\mp 4Ct), \tag{86}$$

for the first metric obtained in Sec. IV, and by

$$\chi\tilde{q} = I\exp(\mp 2Ct) + (2^{1/2}aA^2/3CB^3)\times\exp[\mp 4Ct \mp C(u+z+\bar{z})/2^{1/2}], \tag{87}$$

$$\chi\tilde{p} = -(I/3)\exp(\mp 2Ct) + (2^{1/2}aA^2/3CB^3)\times\exp[\mp 4Ct \mp C(u+z+\bar{z})/2^{1/2}], \tag{88}$$

for the second metric, where

$$I \equiv (A/B)^2(1+E/B^2).$$

From (85)–(88) it is clear that the solutions do not have singularities in general. Furthermore, both metrics satisfy

$$\chi(\tilde{q}-\tilde{p}) = (4I/3)\exp(\mp 2Ct). \tag{89}$$

Finally, we are going to study the properties of the velocity of the fluid of the Kerr–Schild metrics. By using the formulas of Ref. 21, making a change of null tetrad and after some standard and straightforward calculations we obtain for the shear, vorticity, and expansion of the fluid the following expressions:

$$\sigma_{\alpha\beta}:\ A = -\left(\frac{1}{18}\right)^{1/2}\Big\{(1+\epsilon)V + \frac{2\phi_{11}}{\phi_{00}L^2}DH - \frac{\epsilon}{L^2}[\Delta H - 2H(\gamma+\bar{\gamma})] - \frac{H}{L^2\phi_{00}}(1-\epsilon)\left[2D\phi_{11} - \frac{2\phi_{11}}{\phi_{00}}D\phi_{00}\right] - \frac{2H\epsilon}{L^2\phi_{00}}(2\phi_{11}\rho - \mu\phi_{00})\Big\}L^{-1}, \tag{90}$$

$$B = -(1/2L^2)[H\bar{\tau}(1-\epsilon) + (1+\epsilon)\bar{U}], \tag{91}$$

$$C = 0, \tag{92}$$

$$\omega_{\alpha\beta}:\ U = (1/2L^2)(1-\epsilon)(H\bar{\tau} + \bar{U}), \tag{93}$$

$$V=0, \tag{94}$$

$$\theta:\ \theta=\frac{1}{2L}\Big\{(1+\epsilon)(DH+4H\rho)+\frac{2\phi_{11}}{\phi_{00}L^2}DH$$
$$-\frac{\epsilon}{L^2}[\Delta H-2H(\gamma+\bar{\gamma})]$$
$$-(1-\epsilon)\frac{H}{L^2\phi_{00}}\Big[2D\phi_{11}-\frac{2\phi_{11}}{\phi_{00}}D\phi_{00}\Big]$$
$$-\frac{(2\phi_{11}\rho-\mu\phi_{00})(\epsilon H/L^2-3)}{\phi_{00}}\Big\}, \tag{95}$$

where we have used the notation of Ref. 21 and we have put

$$L^2=2\phi_{11}/\phi_{00}+\epsilon H.$$

Moreover, $\epsilon=1$ for case A and $\epsilon=-1$ for case B. These expressions are valid in general. From (93) we see that the solutions obtained in this paper do not have vorticity. This is a direct consequence of assumption (27). On the other hand, they have, in general, shear, expansion, and acceleration. The explicit expressions of these quantities for the explicit metrics of Sec. IV may be easily obtained from (90)–(95). However, the static and spherically symmetric solution is shear-free and expansion-free (of course!). In the next section, we are going to study this particular solution.

VI. A CLASS OF STATIC, SPHERICALLY SYMMETRIC PERFECT-FLUID METRICS

In Sec. IV, we obtained the metric

$$d\tilde{s}^2=-(Ar)^2dt^2+N^{-2}dr^2+r^2(d\theta^2+\sin^2\theta\,d\phi^2)$$
$$+(Ar)^2f(r)(-dt+dr/ArN)^2, \tag{96}$$

where $f(r)$ is a solution of the differential equation (81). By making the following change of the timelike coordinate:

$$dT=dt+[f/ArN(1-f)]dr, \tag{97}$$

the metric (96) becomes

$$d\tilde{s}^2=-(Ar)^2(1-f)dT^2+[1/N^2(1-f)]dr^2$$
$$+r^2(d\theta^2+\sin^2\theta\,d\phi^2), \tag{98}$$

In this form, the metric is manifestly static and spherically symmetric. By using the formulas of previous sections we can get the velocity of the fluid

$$\tilde{u}=Ar(1-f)^{1/2}dT \tag{99}$$

and the density and pressure

$$\chi\tilde{q}=\chi q+(N^2/r)f'+[1-3(r/R)^2]f/r^2, \tag{100}$$

$$\chi\tilde{p}=\chi p-(N^2/r)f'-[1-(3-2a/Ar)(r/R)^2]f/r^2, \tag{101}$$

where χq and χp are the density and pressure of the Schwarzschild interior solution and are given in (72). By the way, we remark that the metric (98) is a generalization of the Schwarzschild interior metric. The Schwarzschild metric is the particular case $f(r)=0$ [as is evident from (98)–(101) or directly from (80)].

If we want to study the properties of the solution (98) we have to solve the differential equation (81). This equation is linear and of second order and, in general, it has four regular singular points.[22,23] This type of equation is called Heun's equation.[22,23] However, there are two cases in which the equation has only three regular singular points (so that it can be reduced to the hypergeometric equation[23]). These cases are defined by $b=0$ and $b=a$. When $b=a$, it may be shown that there is not any regular solution. Then, we do not consider this case here.

Let us begin with the easier case $b=0$. In this case, (81) can be reduced to a hypergeometric equation and, in fact, the general solution of (81) may be expressed by means of elementary functions as follows:

$$f(r)=3C[1-(NR/r)\arcsin(r/R)]+cNR/r, \tag{102}$$

where C and c are arbitrary constants. In order to have a regular solution in $r=0$ we should impose

$$c=0. \tag{103}$$

Keeping this condition in mind, from (100) and (101) we get the density and pressure for this case,

$$\chi\tilde{q}=(3/R^2)[1-2C+(3CNR/r)\arcsin(r/R)], \tag{104}$$

$$\chi\tilde{p}=-R^{-2}[1+(3CNR/r)\arcsin(r/R)]. \tag{105}$$

It follows from (104) and (105) that this solution satisfies the equation of state

$$\chi(\tilde{q}+3\tilde{p})=-6C/R^2=\text{const.} \tag{106}$$

By physical considerations, we must demand $C<0$ so that $\tilde{q}+3\tilde{p}$ is positive and, also, this assures the correctness of the signature for (98) because $f(r)<0$. This special metric is just the static limit of the Wahlquist solution[12] and it was also given by Whittaker.[24]

Now, let us study the general case $b\neq 0$. By simplicity, it is convenient to distinguish several possibilities depending on the different values of b/a. Thus, for example, when $b/a\leqslant\frac{1}{3}$ the regular solution of (81) is

$$f(r)=C(r/R)^2F[(b-a)/2b,(5b-2a)/b;$$
$$5,2,\tfrac{5}{2},\tfrac{5}{2};\ (1-N)/2], \tag{107}$$

where F is the solution of Heun's equation (see Ref. 22). The regular solution of (81) when $\frac{1}{3}<b/a<\frac{1}{2}$ is

$$f(r)=C[2b/(a-b)]^{3/2}(r/R)^2F$$
$$\times[2b/(b-a),(10b-4a)/(b-a);5,2,\tfrac{5}{2},3;\ x],$$

$$x=[2b(1-N)/2(b-a)], \tag{108}$$

where C is an arbitrary constant again. Analogously, the solution when $\frac{1}{2}\leqslant b/a<1$ may be given by means of the Heun's function F.

We shall restrict ourselves to the case $b/a\leqslant\frac{1}{3}$ because all the possibilities are quite similar. From (107) it may be shown

$$f(0)=0,\quad \frac{f}{r^2}(r=0)=\frac{C}{R^2},\quad \frac{f'}{r}(r=0)=\frac{2C}{R^2}, \tag{109}$$

and then, the density and pressure are regular everywhere. Moreover, from (72), (100), (101), and (109), we have

$$\chi\tilde{q}(0)=3(1+C)/R^2,$$
$$\chi\tilde{p}(0)=R^{-2}[-3C+(3b-a)/(a-b)].$$

Therefore, we must choose C as follows:

$$-1<C<(3b-a)/3(a-b)$$

so that $\tilde{q}(0)$ and $\tilde{p}(0)$ are positive.

Bearing this condition in mind, and taking into account the following relation:

$$\chi\tilde{p}(r=R) = \chi p(r=R) = -R^{-2} < 0,$$

we conclude that the pressure is a decreasing function of r and that there exists a value $r = r_0 < R$ such that $\tilde{p}(r_0) = 0$.

Unfortunately, it is very difficult to find out the equation of state for these metrics. However, from (100) and (101) it is evident that the following relation holds in general:

$$\chi(\tilde{q} + \tilde{p}) = \chi(q+p)(1-f) = [2a/ArR^2](1-f).$$

This expression proves that there are no solutions in which the density and the pressure vanish at the same value of r without singularities in the metric [see (98)] and, therefore, the equation of state cannot be that of a polytropic fluid.

[1]J. Martín and J. M. M. Senovilla, J. Math. Phys. **27**, 265, 2209 (1986).
[2]G. C. Debney, R. P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
[3]A. Thompson, Tensor (N. S.) **17**, 92 (1966).
[4]J. Plebanski and A. Schild, Nuovo Cimento B **35**, 35 (1976).
[5]A. H. Taub, Ann. Phys. (NY) **134**, 326 (1981).
[6]B. C. Xanthopoulos, J. Math. Phys. **19**, 1607 (1978); Ann. Phys. (NY) **149**, 286 (1983).
[7]P. C. Vaidya, Pramana **8**, 512 (1977).
[8]Throughout the paper there are two kinds of objects related to the metrics $\tilde{g}$ and g. We denote these, respectively, by a tilde and by no label. Consequently, we raise and lower indices of the tensors with $\tilde{g}$ or g, respectively. It is easily shown that $\tilde{l} = l$ (see Ref. 1).
[9]However, when l is also shear-free there is not, in general, enough freedom in choosing l because of the generalized Goldberg–Sachs theorem (see Ref. 11).
[10]D. Cox and E. J. Flaherty, Jr., Commun. Math. Phys. **47**, 75 (1976).
[11]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Deutscher Verlag der Wissenschaften, Berlin, 1980).
[12]H. D. Wahlquist, Phys. Rev. **172**, 1291 (1968).
[13]D. Kramer, Class. Quantum Grav. **1**, L3 (1984).
[14]Standard notation in Newman–Penrose formalism is used throughout the paper. Our conventions coincide with those of Kramer *et al.* (Ref. 11) except for the spin coefficients where the sign is changed.
[15]J. M. M. Senovilla, Ph.D. thesis, Universidad de Salamanca, 1986.
[16]The expressions (26) are calculated in the tetrad $\tilde{l} = l$, $\tilde{m} = m$, and $\tilde{k}^\alpha = k^\alpha + Hl^\alpha$, (Ref. 1).
[17]J. M. M. Senovilla, in *Actes dels Encuentros Relativistas Españoles 85*, Maó 1985, edited by A. Molina (L'E. T. S. E. I. B., Menorca, 1986).
[18]The metric g does not have any physical meaning when we have obtained the new metric g. Then, we can allow the relationship $q + 3p = 0$ since we only must impose physical restrictions to the new $\tilde{q}$ and $\tilde{p}$.
[19]J. Wainwright, Int. J. Theor. Phys. **10**, 39 (1974).
[20]This null tetrad and the spin coefficients have been obtained by using the method developed in Sec. II (and Ref. 1) starting from the most general tetrad (7)–(11). Therefore, we have needed the conformal factor for the interior Schwarzschild metric, which is given by

$$\phi = r(A \sin \omega t + B).$$

This explicit value has been found by J. Martín and J. M. M. Senovilla (unpublished) and as far as we know, it was unknown in this form up to now.
[21]J. Wainwright, J. Math. Phys. **18**, 672 (1977). We use the formulas (A6)–(A9) of Appendix A of this paper.
[22]E. Kamke, *Differentialgleichungen Losungsmethoden und Losungen* (Chelsea, New York, 1971).
[23]Bateman Manuscript Project, A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953).
[24]J. M. Whittaker, Proc. R. Soc. London Ser. A **306**, 1 (1968).

On spherically symmetric shear-free perfect fluid configurations (neutral and charged). II. Equation of state and singularities

Roberto A. Sussman[a)]
School of Mathematical Sciences, Queen Mary College, Mile End Road, London E1 4NS, England

(Received 6 August 1987; accepted for publication 25 November 1987)

Geometrical and physical properties of the solutions derived and classified in Part I [J. Math. Phys. **28**, 1118 (1987)] are examined in detail. It is shown how the imposition of zero shear restricts the possible choices of equations of state. Two types of singular boundaries arising in these solutions are examined by verifying the local behavior of causal curves approaching these boundaries. For this purpose, a criterion due to C. J. S. Clarke (private communication) is given, allowing one to test the completeness of arbitrary accelerated timelike curves in terms of their acceleration and proper time. One of these boundaries is a spacelike singularity at which causal curves terminate as pressure diverges but matter-energy and charge densities remain finite. At the other boundary, which is timelike if the expansion Θ is finite, proper volume of local fluid elements vanishes as all state variables diverge but causal curves are complete. If Θ diverges at this boundary, a null singularity arises as the end product of the collapse of a two-sphere generated by a given class of timelike curves. The gravitational collapse of bounded spheres matched to a Schwarzschild or Reissner–Nordstrøm exterior is also examined in detail. It is shown that the spacelike singularity mentioned above could be naked under certain parameter choices. Solutions presenting the other boundary produce very peculiar black holes in which the "surface" of the sphere collapses into the above mentioned null singularity, while the "interior" fluid layers avoid this singularity and evolve towards their infinite future.

I. INTRODUCTION

As mentioned in the Sec. I of Part I (Ref. 1), SSSF (spherically symmetric shear-free perfect fluid) solutions characterized by the metric I(11) follow from a strategy of "mathematical simplicity" in which the field equations are considerably simplified by imposing a simple motion on the fluid (shear-free motion). [Equation I(11) denotes Eq. (11) of Part I. All reference to equations of Part I will be made in this form. The same conventions on indices, signature, and units as in Part I are being used.] Having derived and classified the large class of SSSF solutions denoted as ChKQ (charged Kustaanheimo–Qvist) solutions, the natural continuation of the strategy of mathematical simplicity is an examination of the properties of these solutions. However, details concerning specific particular solutions (see Appendices A and C) will be avoided as much as possible by focusing on properties common to all, or at least to a large subclass of ChKQ solutions. Although much of the discussion in the present paper (especially Secs. II, III, and VI) applies to ChKQ solutions in general, it is convenient to concentrate the investigation of these solutions on the subclass in which one of the time-dependent free parameters, the function $L = L(t)$, is set to a constant. This large subclass comprises the "McVittie-type" and "Wyman-type" solutions (see Sec. V of Part I). Other ChKQ solutions will be examined in the subsequent paper (Part III) together with other topics, such as the global and asymptotic properties of all ChKQ solutions. The contents of Part II are summarized below.

In Sec. II a review[2,3] is offered of the invariant characterization of SSSF solutions in terms of a vector field generating a one-dimensional group of conformal motions. The field equations are written in Sec. III in a compact form in terms of the Hubble scale factor of the comoving observers, $H = (g_{rr})^{1/2}$, or the proper radius of the orbits of SO(3), $R = (g_{\theta\theta})^{1/2}$, either one of which is taken as the basic gravitational field variable. It is shown that the field equations contain "curvature terms" related to the curvature of the three-surfaces orthogonal to the four-velocity ("surfaces Σ_t" labeled by constant coordinate time) and to the four-acceleration. In Sec. IV, the McVittie- and Wyman-type solutions are introduced, while three possible choices of time coordinates are presented in Sec. V. In Sec. VI, it is shown how an equation of state (in general, a nonbarotropic equation of state, for the isentropic case see the Wyman solution in Appendix A), whether "imposed" or "obtained," must comply with the above-mentioned curvature terms in the field equations. In general, these curvature terms severely restrict the equations of state compatible with the solutions. In Sec. VII, it is shown how a simple, though unphysical, type of "formal" equation of state can be constructed as a boundary condition. Most authors who have previously studied ChKQ solutions [authors of category (b), see Sec. I of Part I] have examined them as models of collapsing spheres, and so have used a particular case of these boundary conditions.

Regularity conditions are presented in Sec. VIII, identifying the coordinate representation of the boundaries within which scalar curvature invariants and gravitational field variables H and/or R are smooth and bounded ("regularity boundaries"). In order to understand the nature of these boundaries, it is necessary to verify if causal curves approaching them are complete. In order to do this, Sec. IX examines the completeness criteria for two types of causal

a) Present address: Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel.

 0022-2488/88/040945-26$02.50

congruences whose description is simple in the coordinates currently used: the word lines of observers comoving with the matter and "radial" null geodesics. Since the former are accelerated curves, a criterion, due to Clarke (private communication), of their generalized affine parameter (GAP) completeness in terms of their acceleration and proper time is offered and rigorously proved in Appendix B. By verifying the completeness of timelike and null curves approaching the coordinate values associated with the regularity boundaries where curvature scalars diverge, it is found in Sec. X that one of these boundaries [Eq. (49)] is a spacelike scalar curvature singularity in which pressure and pressure gradient diverge but the proper volume of local fluid elements, matter-energy, and charge densities remain finite. This "finite-density" singularity (FD singularity), similar to that arising in some Bianchi models studied by Collins and Ellis,[4] has been reported previously (though usually in a nonrigorous manner) by authors studying particular cases of ChKQ solutions.[5–12]

The other regularity boundary examined in Sec. X [Eq. (48)] is timelike and, since proper volume of local fluid elements vanishes as all curvature scalars diverge, apparently is a coordinate representation of a "big-bang" singularity present in Friedman–Robertson–Walker (FRW) solutions. However, causal curves approaching these coordinate values are complete, and so technically speaking, these values do not mark a singularity but a singular boundary which behaves as a sort of "asymptotically delayed" big bang (AD big bang). This situation corresponds to a finite expansion kinematic parameter Θ, so that infinite volume contraction (or expansion) takes place in infinite proper time. This boundary was mentioned by Mashhoon and Partovi[11] and by Collins[12] in their study of the Wyman solution; however, Mashhoon and Partovi failed to notice that it takes place in the infinite future/past of the observers comoving with the fluid. In both singular boundaries mentioned above, the strong and dominant (though not necessarily the weak) energy conditions are violated in general.

However, if the equation of state is such that Θ diverges for a class of comoving observers, the world lines of the latter are incomplete, collapsing into a null singularity which will be termed a "localized" singularity (L singularity). The world lines of all other classes of comoving observers, either remain complete evolving towards the AD big bang, or terminate in another spacelike singularity which will be called a "finite-volume" singularity (FV singularity). The latter singularity is characterized by diverging Θ, ρ, and p (though charge density and four-acceleration remain finite) with nonzero terminal proper volume of local fluid elements.

With all the information obtained in Sec. X, the case of collapsing spheres matched to a Schwarzschild or Reissner–Nordstrøm exterior is discussed in Sec. XI. Since their evolution terminates at $R>0$, collapsing solutions presenting the finite-density spacelike singularity do not form black holes. This has also been commented by several authors studying particular cases of collapsing ChKQ solutions (see especially Refs. 6 and 7). However, in solutions presenting the other regularity boundary, the matching surface (i.e., the time history of the two-sphere which is the "surface" of the sphere) collapses into the null singularity (L singularity) mentioned previously, but the interior layers continue their evolution towards the AD big bang in their infinite future. In some cases the evolution of the interior layers terminates in the spacelike FV singularity mentioned previously.

As far as I am aware, these peculiar singularities (the L and FV singularities), together with the type of black holes associated with them, have never been reported before. One of these two cases (the case presenting a AD big bang) is probably the only example provided so far in which the collapse into a black hole could be "survived" by observers inside the sphere. The question of censorship of the spacelike and null singularities mentioned above is discussed with the help of qualitative Penrose diagrams. Since the FD spacelike singularity does not involve "shell-crossing" effects, and there exist "bad" choices of parameters which make it naked, it is suggested that it could provide a new type of example of naked singularities in spherically symmetric collapse.[13] An example of a simple collapsing particular charged solution is presented in Appendix C.

II. KINEMATICS AND SYMMETRIES

The metric for SSSF configurations (neutral or charged) is given in comoving coordinates by I(11):

$$ds^2 = -\left[\frac{\dot H/H}{\Theta/3}\right]^2 dt^2 + H^2[dr^2 + f^2(d\theta^2 + \sin^2\theta\, d\phi^2)]\,, \qquad (1)$$

with $\dot H \equiv \partial H/\partial t$, $\Theta = \Theta(t)$ the expansion kinematic parameter, and $f = f(r)$ fixed in general by a choice of radial coordinate r. Such a choice was made in Part I by demanding the radial coordinate to be such that $(f')^2 = 1 - kf^2$, where a prime denotes differentiation with respect to r. This choice leads to

$$f(r) = \begin{cases} r, & k=0\,, \qquad (2a)\\ \sin r, & k=1, \qquad (2b)\\ \sinh r, & k=-1. \qquad (2c)\end{cases}$$

The expansion kinematic parameter Θ will be left for the time being as an unspecified function. As the contents of this section apply to all SSSF solutions, the metric coefficient H can be any function satisfying Eq. I(15) [or I(18)].

The coordinates (t,r,θ,ϕ) in (1) provide a natural representation reflecting the geometry of SSSF solutions. Since the spatial coordinates (r,θ,ϕ) are comoving with the fluid, the four-velocity has the simple form I(2):

$$u^\alpha = U(t,r)\delta_t{}^\alpha\,, \qquad (3a)$$

where

$$U(t,r) \equiv (-g^{tt})^{1/2} = (\Theta/3)/(\dot H/H)\,, \qquad (3b)$$

and $\delta^\alpha{}_\beta$ is the Kronecker delta tensor. Equations (3) define at every point in space-time a unit timelike vector field everywhere tangent to the set of three-surfaces $\Sigma_r(t,\theta,\phi)$ labeled by constant values of r. These three-surfaces are generated by the world lines of observers comoving with the fluid (i.e., integral curves of u^α), and are a set of coaxial world tubes describing the time evolution of the orbits of SO(3) (the concentric two-spheres generated by the Killing vectors of spherical symmetry). Thus each one of these two-spheres is completely characterized by a constant value of r which "represents" one class of equivalence of comoving observers in an orbit of SO(3). Since angular coordinates are ignorable, the evolution of the fluid can be completely described by two-dimensional coordinate patches parametrized by (t,r) in which each point represents one of the above-mentioned two-spheres with proper surface:

$$\mathscr{S}(t,r) = 4\pi R^2(t,r) = 4\pi[f(r)H(t,r)]^2 . \tag{4}$$

The metric coefficient $R = fH = (g_{\theta\theta})^{1/2}$ is then invariantly characterized as the proper radius, or "curvature" radius of these comoving two-spheres.

Since spherically symmetric fluids are irrotational, the three-surfaces $\Sigma_t(r,\theta,\phi)$ labeled by constant values of the time coordinate can be defined invariantly as spacelike three-surfaces everywhere orthogonal to the four-velocity field.[14] Therefore, the rest frames of the comoving observers lie along these surfaces, and the projection tensor has the simple form

$$h_{\alpha\beta} = g_{\alpha\beta} + u_\alpha u_\beta = g_{ij}\delta_\alpha{}^i\delta_\beta{}^j , \tag{5}$$

so that the projection of any tensor quantity along these rest frames is given simply by the spatial components of the tensor. While the surfaces Σ_r clearly have topology $S^2 X\mathbb{R}$, the topology of the surfaces Σ_t is not obvious. With the coordinate choice (2), the induced metric of the latter is conformal to that of three-surfaces of constant curvature. Whether this fact is a mere coordinate effect or an indication of the topology of these surfaces will be discussed in Sec. II of Part III.

The kinematical parameters associated with SSSF solutions are the expansion Θ and the four-acceleration a_α appearing in the usual decomposition of $u_{\alpha;\beta}$ (see Refs. 14 and 15):

$$u_{\alpha;\beta} = (\Theta/3)h_{\alpha\beta} - a_\alpha u_\beta , \tag{6a}$$

where

$$\Theta/3 \equiv u^\alpha{}_{;\alpha} , \tag{6b}$$

and

$$a_\alpha \equiv u_{\alpha;\beta}u^\beta . \tag{6c}$$

For the metric (1) and using (3) and (4), the expansion Θ can be expressed as

$$\frac{\Theta}{3} = \frac{d}{d\tau}[\ln H] = \frac{d}{d\tau}[\ln R] , \tag{7}$$

where

$$\frac{d}{d\tau} \equiv (-g^{tt})^{1/2}\frac{\partial}{\partial t} = U(t,r)\frac{\partial}{\partial t} \tag{8}$$

is the normalized proper time derivative for the comoving observers.

The absence of shear requires that, for all comoving observers, the relative distances of neighboring fluid particles change isotropically as measured in the rest frames of these observers.[14] This constraint requires Θ to be constant in these rest frames, and thus constant along the surfaces Σ_t. Also, the four-acceleration forcing the fluid to evolve without shear must have the form

$$a_\alpha = \delta_\alpha{}^r\mathscr{A}(t,r) , \tag{9a}$$

where $\mathscr{A}$ is the gradient:

$$\mathscr{A}(t,r) \equiv -[\ln U]' = [\ln \dot{H}/H]' . \tag{9b}$$

The specific form of Θ and a_α conveys a sort of "kinematic isotropy" that follows directly from the kinematical constraint of imposing zero shear on the fluid motion. This kinematic isotropy can be characterized invariantly by the action of the vector field[2,3]:

$$\lambda^\alpha \equiv \lambda(t,r)u^\alpha , \tag{10a}$$

where

$$[\ln\lambda]_{,\alpha} = a_\alpha \Rightarrow \lambda = U^{-1} . \tag{10b}$$

This vector field satisfies

$$\lambda_{(\alpha;\beta)} = (\Theta/3)\lambda h_{\alpha\beta} \Rightarrow \mathscr{L}_\lambda h_{\alpha\beta} = (2\Theta/3)h_{\alpha\beta}, \tag{11}$$

where $\mathscr{L}_\lambda$ denotes the Lie derivative operator along the vector λ^α. Hence, for each comoving observer, λ^α generates a one-parameter group of conformal motions along u^α that locally preserve relative orientations of neighboring fluid particles in the rest frames (the surfaces Σ_t) as the latter evolve in time. Thus, as mentioned by Tauber and Weinberg,[2] these three-surfaces change locally by a sort of "rigid magnification."

It is worthwhile examining the kinematic isotropy generated by (10a) in the two limiting particular cases of SSSF solutions: Static particular solutions for which $\Theta = 0$ but $a_\alpha \neq 0$, and the FRW (Friedman–Robertson–Walker) solutions in which $a_\alpha = 0$ but $\Theta \neq 0$. In the latter case, $\lambda = \lambda(t)$ can be normalized by rescaling the time coordinate, and so the vector λ^α can be identified with the four-velocity u^α. In this case, λ^α is a conformal Killing vector and it is "the same" vector everywhere along the surfaces Σ_t, in the sense that it is invariant under the action of the three-dimensional (3-D) transitive group of isometries Bianchi I, V, and IX (depending on the value of $k = 0, \pm 1$) whose orbits are the surfaces Σ_t.[16] This feature is related to the fact that for the FRW solutions, these surfaces are surfaces of constant proper time; therefore relative distances along the rest frames change isotropically and at the same proper time rate for all comoving observers [i.e., these surfaces are rigidly "magnified" by the same scale factor $H(t)$ everywhere]. For the static limit of SSSF solutions, the vector λ^α in (10a) becomes the timelike Killing vector field characterizing spherically symmetric static solutions.

For SSSF solutions that are neither static nor FRW, the imposition of zero shear requires the four-acceleration to have the specific form given by (9). This four-acceleration deviates all fluid particles from geodesic motion in such a way that the required kinematic isotropy occurs. Hence, because of the dependence of the four-velocity on r, the surfaces of constant proper time Σ_τ do not coincide with the surfaces Σ_t. From (1) and (3), the proper time $\tau(t,r)$ for comoving observers labeled by different constant values of r is

$$\tau(t,r) = \int \frac{dt}{u(t,r)} \quad (r \text{ constant}) , \tag{12}$$

so that $t = t(\tau,r)$ and thus $\Theta(t) = \Theta(\tau,r)$. Although relative distances change at the same proper time rate along the surfaces Σ_t (Θ must be constant along these surfaces), the scale factors are different for different r. These scale factors are given by the metric coefficient $H = (g_{rr})^{1/2}$, which from (4) and (7) can be identified as a position-dependent Hubble scale factor. The relation between the Hubble scale factor H at different surfaces Σ_r labeled by r_0 and r_1 is proportional to the ratio of the proper time along r_0 and r_1. This relation follows from (7) and (12), and is given by

$$\frac{d\tau_0}{d\tau_1} = \frac{d(\ln H_0)}{d(\ln H_1)}, \tag{13}$$

where the subindices "0" and "1" in τ and H indicate evaluation of these quantities at $r = r_0$ and $r = r_1$.

The difference between FRW solutions and other nonstatic SSSF solutions can be stated in more precise geometric terms. In the former solutions, each 3-D surface Σ_t is the orbit of the group under which it is invariant. For each surface Σ_t in the latter solutions, λ^a is only invariant under the action of SO(3) whose 2-D orbits (i.e., the two-spheres with proper radius R) are contained in the surfaces Σ_t. Since in these solutions the only symmetry group acting (in an intransitive manner) in the surfaces Σ_t is SO(3), these surfaces are only isotropic (in the geometric sense) with respect to observers comoving along a locus which marks the common center of the orbits of this group (see Sec. VIII and Sec. V of Part III). At this locus, usually (but not necessarily) labeled by $r = 0$, the four-acceleration vanishes and observers there do not detect any preferential direction. However, this feature happens in all spherically symmetric space-times in which such a "center" of symmetry exists, since any other possible kinematic effect, such as shear or viscosity, will depend only on r along the surfaces Σ_t and must vanish at the value of r marking the center. For observers not comoving along a center in nonstatic SSSF solutions (other than FRW), the four-acceleration gives a preferential direction along $\partial/\partial r$. Therefore, the kinematical isotropy of these solutions does not manifest itself in the form of a simple isotropical Hubble law as the geometric isotropy of the FRW solutions does. This aspect will be discussed in Sec. XII of Part III.

III. THE FIELD EQUATIONS

As Appendix A of Part I shows, the Einstein–Maxwell field equations for the metric (1) relate the state variables ρ, p, and q with the expansion Θ, the Hubble scale factor, and its derivatives $\dot{H}$, H', and H''. Since $R = fH$, these equations can be easily rewritten in terms of R and its derivatives. Therefore, the straightforward geometric interpretation of H and R discussed in the previous section suggests that either one of these quantities should be identified as the "metric potential" or gravitational field variable of the SSSF solutions. The choice of H or R could depend on the approach one is giving to a particular solution. For example, in a cosmological context H could be a better choice; but if the solution is being considered as modeling a star, then R is a better-suited variable to appear in equations of stellar structure. For charged solutions, the function $E(r)$ appearing in the Maxwell equation I(7b),

$$4\pi q(t,r) = fE'/R^3, \tag{14a}$$

is related to the electric field measured in the rest frame of the comoving observers,

$$F_{tr} = -(-g^{tt})^{1/2}E/f, \tag{14b}$$

which is given by I(7a).

In order to simplify the remaining field equations, it is necessary to eliminate derivatives, $\dot{H}$, H', and H'' in terms of H (or R) and E. As discussed in Appendix A of Part I, the derivative H'' can be eliminated by using the field equations in the form of the constraint $(G^r{}_r - G^\theta{}_\theta) = 8\pi(T^r{}_r - T^\theta{}_\theta)$, which leads to the "equation of pressure isotropy" I(15):

$$\left[\frac{H}{H}\right]' - \frac{f'H'}{fH} - \left[\frac{H'}{H}\right]^2 + \frac{3J}{f^3H} - \frac{2E^2}{f^4H^2} = 0, \tag{15}$$

where $J = J(r)$ is the same function appearing in Eq. (A5) in Appendix A of Part I. If the first and second integral of (15) are known, first-order derivatives like $\dot{H}$ and H' could also be eliminated in terms of H and E. This will be done in subsequent sections for the particular case of the ChKQ solutions, using the first and second integrals of Eq. (15) [or in the form of I(18)] obtained in Part I.

From Eq. (7), Θ is related to the proper time derivative of H (or R) along the surfaces r. Eliminating second-order derivatives of H (or R) with the help of (15), the field equation $G^t{}_t = 8\pi T^t{}_t$ becomes a first-order equation of motion for the comoving observers along the surfaces Σ_r. This field equation can be given in terms of H or R as

$$\begin{aligned}\left[\frac{dH}{d\tau}\right]^2 &= \left[\frac{\Theta H}{3}\right]^2\\ &= \frac{1}{H}\left[\frac{8}{3}\pi\rho H^3\right] - k + \frac{2J}{f^3H}\\ &\quad - \frac{E^2}{f^4H^2} + \frac{2f'H'}{fH} + \left[\frac{H'}{H}\right]^2,\end{aligned} \tag{16a}$$

$$\begin{aligned}\left[\frac{dR}{d\tau}\right]^2 &= \left[\frac{\Theta R}{3}\right]^2\\ &= -\left[1 - \frac{2}{R}\left(\frac{4}{3}\pi\rho R^3 + J\right) + \frac{E^2}{R^2}\right] + \left[\frac{fR'}{R}\right]^2,\end{aligned} \tag{16b}$$

where $J(r)$ is given by Eq. (A5) of Part I. The form (16a) of this field equation appears as a Friedman-like equation, while the form (16b) looks like a first-order stellar structure equation.

A useful relation between ρ' and the functions J and E follows by differentiating either one of equations (16) with respect to r, and eliminating H'' (or R'') with the help of (15), leading to

$$\tfrac{4}{3}\pi\rho' R^3 = -J' + EE'/R. \tag{17}$$

The specific form of ρ' illustrates how for SSSF solutions, for which (15) must be satisfied, the form of ρ along the surfaces Σ_t is severely restricted by the intrinsic curvature of these surfaces. This follows from the relation[14] between ${}^{(3)}\mathscr{R}$, the Ricci scalar of the surfaces Σ_t, ρ, and Θ:

$$[\Theta/3]^2 = \tfrac{8}{3}\pi\rho + \tfrac{1}{6}\,{}^{(3)}\mathscr{R}, \tag{18}$$

allowing one to identify, with the help of (16), the form of ${}^{(3)}\mathscr{R}$ as a function of H, J, and E. Bearing in mind that Θ is constant along the surfaces Σ_t, Eqs. (16)–(18) show that ρ is constrained to differ from the curvature scalar of each surface Σ_t, by a constant $[\Theta(t), t = \text{const}]$, albeit a different constant at each surface Σ_t. Since a solution of Eq. (15) determines H', it fixes ${}^{(3)}\mathscr{R}$ and ρ along each surface Σ_t up to (at most) three arbitrary constants: two integration constants of (15) and Θ at $t = \text{const}$. In general, these constants will be different along different surfaces Σ_t defining $\Theta(t)$ plus two arbitrary functions of time associated with the integration of (15).

The field equation $G^r{}_r = 8\pi T^r{}_r$ can be obtained from Eqs. (16) by eliminating p from the contracted Bianchi identity $u_\alpha T^{\alpha\beta}{}_{;\beta} = 0$,

$$\frac{d\rho}{d\tau} = -(p+\rho)\Theta\,, \tag{19}$$

and performing the time derivative of ρ using (16). This leads to the following Raychaudhuri equation:

$$\frac{1}{R}\frac{d^2R}{d\tau^2} = \frac{\Theta^2}{9} + \frac{d}{d\tau}\frac{\Theta}{3} = -4\pi\left(p + \frac{\rho}{3}\right) + a^\alpha{}_{;\alpha}\,, \tag{20}$$

where

$$a^\alpha{}_{;\alpha} = \Psi_{(2)} + (fR'/R)(f\mathscr{A}/R^2) \tag{21a}$$

and

$$\Psi_{(2)} \equiv (1/R^3)[-J + EE'/R]\,, \tag{21b}$$

with $\Psi_{(2)}$ and $\mathscr{A}$ being the invariant conformal scalar "psi-two" computed in Appendix A of Part I and the magnitude of the four-acceleration given by (9b). Equation (16) is a first integral of (20), and so any solution to the former will be a solution to the latter if (19) holds. By inserting the contracted Bianchi identity $h_{\alpha\beta}T^{\alpha\gamma}{}_{;\gamma} = 0$,

$$-p' + EE'/R = (\rho + p)\mathscr{A}\,, \tag{22}$$

into (21a), Eq. (20) becomes the stellar structure equation derived by Misner and Sharp[17] for the neutral case and by Beckenstein[18] for the charged case. From this point of view, one can identify

$$M(t,r) \equiv \tfrac{4}{3}\pi\rho R^3 + J \tag{23}$$

appearing in (16b) as a sort of "mass function,"[7,19,20] so that Eq. (16b), at each surface Σ_r, is formally analogous to the "energy equation" of a charged relativistic "particle" with time-dependent mass given by (23) and charge given by $E(r)$ in a central field under the effect of the "potential" (fR'/R). This type of formal analogy will not be developed any further.

The basic equations characterizing SSSF solutions are (15) and (16). The former is a sort of "transversal" geometric constraint expressing how ρ and the intrinsic curvature of the surfaces Σ_t are restricted by the imposition of zero shear in an adiabatic fluid flow. Equation (16), on the other hand, is an evolution equation whose solution indicates how this 3-D curvature changes along the surfaces Σ_r in a way compatible with (15). However, even with a solution of (15), Eq. (16) cannot be integrated unless the dependence of ρ on r and on H (or R) along the surfaces Σ_r is specified. This indeterminacy is the consequence of not having prescribed (yet) an equation of state, and will be discussed in Secs. VI and VII for the specific case of ChKQ solutions.

For charged solutions, the electric charge density q is connected to the field variables E and H (or R) by the Maxwell equation (14a). Since E' is completely fixed along the surfaces Σ_t by a solution of (15), q varies as H^{-3} along the surfaces Σ_r. From a physical point of view, E should be obtained by integrating (14a) given an initial charge distribution; however, Eq. (15) might be analytically solvable only for restricted functional forms of $E(r)$ and $J(r)$. The first and second integrals of (15) obtained in Part I will be reviewed in the next section.

IV. McVITTIE-TYPE AND WYMAN-TYPE SOLUTIONS

The ChKQ solutions, derived and classified in Part I, form a large subclass of SSSF solutions. Therefore, they are described by the metric (1) with H satisfying the constraint (15) [simplified in the form of Eq. I(18)] under the restrictions given by I(22a) and I(22b). These restrictions fix the functions J [Eq. (A5) in Appendix A of Part I] and E as $J = \mu(fh)^5$ and $E = \pm\epsilon(fh)^3$, where the functions f and h are defined by Eqs. I(17) and I(23) in terms of $y(r)$, defined by I(16b).

A first integral of (15) is given by Eqs. I(21), which, translated in terms of H or R, is Eq. I(40):

$$\frac{fR'}{R} = f' + f\frac{H'}{H} = \frac{(fh)'}{h} \pm (fh)^2\mathrm{Q}^{1/2}\,. \tag{24}$$

This Q is related to the quartic Q of I(24) by

$$\mathrm{Q} \equiv W^{-2}Q = \Delta - 2\mu W + \epsilon^2 W^2 + LW^{-2}\,, \tag{25a}$$

$$W \equiv h/H = fh/R\,, \tag{25b}$$

where the constants Δ, μ, ϵ, and $L = L(t)$ are defined in I(21c). As discussed in Sec. IV of Part I, ChKQ solutions can be divided in three subclasses, depending on the form of the function $L(t)$ appearing in (25a): (1) if $L = \text{const}$, one has the McVittie-type $(L = 0)$ and Wyman-type $(L = \text{const} \neq 0)$ solutions (see Sec. IV Of Part I); (2) the conformally flat subclass with $L = L(t)$, but $\epsilon = \mu = 0$ (i.e., $\Psi_{(2)} = 0$); and (3) the most general case with $L = L(t)$ and $\Psi_{(2)} \neq 0$, i.e., ϵ and μ nonzero and, for neutral solutions, $\mu \neq 0$. In order to simplify the study of ChKQ solutions, in this paper I will consider only solutions belonging to case (1), while cases (2) and (3) will be examined in Secs. VII and IX of Part III. The McVittie- and Wyman-type solutions, will be referred to hereafter as M- and W-type solutions.

Using Eqs. (24) and (25), Eqs. (16) can be written completely in terms of H (or R) and the functions J and E. The Friedman-like equation (16a) now becomes

$$\begin{aligned}\left[\frac{dH}{d\tau}\right]^2 = \frac{1}{H}\left[\frac{8}{3}\pi\rho H^3\right] - k + (fh)^2LH^2 \\ + f^2\left[\frac{h_y}{h}\right]^2 + 2ff_y\left[\frac{h_y}{h}\right] + \Delta(fh)^4 \\ \pm 2(hH)^2\left[ff_y + f^2\left[\frac{h_y}{h}\right]^2\right]\mathrm{Q}^{1/2}\,,\end{aligned} \tag{26}$$

where

$$h_y \equiv \frac{dh}{dy} = f^{-1}h'\,,$$

where (16b) becomes a similar expression once (24) is explicitly substituted. The three first terms in the right-hand side of (26) are analogous to those found in a Friedman equation for the FRW solutions. The third term, which is a constant along the surfaces Σ_r (though a different constant for different comoving observers), is analogous to a sort of position-dependent "cosmological constant." If $L = \text{const}$, Eqs. (24) and (25) are sufficient to write the Raychaudhurri equation (20) without first-order derivatives, since in this case the function $\mathscr{A}(t,r)$ defined by (9b) can be obtained from (25) with the help the integrability condition

$$[\dot{H}/H]' = [H'/H] \tag{27}$$

leading to

$$\mathscr{A}(t,r) = \frac{fh^2}{Q^{1/2}}[\epsilon^2 W^2 - \mu W - LW^{-2}]$$
$$= \frac{fh^2}{H}\frac{\epsilon^2 h^2 - \mu hH - Lh^{-2}H^4}{[\Delta H^2 - 2\mu hH + \epsilon^2 h^2 + Lh^{-2}H^4]^{1/2}}, \tag{28}$$

where W is given by (25b). With the help of Eq. (28), the term $a^\alpha_{;\alpha}$ appearing in (20) and (21) can be written explicitly. The M- and W-type solutions are characterized by six constant parameters a, b, c, ϵ, μ, and L according to which these solutions have been classified in the tables of Part I. Therefore, each specific combination of these parameters leads to a specific form for the functions in (24), (25), and (28) and the field equations (14a), (20), and (26). However, it is still necessary to compute the derivative $\partial H/\partial t$ in order to be able to fix the time coordinate and have a link between Θ and the evolution of H in terms of the chosen time coordinate. These aspects will be discussed in the following section.

V. THE TIME COORDINATE

While the radial dependence of H is completely fixed by the integration of the constraint equation (15), the time dependence of H is contained in (at most) two arbitrary functions of time: $T(t)$ and $L(t)$, which appear in I(24a) as integration "constants" of (15). These functions are the time-dependent free parameters of the solutions, and will be referred to as the "t parameters." The t parameters appear in the explicit forms of H given by I(28), I(30), I(32), I(43), I(45), and I(46) obtained by inverting the second integral of (15) in the form of I(24a).

As shown by Eq. I(29), the t parameter $T(t)$ can always be rewritten in terms of the elliptic integral in I(24a) evaluated at $r = r_0 \geqslant 0$ fixed but arbitrary. Since $T(t)$ is the only t parameter of M- and W-type solutions, if I(29) is used, $H_0(t)$ replaces T as the t parameter [see Eqs. I(30)]. Hence, the t parameter $T(t)$, or any rescaling of it, can always be related to the Hubble scale factor H along an arbitrary surface Σ_r. Though the convenience of having the t parameter given as $T(t)$ or as $H_0(t)$ might depend on the specific problem being considered.

In Part I, the time coordinate was left unspecified up to a rescaling of the form $t = t(t^*)$. In the metric (1), the expansion Θ has been kept so far as an unspecified function, though as mentioned in Part I, its appearance in g_{tt} cannot be "absorbed" into a new time coordinate $t^* = \int(3/\Theta)\,dt$ before computing $\dot{H}/H$ explicitly. Once this is done, different choices of time coordinate will arise by demanding specific relations between Θ and the derivatives of the t parameter. However, because of the geometric interpretation of Θ, H, and R, the condition

$$\Theta(t) = 0 \Rightarrow \dot{H}(t,r) = 0 \tag{29}$$

must hold for whatever choice of time coordinate and t parameters. The converse of (29) does not hold in general, as will be discussed in Secs. VIII and X, it leads to $|U(t,r)| \to \infty$, where U is given by Eqs. (3).

From I(24a) it is possible to compute $\partial H/\partial t$ explicitly by following the rules of differentiation of an elliptic integral $\mathrm{F}[\Psi,\eta]$ which it treated as a function of the independent variables; $\Psi(W,\eta)$ and η (Ref. 21). Since the time dependence of the modulus η is contained in the t parameter $L(t)$ (see Sec. II of Part I), if this function is set to a constant, η will also be constant and $\mathrm{F} = \mathrm{F}[\Psi(W)]$. In this case, comprising M- and W-type solutions, depending on the choice of T or H_0 as t parameter, the derivative $\partial H/\partial t$ turns out to be

$$\dot{H}/H = -Q^{1/2}\,\dot{T} = [Q/Q_0]^{1/2}(\dot{H}_0/H_0), \tag{30}$$

where Q is given by (25a). This derivative is computed for the cases with $L = L(t)$ in Secs. VII and IX of Part III. By inserting Eq. (30) in the metric (1), various choices of time coordinate for the ChKQ solutions can be made. These choices, each of which can be associated with a specific form of the t parameter, are the following.

(1) Keep $T(t)$ as the t parameter and choose t such that

$$\Theta = -3\dot{T} \tag{31a}$$

$$\Rightarrow ds^2 = -Q(t,r)\,dt^2 + H^2(t,r)\times[dr^2 + f^2(d\theta^2 = \sin^2\theta\,d\phi^2)]. \tag{31b}$$

(2) Choose t to be the proper time for comoving observers along r_0, this choice leads to

$$\Theta/3 = \dot{H}_0/H_0 = -[Q_0]^{1/2}\,\dot{T} \tag{32a}$$

$$\Rightarrow ds^2 = -\frac{Q(t,r)}{Q_0(t)}dt^2 + H^2(t,r)\times[dr^2 + f^2(d\theta^2 + \sin^2\theta\,d\phi^2)]. \tag{32b}$$

Obviously, the t parameter H_0 is better suited for this coordinate choice, however, T can be used also provided that Θ is given by (32a).

(3) Choose the t parameter in either one of the forms, T or H_0, as time coordinate. Then the expansion $\Theta(T)$, or $\Theta(H_0)$, becomes the t parameter. If T is chosen as time coordinate, the metric becomes

$$ds^2 = -\frac{Q(T,r)}{[\Theta(T)/3]^2}dt^2 + H^2(T,r)\times[dr^2 + f^2(d\theta^2 + \sin^2\theta\,d\phi^2)], \tag{33}$$

with

$$\Theta(T) \neq 0.$$

If H_0 is chosen as time coordinate, the metric takes a form similar to (33).

Each one of the coordinate choices presented above is entirely equivalent, though one coordinate choice can be more useful than another for the study of a given aspect of the solutions. For example, if it is convenient to single out the comoving observers along a particular surface Σ_r, then the choice (32) with t being the proper time at that surface Σ_r can be more convenient. Such a situation could arise in the case of a bounded sphere, in which case the special surface Σ_r would be the surface at which the sphere is matched with the external Schwarzschild or Reissner–Nordstrøm field (see Sec. XI). Ideally, H_0, or $R_0 = f_0 H_0$ could be the t parameter, but as Eqs. I(30) show $H(H_0,r)$ can be quite cumbersome, and so this coordinate choice will have, in general, complicated metric coefficients. The coordinate choice (31) does not single out any surface Σ_r and leads to the simplest form of the metric; however, the equation determining $T(t)$ could be more complicated. Since it can be helpful if one needs to perform calculations involving the metric, this choice is the one usually followed by authors who have studied particular cases of the ChKQ solutions. The choice (33) has a coordinate singularity at $\Theta(T) = 0$, but has the advantage that the metric coefficient H is already determined as a function of the coordinates T and r. However, $\Theta(T)$, which is now the t parameter appearing in the metric, might be a complicated function of T or H_0. This choice has been favored in the study of the Wyman solution and its charged version.[11,12,22] The three types of time coordinate presented above will be used throughout this paper and Part III, specifying in each case which choice has been made.

VI. THE EQUATION OF STATE

As mentioned in Sec. II, the determination of the time evolution of the fluid by solving the field equations requires one to supply to these equations extra information contained in an "equation of state," i.e., a constraint relating the state variables ρ and p. However, it is desirable to choose an equation of state that not only determines the time evolution of the fluid, but at least satisfies the "weak energy condition,"[14] and bears some minimal relation to the physical properties of the fluid. Since the case of SSSF solutions satisfying a barotropic equation of state has been extensively discussed in the literature (Refs. 11, 12, 19, 20, and 22, and also Refs. 1, 5, and 8–10 of Part I), this section will concentrate mainly on nonbarotropic equations of state. The Wyman solution, satisfying a barotropic equation of state is discussed in Appendix A. The guidelines to be followed in order to prescribe or obtain an equation of state, together with some thermodynamical properties of the solutions, will be discussed in this section. A kinetic theory approach will be left for Sec. XI of Part III. For the sake of simplicity, the coordinate choice (33) will be used and only neutral fluid configurations will be considered.

For a perfect fluid configuration, one defines the particle number and entropy densities, N and S, respectively, which satisfy

$$(Nu^\alpha)_{;\alpha} = \frac{dN}{d\tau} + N\Theta = 0\,, \tag{34a}$$

$$\frac{dS}{d\tau} = 0\,. \tag{34b}$$

Equation (34a) is a continuity equation implying the conservation of the number of baryons in the fluid, while (34b) simply states that the entropy density is constant along the world lines of fluid flow. An equation of state can be formally given by specifying the matter-energy density considered as a primary thermodynamical potential[23]:

$$\rho = \rho(N,S). \tag{35}$$

Given an equation of state in the form (35), other thermodynamical state variables, such as pressure $p(N,S)$ and temperature $\mathscr{T}(N,S)$ can be obtained through the first law of thermodynamics: $d\rho = [(\rho + p)/N]dN + [N\mathscr{T}]dS$, which can be applied to SSSF solutions by "translating" the thermodynamical variables (or parameters) S and N into the coordinates used in (1). For this representation, Eqs. (34) become

$$N^{-1} = \tfrac{4}{3}\pi H^3\,, \tag{36a}$$

$$S = S(r)\,, \tag{36b}$$

which relate the "thermodynamical" coordinates (N,S) to the "geometric" coordinates (H,r). Thus the surfaces of constant entropy and particle number densities are the surfaces Σ_r and $H = \text{const}$, respectively. Equation (36a) conceals an arbitrary function of r by assuming that the initial particle number density is given by $[(4/3)\pi H]^{-1}$ evaluated in an arbitrary initial surface Σ_t. In Eq. (36b), $S(r)$ is an unspecified function because for a perfect fluid the only information given about it by the second law of thermodynamics [in (34b)] is that it is conserved along the fluid world lines, and thus, like the initial particle number density, it is not essential to determine the motion of the fluid. However, in order to translate into geometric coordinates the information conveyed in the functional dependence of thermodynamical variables on S, the function $S(r)$ must be somehow known or inferred.

Using Eqs. (36), the first law of thermodynamics can be projected along the directions parallel and orthogonal to u^α. Considering that $dN = \dot{N}\,dt + N'\,dr$ and $dS = S'\,dr$, the relations $(\partial\rho/\partial N)_s = (\rho + p)/N$ and $(\partial\rho/\partial S)_N = \mathscr{T}S$ become the Bianchi identity (19) plus

$$-\tfrac{4}{3}\pi\rho' H^3 + (\tfrac{4}{3}\pi H^3)'(p+\rho) = \mathscr{T}S'\,, \tag{37a}$$

which for M- and W-type solutions becomes

$$-\frac{J'}{R^3} + \left[\frac{h'}{h} - f h^2 Q^{1/2}\right]\left[{}^{(3)}\mathscr{R} + a^\alpha{}_{;\alpha} + \frac{1}{3}\frac{d\Theta}{d\tau}\right] = \frac{\mathscr{T}S'}{H^3}\,, \tag{37b}$$

where Eqs. (16), (17), (20), (25), and (36) have been used to eliminate ρ', H', $p+\rho$, and N. Equations (37) convey to the thermodynamical state variables $\mathscr{T}$ and S the "transverse" geometric constraints that shear-free motion imposes on the surfaces Σ_t, and thus it is in itself a strong constraint restricting possible choices of equations of state.

If $S' = 0$ in (37a), one has the conditions that the parameters characterizing ChKQ solutions must satisfy in order for these solutions to be compatible with a barotropic equation of state $p = p(\rho)$. For the particular case of FRW solutions, $S' = 0$ in (37a) is trivially satisfied implying that these solutions always admit a barotropic equation of state. For nonstatic ChKQ solutions other than FRW, inserting

$S'=0$ in Eq. (37a) leads directly to the parameters characterizing the Wyman solution (see Appendix A). If $S'\neq 0$, an equation of state of the form (35) can be either prescribed or obtained in connection with Eq. (16). If this equation is treated as an equation of motion, then it cannot be integrated unless one knows how ρ depends on H (or R) along every surface Σ_r, i.e., $\rho(H,r)$. Conversely, if Eq. (16) in the form (18) is used a sort of "definition" of ρ, then this definition is undetermined unless one specifies Θ, which implies specifying *a priori* how the fluid evolves in time and then finding out the matter-energy density associated with this motion. Depending on how one approaches this field equation, an equation of state of the form (35) can be either "prescribed" or "obtained." That is, based on (16), two strategies are possible: (a) the motion of the fluid is obtained by finding Θ from a prescribed ρ, and (b) the matter-energy density is obtained from a prescribed Θ (i.e., a prescribed motion of the fluid). Obviously, (a) is a physical strategy, and so, an equation of state which follows this strategy is a "physical equation of state," while (b) is a "kinematical strategy" associated to a "formal equation of state." If a physical equation of state is prescribed, it determines $p(N,S)$ and $\mathscr{T}(N,S)$ through the first law of thermodynamics and can be translated into geometric terms by Eqs. (40). If a formal equation of state is obtained by a specific choice of Θ and S, then Eq. (37b) is a sort of "geometric definition" of $\mathscr{T}S'$.

Although the prescription of any function Θ for arbitrary $S(r)$ is sufficient to have the field equations determined, it is desirable to avoid dealing with fluids divorced from any physical concern. Therefore, the guideline to obtain a formal equation of state should follow some sort of physical justification, such as choosing Θ from a boundary condition or to comply with a given form of $S(r)$. Since there are very few examples in the literature[24] of nonbarotropic physical equations of state that could be applied to simple thermodynamical systems like a perfect fluid configuration, practically all authors studying SSSF solutions (except FRW solutions) have used formal equations of state, either obtained as boundary conditions or, in the case of authors studying the Wyman solution,[11,12,22] by constructing an equation of state compatible with $S=$ const. Such an equation of state is, of course, a formal one. See Appendix A.

Whether "physical" or "formal," an equation of state must be compatible with the field equations. This compatibility can be tested by Eq. (37), which together with the Bianchi identity (19) forms a coupled system of integrability conditions of ρ, relating the state variables ρ, $\mathscr{T}$, and S with Θ and $d\Theta/d\tau$, and the "curvature terms" $a^{\alpha}{}_{;\alpha}$ and ${}^{(3)}\mathscr{R}$, which are fully determined by the constraint equation (15) irrespective of the choice of equation of state. Because of the presence of these "curvature" terms, it might be impossible to propagate along u^{α} a given set of physically motivated initial conditions specified by relations between the state variables (p and ρ, or S and $\mathscr{T}$) in an initial surface Σ_t, and at the same time to have a fluid performing shear-free motion so that Eqs. (15), (16) [or (19)], and (37) are satisfied at every subsequent surface Σ_t. For example, since for both types of equations of state S and $\mathscr{T}$ appear in (37) intertangled in the product $\mathscr{T}S'$, these variables can be separated by inserting in (37) either an empirical or a theoretical relation between them. However, such a physically motivated relation could very well be incompatible with the curvature terms ${}^{(3)}\mathscr{R}$ and $a^{\alpha}{}_{;\alpha}$. Since perfect fluid thermodynamics do not give any further insight into the relation between $\mathscr{T}$ and S, the separation of these variables in the product $\mathscr{T}S'$ requires one to study the thermodynamics of fluids in small deviations from thermal equilibrium and/or kinetic theory. The former approach will be discussed below, while the latter will be left for Sec. XI of Part III.

At first-order deviations from equilibrium,[25–27] the heat flux vector q^{α} is given by

$$q^{\alpha}=-\kappa h^{\alpha\beta}[\mathscr{T}_{,\beta}+\mathscr{T}a_{\beta}], \tag{38}$$

where $\kappa=\kappa(N,S)$ is the coefficient of thermal conductivity and a_{β} is the four-acceleration defined by (9), or (28) for M- and W-type solutions. If the energy-momentum tensor is that of a perfect fluid, then q^{α} must vanish leading to the following two options:

$$(1)\quad \kappa=0,\quad \mathscr{T}\ \text{undetermined}, \tag{39a}$$

$$(2)\quad \kappa>0,\quad (\ln\mathscr{T})_{,\beta}+a_{\beta}=0. \tag{39b}$$

Condition (39b) reduces for static space-times to the "Tolman law,"[28] which governs the temperature gradient of such configurations. For SSSF solutions, this condition is $\mathscr{T}=U(t,r)$, with U given by (3b), so that the vector λ^{α} in (10) could be identified as $\lambda^{\alpha}=\mathscr{T}u^{\alpha}$. For the particular case of M- and W-type solutions, condition (39b) reads

$$\mathscr{T}(T,r)=\Theta(T)/3/[Q(T,r)]^{1/2}, \tag{40}$$

which fixes $\mathscr{T}$ for $\kappa>0$. If (40) is inserted in (37b), the latter equation becomes a "definition" of S' from which $S(r)$ is found by integration. However, Eq. (37b) is incompatible with having $S=S(r)$ [i.e., incompatible with (34b)], as it can be verified by writing up explicitly the terms ${}^{(3)}\mathscr{R}$, $a^{\alpha}{}_{;\alpha}$, and J' appearing in (37b). This incompatibility occurs for all nonstatic ChKQ with $S'\neq 0$, and it is a simple example of how the imposition of shear-free motion, which completely fixes ${}^{(3)}\mathscr{R}$ and $a^{\alpha}{}_{;\alpha}$, restricts the forms of state variables and equations of state. Therefore, the only option for these solutions is to assume that $\kappa=0$, which is a strong restriction on their range of applicability as models of physical materials. Possible kinetic theory implications of having $\kappa=0$ will be discussed in Sec. XI of Part III.

VII. BOUNDARY CONDITIONS AS FORMAL EQUATIONS OF STATE

The simplest type of formal equation of state is provided by suitable boundary conditions which are sufficient conditions to have the field equations fully determined. This choice of equation of state follows from a kinematical strategy, and is the one usually followed by authors who have examined ChKQ solutions. From Eq. (18), a formal equation of state can be obtained by specifying $\Theta=\Theta(t)$, and a simple way to do so consists in finding ρ from a given set of physical conditions in an arbitrary surface Σ_r which will be called the "boundary Σ_r surface." Using the coordinate choice (32) and choosing a boundary Σ_r surface corresponding to a fixed but arbitrary $r = r_0 \geqslant 0$, a physically motivated boundary condition can be specified by a reasonable barotropic equation of state $p_0 = p_0(\rho_0)$ valid only at r_0. This type of equation of state will be referred to hereafter as a "localized equation of state." Given such an equation of state, the Bianchi identity (19) at r_0,

$$\frac{d\rho_0}{p_0(\rho_0) + \rho_0} = -\frac{3\, dH_0}{H_0}, \tag{41}$$

can be integrated yielding: $\rho_0 = \rho_0(H_0)$. Knowing how ρ_0 depends on H_0, the time evolution field equation (16) restricted to r_0 can be integrated. This equation can be expressed either in terms of H_0 as

$$\begin{aligned}(\dot{H}_0)^2 = {} & (1/H_0)(\tfrac{8}{3}\pi\rho_0 H_0^3) - k \\ & + [f_0^2 (h_y/h)^2 + 2(1 - ky_0)(h_y/h)_0] \\ & + L(f_0 h_0)^2 H_0^2 \\ & + 2h_0^2[1 - ky_0 + f_0^2(h_y/h)_0][Q_0]^{1/2},\end{aligned} \tag{42a}$$

or in terms of $R_0 = f_0 H_0$ as

$$\begin{aligned}(\dot{R}_0)^2 = {} & -[1 - (2/R_0)(\tfrac{4}{3}\pi\rho_0 R_0^3 + J_0) + E_0^2/R_0^2] \\ & + [1 - ky_0 + f_0^2(h_y/h)_0 \pm (f_0 h_0)^2 [Q_0]^{1/2}]^2,\end{aligned} \tag{42b}$$

where $h_y \equiv dh/dy$, and $(h_y/h)_0$ indicates that the quantity in brackets is evaluated at r_0. If $r_0 = 0$ (if space-time is regular there, see Sec. VIII) is chosen as the boundary surface, Eq. (42a) simplifies to

$$\begin{aligned}(\dot{H}_c)^2 = {} & (1/H_c)(\tfrac{8}{3}\pi\rho_c H_c^3) - k - 2(h_y/h)_c \\ & + 2h_c^2[Q_c]^{1/2},\end{aligned} \tag{42c}$$

where the subindex c indicates evaluation at $r = 0$. Equations (42) look like Friedman equations for FRW solutions, however, there are extra curvature terms that come from evaluating derivatives like H' at fixed r. Since these derivatives vanish for the FRW solutions, these curvature terms do not appear in the corresponding Friedman equations.

Once Eqs. (42) are integrated yielding $H_0 = H_0(t)$, the time evolution of comoving observers at the boundary Σ_r surface is known. However, since H can be given as $H_0 = H(H_0, r)$ [see Eqs. I(30) and (13)], the time evolution of all comoving observers with $r \neq r_0$ can also be found. From another angle, a localized equation of state fixes Θ through either one of Eqs. (42) and thus determines ρ through (18). The resulting formal equation of state can be obtained in the form (35) by translating $\Theta(t)$ into "thermodynamical" coordinates, which involves eliminating t in terms of H and r and choosing $S(r)$ so that $\rho = \rho(H,r) = \rho(N,S)$. The prescription of a localized equation of state for a given ChKQ solution is a mathematically convenient equation of state but because of the presence of curvature terms such as ${}^{(3)}\mathscr{R}$ and $a^{\alpha}{}_{;\alpha}$, it is unlikely to be physically realistic for $r \neq r_0$.

The choice of boundary Σ_r surface depends on the specific problem under consideration. For the purpose of using ChKQ solutions as bounded spheres matched to a Schwarzschild or Reissner–Nordstrøm exterior, r_0 should label the matching surface (the surface of the sphere) so that the time coordinate coincides with the proper time of observers comoving with this surface (see Sec. XI and Appendix C). Most previous work on ChKQ solutions is concerned with their use as bounded spheres, though the time coordinate choice usually favored is (31). In Sec. XI ChKQ solutions modeling bounded spheres will be concentrated on. For unbounded fluid configurations, the choice $r = 0$ (if space-time is regular at this locus, see Sec. VIII) seems to be better suited. For M- and W-type solutions, L is a constant, and then the prescription of a localized equation of state determines the only t parameter H_0 [or $T(t)$ depending on the coordinate choice]. However, if $L = L(t)$, then this extra t parameter allows one to prescribe extra conditions on the state variables (see Secs. VII and IX of Part III). In the following sections, and unless it is specified otherwise, ChKQ solutions will be examined assuming that a given localized equation of state has been prescribed. Although these equations of state are purely formal, they are mathematically simple and are useful inasmuch as they provide one with examples of ChKQ solutions in which the time evolution is fully determined.

VIII. REGULARITY CONDITIONS

Since ChKQ solutions, as spherically symmetric space-time manifolds, can be decribed in 2-D coordinate patches (t,r) (see Sec. III of Part III), all relevant quantities associated with these solutions, whether metric coefficients, state variables ρ, p, and q, or curvature scalars, can be treated as real functions of t and r. However, coordinate values t and r are restricted by regularity conditions ensuring that these functions are smooth and bounded, i.e., at least C^1 functions. Some of these conditions arise from the specific functional form of curvature terms present in the field equations independently of the choice of equation of state, and thus are basic regularity conditions. Other conditions follow from a specific choice of equation of state and thus are supplementary regularity conditions. Both types of conditions will be discussed below.

A. Regularity conditions for the metric

Leaving aside the coordinate singularity arising as $\Theta = 0$ using the time coordinate choice (33), the regularity of the metric (1) follows from the regularity of H and $R = fH$, treated as non-negative functions of t and r. In Part I, the most general form of H was that of a quotient of elliptic functions which can be given generically as

$$H(t,r) = h(r)\Xi(t,r)/\Pi(t,r), \tag{43}$$

which indicates that a necessary condition for the metric coefficients in (1) to be smooth and bounded is

$$|\Pi(t,r)| > 0. \tag{44}$$

Since the function $\Pi(t,r)$ becomes a constant if $L = 0$ (see tables of Part I), that is M-type solutions, condition (44) can only fail to hold in solutions with $L \neq 0$ (though there are exceptions to this rule, see Sec. IV of Part III). However, besides condition (44), H (and/or R) could diverge if the functions h and $X \equiv \int h^2\, dy$ given by I(23) and I(25) diverge. These functions, which appear as arguments of Π and

Ξ in the expressions of the type (43) derived in Part I, will be smooth and bounded if, for each category $X_{(1)}$ to $X_{(5)}$ in Eq. I(25), the values of the parameters (a,b,c) characterizing h are restricted so that $ay^2 + 2by + c > 0$ for $r \geqslant 0$ [or $y \geqslant 0$, see Eq. I(16b)].

B. Regularity at the center

In a spherically symmetric space-time, a regular center can be defined[17] as the set of events corresponding to fixed points of the group SO(3), that is, the world line of the common center of the orbits of SO(3) labeled by a fixed coordinate $r = r_c$. In solutions in which a regular center exist, the conditions[17]

$$R(t,r_c) = \dot{R}(t,r_c) = 0\,, \tag{45a}$$

$$[fR'/R]_{r=r_c} = 1 \tag{45b}$$

will hold for $r = r_c$. For solutions in which $h(0)$ and $X(0)$, and so $H(0,t)$ and $\dot{H}(0,t)$, are bounded, one can identify the locus $r_c = 0$ as a regular center. Since $R = fH$, $f(0) = 0$ and $f'(0) = 1$ clearly hold if $H(0,t)$ and $\dot{H}(0,t)$ are bounded. If $k = 1$ as in (2b), then $f = 0$ and $f' = 1$ for $r_c = \pi$, and so conditions (45) also hold also at this locus. This effect occurring for solutions with $k = 1$ will be discussed in Secs. II and III of Part III. Throughout this paper it will be assumed that (a,b,c) have been chosen so that conditions (45) hold, and this requires specifically avoiding solutions corresponding to $X_{(1)}$ with $c = 0$, $X_{(2)}$ with $b > 0$ and $c = 0$, and $X_{(4)}$ with $b = 0$ [see Eqs. I(25)], in which h and X (and so H) diverge as $r \to 0$. Solutions in which h and X diverge at $r = 0$, and so conditions (45) are violated at this locus, will be examined in Secs. V and VI of Part III (see also Appendix B).

C. Regularity of curvature scalars

Curvature scalars formed with contractions of the Riemann tensor, such as $g_{\alpha\beta}\mathscr{R}^{\alpha\beta}$, $\mathscr{R}_{\alpha\beta}\mathscr{R}^{\alpha\beta}$, or $\mathscr{R}_{\alpha\beta\gamma\delta}\mathscr{R}^{\alpha\beta\gamma\delta}$, which can be expressed as algebraic combinations of ρ, p, and q, will be regular if these functions are smooth and bounded. These state variables can be computed from (14a), (16) [or (18)], (20), and (21), which contain "curvature" terms, such as ${}^{(3)}\mathscr{R}$ and $a^{\beta}{}_{;\beta}$, whose form as functions of H is independent of the choice of equation of state. An examination of the field equations shows that all state variables ρ, p, and q diverge as $H \to 0$, and since R could vanish regularly at $r = 0$ [because of (45a)], the condition that H vanishes can be expressed in terms of R as $R \to 0$ for $r > 0$. Therefore, one basic regularity condition involving the state variables is

$$\left.\begin{matrix} R(t,r) > 0\,, & \text{for } r > 0 \\ H(t,r) > 0 & \end{matrix}\right\} \Rightarrow |\Xi(t,r)| > 0\,. \tag{46}$$

From (8) and (9b), the term $d\Theta/d\tau$ in the Raychaudhuri equation (20) and the magnitude of the four-acceleration $\mathscr{A} = |a_\beta a^\beta|^{1/2}$ diverge if the metric coefficient $U^{-1} = (-g_{tt})^{1/2}$ vanishes. This requires, for ChKQ solutions in general, the converse of condition (29) to hold, that is, the fact that $\dot{H}/H$ can vanish for $\Theta \neq 0$. For M- and W-type solutions, in which $\dot{H}/H$ is given explicitly by (30), the converse of condition (29) holds if $Q = 0$ with Q given by (26a). Since Eqs. (20)–(22) define p and p', the pressure and pressure gradient also diverge for any equation of state as long as there are t and r values for which Q vanishes. However, the matter-energy and charge densities remain finite as $Q \to 0$. Thus a second basic regularity condition ensuring that p, p', and curvature scalars remain bounded is that $\dot{H}/H$ should only vanish if Θ does. For M- and W-type solutions, this condition is

$$|Q(t,r)| > 0\,. \tag{47}$$

For solutions with time dependent L, the converse of condition (29) takes a more complicated form and will be discussed in Secs. VII and IX of Part III. The behavior of curvature scalars as $\Pi \to 0$ and as $r \to 0$ in the cases where conditions (45) are violated will be discussed in Secs. V and VI of Part III (see also Appendices A and B of Part III).

D. Regularity conditions following from a choice of equation of state

Regularity conditions (45)–(47) are necessary and sufficient. As shown in Appendix C, whether these conditions hold or not depends on the parameters $(\epsilon,\mu,\Delta,L,a,b,c)$ characterizing specific particular solutions. However, for those solutions in which H and/or Q could vanish and so (46) and/or (47) would be violated, an equation of state can be chosen in such a way that (t,r) values are further restricted and the fluid evolves in such a way that H and/or Q do not vanish. Such a choice of equation of state provides further

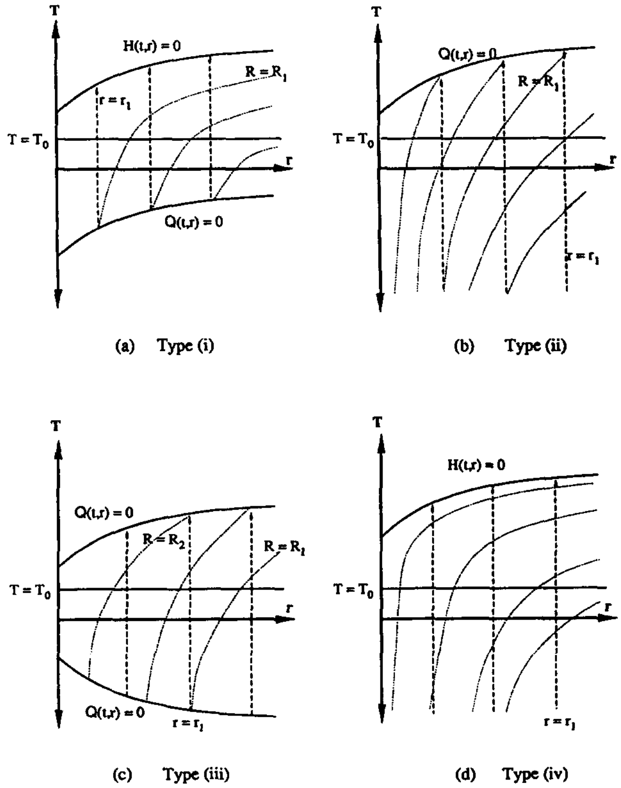


FIG. 1. Four types of domain of regularity. The evolution of fluid layers (world lines of comoving observers, vertical dotted lines) for the types (i) to (iv) of domain of regularity are displayed from (a) to (d). Dotted curves denote hypersurfaces of constant R. The arrows indicate a collapsing fluid motion, $\Theta < 0$ (towards decreasing R). For an expanding fluid, $\Theta > 0$, the arrows would point downward (toward increasing R). If $\Theta = 0$ at $T = T_0$, the fluid bounces and the arrows would reverse at this surface Σ_T. See Sec. VIII.

supplementary regularity conditions by setting Θ [the part of ρ and p not fixed by solving Eq. (15)] to be such that $H>0$ and/or $|\mathrm{Q}|>0$ for all (t,r). This situation requires Θ to vanish at a given surface Σ_t so that fluid layers "bounce" (see Fig. 1), and so it is analogous to what happens in the FRW solutions when negative pressure is introduced through the "cosmological constant," thus ensuring that the only regularity condition $H(t)>0$ holds for all t. However, in the latter solutions ${}^{(3)}\mathscr{R}$ diverges as $H\to 0$, and there is no acceleration term $a^{\alpha}{}_{;\alpha}$ which could diverge at $H\neq 0$ independently of the specification of an equation of state.

E. Domain of regularity

The range of (t,r) values ($r\geqslant 0$) in which the necessary regularity conditions (44), (46), and (47) hold is a domain of regularity characteristic of each ChKQ solution with a "regular center" [though condition (47) is only valid for M- or W-type solutions] independently of the choice of equation of state. If these conditions are phrased in terms of W defined by (26b), one has for M- and W-type solutions: $W>0$ and $Q(W)>0$, which are just the requirements fixing the range of W as integration variable in I(24a), Fig. 1 of Part I. Since the range of W, which directly depends on the roots of Q (i.e., which in turn are the same as the roots of Q), has been classified for ChKQ solutions in four basic types in Fig. 1 of Part I, these types are essentially the four types of domain of regularity of the solutions. As $W=fh/R$, the information contained in the types of domain of regularity can be conveyed in terms of allowed values of R for comoving observers with $r>0$ [$R=0$ for $r=0$ if conditions (45) hold]. Thus infinite values of the integration variable W can be identified with the boundary of the domain of regularity characterized by $R\to 0$. Similarly, the limits of integration of I(24a) given by $W\to A$ (A is a root of Q) and $W\to 0$ can be, respectively, identified with the boundaries of the domain of regularity given by $\mathrm{Q}\to 0$ and $R\to\infty$.

As a result of the invariant meaning of R discussed in Sec. II, the range of allowed values of this function is an important invariant characterization of each solution. This range of R is shown in Fig. 1 for the four types of domain of regularity (i)–(iv). Consider a typical fluid layer labeled by, say, $r=r_1>0$, if the solution is type (i) [Fig. 1(a)], R is constrained to vary between 0 and f_1h_1/A, where A is a root of Q. For solutions of type (ii) and (iv) [Figs. 1(b) and 1(d)], one has $R>f_1h_1/A$ and $R>0$, respectively, while in solutions of type (iii) [Fig. 1(c)] fluid layers evolve between two branches of $Q=0$ (i.e., $f_1h_1/B<R<f_1h_1/A$, where A and B are two consecutive roots of Q). Coordinate representations of the domain of regularity and of the boundaries $\mathrm{Q}=0$ and $H=0$ for various solutions classified in Tables III and VI of Part I are provided in Table I.

IX. COMPLETENESS OF CAUSAL CURVES

As mentioned in the previous section, if conditions (45) hold, curvature scalars formed by contracting the Riemann tensor diverge at the coordinate values given by the following constraints:

$$H(t,r)=0\Rightarrow (g_{rr})^{1/2}=0\,, \tag{48}$$

$$\mathrm{Q}(t,r)=0\Rightarrow (-g_{tt})^{1/2}=0\,, \tag{49}$$

while curvature scalars are bounded but the metric coefficients H and R diverge at those coordinate values for which

$$\Pi(t,r)=0\,, \tag{50}$$

where Π follows from the generic form of H given by (43). The coordinate surfaces (48)–(50) will be referred to as "regularity boundaries" (or simply as "boundaries") as they limit the domain of regularity of M- and W-type solutions. However, further investigation connected with completeness of causal curves is necessary in order to understand the invariant nature of these boundaries. This investigation will be carried on in this section and specifically will be applied to Eqs. (48) and (49) in the following section. For solutions in which conditions (45) are violated, $r=0$ marks an extra regularity boundary, this boundary plus that given by (50) will be studied in Part III. Since for certain equations of state, Θ could be such that Eqs. (48) and/or (49) do not hold (supplementary regularity conditions), it will be assumed in studying either one of these boundaries that an equation of state has been chosen so that comoving observers reach the corresponding boundary.

Since curvature scalars diverge at the regularity boundaries (48) and (49), it appears that these singular boundaries are coordinate representations of scalar curvature singularities. In order to verify this conjecture, the following criterion[29] will be used.

Let $\mathscr{M}$ be the space-time manifold, and let $\gamma(\xi,\beta)$: $\mathbb{R}\to\mathscr{M}$ be a C^1 causal (timelike or null) congruence parametrized by β and by a suitable affine parameter ξ along the curves. Suppose that either one of (48) or (49) is approached along γ. If the conditions (a) curvature scalars diverge and (b) the affine parameter ξ tends to a finite limit hold, then (48) and/or (49) are coordinate representations of a scalar curvature singularity.

Condition (a) above obviously holds for (48) and (49), however (b) must still be verified, and for this purpose, it is convenient to identify γ with causal congruences whose tangent vectors have a simple representation in the coordinates used in (1). The best candidates are the world lines of comoving observers labeled by $r=$ const and with u^{α} as tangent vector, and radial null geodesics whose tangent vector k^{α} satisfies in general the null geodesic equation

$$\frac{dk^{\alpha}}{dv}+\Gamma^{\alpha}_{\beta\sigma}\,k^{\beta}k^{\sigma}=\Omega(v)k^{\alpha}\,, \tag{51a}$$

where

$$\frac{d}{dv}\equiv k^{t}\frac{\partial}{\partial t}+k^{r}\frac{\partial}{\partial r}\,. \tag{51b}$$

The function Ω in (51a) vanishes if the parameter v along the geodesics is an affine parameter.

However, even for these simple causal curves the verification of the incompleteness condition (b) is not trivial. This is so because the world lines of comoving observers are not geodesics and the proper time τ is not an affine parameter along these curves. Also, the components of k^{α} might be complicated functions if the null geodesics are affinely parametrized. In order to deal with this situation, adequate affine parameters will be sought for timelike and null curves separately.

A. Timelike curves

Since integral curves of u^α are not geodesics, the generalized affine parameter[30] (GAP) along u^α can be used to verify condition (b). However, the GAP requires a parallely propagated frame and so it is difficult to compute in the coordinates used in (1). In order to deal with this situation, and having in mind cosmological applications in general, it is desirable to be able to express GAP incompleteness of nongeodesic timelike curves in terms of invariant quantities of physical interest, such as proper time and four-acceleration. Such a GAP incompleteness criterion has not been given before in the literature and is offered in the following.

Proposition (Clarke): Let $\gamma:(0,\tau_1] \to \mathscr{M}$ be a C^1 timelike curve parametrized by its proper time τ, and let $\mathscr{A}(\tau)$ be the magnitude of the four-acceleration $a^\alpha \equiv u^\alpha{}_{;\beta} u^\beta$ at proper time τ:

$$\mathscr{A}(\tau) \equiv [g_{\alpha\beta}(\tau) a^\alpha a^\beta]^{1/2}. \tag{52a}$$

If there exist constants τ_2 and l with $0 \leqslant \tau_2 \leqslant \tau_1$ and $0 < l < 1$, such that

$$\mathscr{A}(\tau)|\tau_1 - \tau| \leqslant l \tag{52b}$$

holds for all $\tau_2 \leqslant \tau \leqslant \tau_1$, then γ is GAP incomplete.

This proposition is rigorously proved in Appendix B for any nongeodesic timelike congruence. However, in order to apply it to M- and W-type solutions, (52b) must be tested for the world lines of comoving observers for which $\mathscr{A}$ is given by (28), while τ follows from (12) which reads explicitly:

$$\tau(t,r) = \int_t \left(\Delta - \frac{2\mu h}{H} + \frac{\epsilon^2 h^2}{H^2} + \frac{LH^2}{h^2}\right)^{1/2} dt. \tag{53}$$

This integral must be computed keeping r constant. Since (28) and the integrand in (53) are complicated functions, and all that is needed is to verify (52b) as comoving observers approach (48) or (49), it will be sufficient to consider the leading terms in the expansion of τ and $\mathscr{A}$ around coordinate values satisfying these constraints. For this purpose, H along the world line of a comoving observer labeled by (say) $r = r_1$ can be expanded around $t = t_1$ such that either one of (48) or (49) holds. This expansion is at first order:

$$H \approx H_1 + \dot{H}_1(t - t_1), \tag{54a}$$

in which

$$\dot{H}_1 = (\Theta_1/3) H_1 [\mathrm{Q}_1]^{1/2} \tag{54b}$$

has been computed from (30). The subindex 1 in Eqs. (54) indicates evaluation at (t_1, r_1).

B. Null geodesics

The affine parameter along null geodesics ϑ is related to the parameter v in (51b) by the differential equation[31]

$$\frac{d\vartheta}{dv} = \exp \int_v \Omega(v)\, dv \tag{55}$$

in which the integral must be evaluated along a null geodesic parametrized as $(t(v), r(v))$, where t and r are related by

$$\left[\frac{dt}{dr}\right]_{\mathrm{null}} = \frac{\pm H}{\mathrm{Q}^{1/2}}. \tag{56}$$

The conformal structure of the regularity boundaries (i.e., whether they are timelike, null, or spacelike surfaces) can be determined by a qualitative examination of Eq. (56) which, for $r = \text{const}$, provides the slopes of the light cones in the (t,r) coordinates as comoving observers approach the regularity boundaries.

The parameter v must be suitably chosen in order to facilitate the evaluation of ϑ in (55). A convenient choice turns out to be that in which the components of a future directed null vector k^α are

$$k^t = \frac{dt}{dv} = \frac{1}{\mathrm{Q}^{1/2}}, \tag{57a}$$

$$k^r = \frac{dr}{dv} = \frac{\pm 1}{H}, \tag{57b}$$

where $\pm$ in (57b) indicates increasing/decreasing r values along the null curve. With this choice, k^α satisfies Eq. (56) with Ω given by

$$\Omega(t,r) = \Theta/3 \pm \mathscr{A}/H, \tag{58}$$

where $\mathscr{A}$ is given by (28), and $t = t(v)$, $r = r(v)$ obtained by integrating Eqs. (57) must be inserted in (58) so that the integral in (55) can be evaluated. As with the timelike case, it is sufficient to consider only leading terms in the expansion of Ω near (48) or (49) in order to verify the convergence of ϑ. For example, H along a null geodesic can be expanded in first order terms as

$$H \approx H_1 + \dot{H}_1(t - t_1) + H_1'(r - r_1), \tag{59a}$$

where H_1 is given by (54b), $H_1' \equiv H'(t_1, r_1)$ can be calculated from Eq. (25), and $(t - t_1)$ and $(r - r_1)$ are related at first order by

$$(t - t_1) \approx (\pm H_1/[\mathrm{Q}_1]^{1/2})(r - r_1). \tag{59b}$$

In expansions such as (54a), (59a), and (59b), one must be careful to have the $\pm$ signs in the coefficients $\dot{H}_1$ and H_1' correctly computed. In the following section, the completeness of timelike and null curves will be studied for each regularity boundary (48) and (49).

X. SINGULARITIES

A. The finite-density singularity

The boundary $\mathrm{Q} = 0$ might occur in solutions of types (i) or (ii). The coordinate values (t,r) corresponding to this boundary are finite in general, and correspond to finite values of Θ, H, and R. In order to verify the convergence of τ, the integral in (53) can be evaluated for $r = r_1$ by expanding H and Q around a value $t = t_1$ with $H_1 \equiv H(t_1, r_1)$ such that $\mathrm{Q}(t_1, r_1) = 0$. For a collapsing configuration ($\Theta < 0$, finite), these expansions are

$$H \approx H_1 + \tfrac{1}{2}\ddot{H}_1(t - t_1)^2, \tag{60a}$$

$$\mathrm{Q}^{1/2} \approx (3/\Theta_1)(\ddot{H}_1/H_1)(t - t_1), \tag{60b}$$

with

$$\ddot{H}_1/H_1 = [\Theta_1/3]^2[\mu h_1 H_1 - \epsilon^2 h_1^2 + L h_1^{-2} H_1^4], \tag{60c}$$

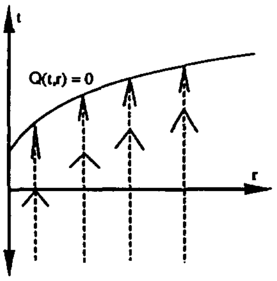


FIG. 2. Null geodesics near the boundary $Q(t,r)=0$. As $Q\to 0$ along the world lines of comoving observers (vertical dotted lines), the slopes of the light cones diverge: $(dt/dr)_{\rm null}\to\pm\infty$, indicating that this boundary is spacelike.

hence, at first order approximation, (53) is

$$\tau-\tau_1\approx[2\Theta_1/3]^{-1}(\ddot{H}_1/H_1)(t-t_1)^2\,, \tag{61}$$

indicating that τ is finite as $Q\to 0$. The magnitude of the four-acceleration, given by (28), is at first order around $t=t_1$:

$$\mathscr{A}\approx -f_1h_1^2/(\Theta_1/3)(t_1-t)\,. \tag{62}$$

From Eqs. (61) and (62), $\tau^{-1}\sim|t-t_1|^{-2}$ while $\mathscr{A}\sim|t-t_1|^{-1}$, thus if τ_1 in (52b) is identified as $\tau_1=\tau(t_1)$, it follows that the GAP incompleteness condition holds for $Q=0$.

From Eq. (56), $(dt/dr)_{\rm null}\to\pm\infty$ as $Q\to 0$, so that the slopes of the light cones along comoving observers become "vertical" in the (t,r) coordinates. Since the "vertical direction" ($r=\text{const}$) in these coordinates corresponds to the timelike world lines of comoving observers and $Q=0$ does not coincide with surfaces Σ_r, then this singular boundary can be characterized as a spacelike singularity. This aspect is illustrated in Fig. 2.

In order to verify the completeness of null geodesics, one must test the convergence of the integral in (55) as $Q\to 0$ along these curves. Using the parametrization given by Eqs. (56), near $Q=0$ one has approximately $dr\approx 0$, $r\approx r_1$ and $dv\approx Q^{1/2}\,dt$, the integral in (55) becomes at first order

$$\int\Omega\,dv\approx\pm f_1h_1^2\int_t^{t_1}\left[-\frac{\mu h_1}{H}+\frac{\epsilon^2h_1^2}{H^2}-\frac{LH^2}{h_1^2}\right]dt\,, \tag{63}$$

where H is given by (60a). Hence, the affine parameter ϑ defined by Eq. (55) is finite as $Q\to 0$, and the coordinate values for which $Q=0$ do represent a scalar curvature singularity at which causal curves terminate. It is easily verified

TABLE I. Domain of regularity and singular boundaries in M- and W-type solutions. This table provides the coordinate representation of the domain of regularity and singular boundaries for those M- and W-type solutions with $|\Pi(T,r)|>0$. This information for solutions presenting the boundary $\Pi=0$ is given in Table II of Part III. Constant parameters and functions h, u, and X for each solution correspond to the forms given in Tables III, VI, and VIII of Part I.

Classification scheme	Type	Domain of regularity	$Q(T,r)=0$	$H(T,r)=0$
NMcV($r3$)(X4,5)	(iv)	$[T+X]^2>0$		$T=\pm X$
NMcV($r2$)(X1,2)	(ii)	$T>\mu u/2$	$T=\mu u/2$	
	(iv)	$T>\lvert\mu\rvert u/2$		$T=\lvert\mu\rvert u/2$
NWy($r2$)(X1,2)	(iv)	$T+X>\pi/\sqrt{\Delta}$		$T=-X+\pi/\sqrt{\Delta}$
	(ii)	$T+X>(2/\sqrt{\lvert\Delta\rvert})\coth^{-1}(\sqrt{3})$	$T=(2/\sqrt{\lvert\Delta\rvert})\coth^{-1}(\sqrt{3})$	
ChMcV($r4$)(X4,5)	(iv)	$T+X>0$		$T=-X$
ChMcV($r3$)(X4,5)	(i)	$\epsilon/\mu<T+X<\sqrt{2}\epsilon/\mu$	$T=\sqrt{2}(\epsilon/\mu)-X$	$T=\epsilon/\mu-X$
	(ii)	$T+X>\sqrt{2}\epsilon/\mu$	$T=\sqrt{2}(\epsilon/\mu)-X$	
ChWy($r3$)(X1,2)	(i)	$\sqrt{12/5}(\epsilon/\mu)<T+X<2(\epsilon/\mu)$	$T=\sqrt{12/5}(\epsilon/\mu)-X$	$T=2(\epsilon/\mu)-X$
ChMcV($r2,r2$)(X1,2)	(i)	$0<T+\mu u(y)<\mu u(y)$	$T=0$	$T=-\mu u(y)$
	(ii)	$T>0$	$T=0$	
	(iv)	$T>\lvert\mu\rvert u(y)$		$T=\lvert\mu\rvert u(y)$
ChWy($r2,r2$)(X1,2)	(i)	$-u^{-1}<T<[u-\Delta/2\epsilon^2]/[(\Delta/2\epsilon^2)u-1]$	$T=[u-\Delta/2\epsilon^2]/[(\Delta/2\epsilon^2)u-1]$	$T=u^{-1}$
ChMcV($r2$)(X1,2)	(i)	$-[\mu+\sqrt{\Delta}\epsilon]u/2<T<-\sqrt{\mu^2-\Delta\epsilon^2}u/2$	$T=-\sqrt{\mu^2-\Delta\epsilon^2}u/2$	$T=-[\mu+\sqrt{\Delta}\epsilon]u/2$
	(ii)	$T>\sqrt{\mu^2-\Delta\epsilon^2}u/2$	$T=\sqrt{\mu^2-\Delta\epsilon^2}u/2$	
ChMcV($r2$)(X3)	(i)	$0<\sin T+u\cos T<[\sqrt{\alpha}-2\mu h]/\alpha\sqrt{\lvert\Delta\rvert}$ $\alpha\equiv\mu^2+\lvert\Delta\rvert\epsilon^2$	$\sin T+u\cos T=[\sqrt{\alpha}-2\mu h]/\alpha\sqrt{\lvert\Delta\rvert}$	$\sin T+u\cos T=-\mu h/\alpha\sqrt{\lvert\Delta\rvert}$
ChWy($r2$)(X1,2)	(i)	$2\sqrt{u_0}<\lvert T+X\rvert<2\sqrt{u_0}\beta$ $\beta\equiv[81\alpha-1]/[81\alpha-4]\quad\alpha\equiv 1+(1+16/2187)^{1/2}$	$T=2\sqrt{u_0}\beta-X$	$T=2\sqrt{u_0}-X$
ChWy($r2$)(X1,2)	(i)	$0<(1/\sqrt{2})\cos V<1+\cos V$ $0<V=\sqrt{\Delta/2}[T+X]<\pi$	$V=\cos^{-1}[\sqrt{2}/(1-\sqrt{2})]$	$V=\pi/2$
ChWy($r2$)$I\delta_+$(X1,2)	(i)	$[\alpha_1+\sqrt{\alpha_1^2+4}]u/2<T<\gamma_+u/2$ $\gamma_\pm\equiv\alpha_1W_\pm-\alpha^2A+[(\alpha_1^2-4)W_\pm^2$ $+2A(4-\alpha_1\alpha_2)W_\pm+(\alpha_2^2-4)A^2]^{1/2}$	$T=\gamma_+u/2$	$T=[\alpha_1+\sqrt{\alpha_1^2+4}]u/2$

that the same results hold for neutral solutions ($\epsilon = 0$), provided that $Q = 0$ occurs.

The singularity marked by $Q = 0$ is essentially different from a FRW big bang, as it is not characterized by the vanishing of proper volumes of local fluid elements (Θ, H, and $R = fH$ are finite), also matter-energy and charge densities ρ and q remain finite, and so the dominant and strong energy conditions are violated, though not necessarily the weak energy condition if $p \to +\infty$. This singularity, which can be associated with the "blowing up" of the terms $d\Theta/d\tau$ and $a^{\alpha}{}_{;\alpha}$ in the Raychaudhurri equation, occurs if the parameters ϵ, μ, Δ, and L are chosen so that Q [and so, Q in I(24b)] has real positive roots [solutions of types (i) and (ii) in Table I]. It has no parallel in more familiar solutions such as FRW or Tolman–Bondi solutions. However, some Bianchi models examined by Collins and Ellis[4] do exhibit similar singularities, which these authors have denoted as "finite-density singularities." This nomenclature will be adopted in this paper, and so unless stated otherwise, any mention of a finite-density singularity (FD singularity) in the following sections and in Part III will be understood to refer to the singularity marked by Eq. (49), or of a similar kind (see Secs. VII and IX of Part III).

B. An "asymptotically delayed" big bang

The boundary $H(t,r) = 0$ marked by Eq. (48) seems to be the coordinate representation of a big-bang singularity present in FRW solutions. This is so because the Hubble scale factor and the proper radii of comoving shells of fluid (given by H and $R = fH$) vanish, and matter-energy and charge densities, pressure and pressure gradient diverge (in general) as these coordinate values are approached. However, such a resemblance must be reexamined closely. In FRW solutions, one has $H = H(t)$, and the big bang is a spacelike singularity labeled by a (singular) Σ_t surface, say $t = t_0$, at which $H(t_0) = R(t_0) = 0$ as $|\Theta(t_0)| \to \infty$ (see Fig. 4). This situation will be referred to henceforth, throughout this paper and Part III, as the standard big bang. Thus, once the equation of state is set in such a way that (dominant or strong) energy conditions hold, following singularity theorems,[30] a standard big bang develops in FRW solutions. As there are no "curvature" terms, such as $\mathscr{A}$, and ${}^{(3)}\mathscr{R} = k/H^2$ also diverges at $t = t_0$, there are no other singularities in these solutions.

However, the kinematical restriction of zero shear requires the expansion Θ to be constant along the surfaces Σ_t orthogonal to u^{α}, that is, $\Theta = \Theta(t)$ in SSSF solutions. Thus if $H = H(t,r)$, it is not possible (in general, see Secs. VII and IX of Part III for counterexamples) to have $\Theta \to \pm\infty$ coinciding with the boundary $H(t,r) = 0$. It is even possible to choose an equation of state in such a way that Θ remains finite throughout this boundary, as indeed it happens in the Wyman solution (see Appendix A in this paper and Appendix A in Part III). The completeness of curves reaching the boundary $H = 0$ with Θ everywhere finite will be considered first, while in the case when Θ diverges along a value (t_0, r_0) satisfying $H(t_0, r_0) = 0$ will be discussed later.

Besides having Θ finite, another important difference with a standard big bang is that $H(t,r) = 0$ is not spacelike.

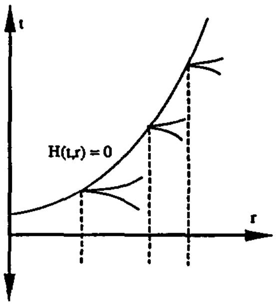


FIG. 3. Null geodesics near the boundary $H(t,r) = 0$. As $H \to 0$ along the world lines of comoving observers (vertical dotted lines), the slopes of the light cones tend to zero: $(dt/dr)_{\text{null}} \to 0$, indicating that this boundary is timelike.

From Eq. (56), $(dt/dr)_{\text{null}} \to 0$ as $H \to 0$, while for $H > 0$, $(dt/dr)_{\text{null}}$ is either negative or positive. That is, the slopes of the light cones in a (t,r) coordinate diagram become horizontal (i.e., the cones "open up") as $H \to 0$. Since $H = 0$ does not coincide with a surface Σ_t, then $H = 0$ will be intersected by past-directed and future-directed null geodesics, and thus it is a timelike boundary. This situation is illustrated in Fig. 3.

In order to evaluate τ in (53), the expansion (54a) of H for fixed $r = r_1$ around the coordinate value $t = t_1$ corresponding to $H = 0$ is, at first-order approximation,

$$H = (\epsilon h_1 \Theta_1/3)(t - t_1), \tag{64}$$

where Θ_1 is finite and the charge density has been assumed to be positive (i.e., $\epsilon > 0$). Near $H = 0$, the integral in (53) is approximately

$$\tau - \tau_1 \approx \int_t^{t_1} \frac{\epsilon h_1}{H} dt \approx \frac{3}{\Theta_1} \int_t^{t_1} \frac{dt}{t - t_1}, \tag{65}$$

which clearly diverges logarithmically. For neutral solutions which are not conformally flat ($\epsilon = 0$, but $\mu < 0$), one obtains the same result: τ diverges logarithmically as H vanishes if Θ is finite. Therefore, since the GAP length along nongeodesic timelike curves is longer than the proper time length, the congruence of world lines of comoving observers is complete at the regularity boundary (48). Also, timelike geodesic congruences reaching this boundary will be complete, since their affine parameter lengths are necessarily longer than proper time length τ along nongeodesic curves.

In order to test the completeness of null geodesics, consider a collapsing configuration ($\Theta < 0$). According to Eqs. (57), near $H(t_1, r_1) = 0$ one has $dt \approx 0$, $t \approx t_1$ and $dv \approx \pm H\, dr$. Hence, the integral in (55) becomes approximately

$$\int_v \Omega\, dv \approx \int_r^{r_1} \left[\frac{\Theta_1 H}{3} + \frac{\epsilon f h^3}{H}\right]_{t = t_1} dr, \tag{66a}$$

in which Θ_1 is finite and the $\pm$ signs in $\mathscr{A}$ have canceled with those of dr. From (58), H is approximately

$$H \approx \pm \epsilon f_1 h_1^3 (r - r_1), \tag{66b}$$

with $\pm$ indicating that H increases or decreases for increasing r. Since $\mathscr{A} \to +\infty$ as $H \to 0$, Eqs. (66) imply that the affine parameter along null geodesics, computed from (55), diverges as these curves reach $H = 0$ at coordinate values with finite Θ.

The fact that τ and ϑ diverge as $H \to 0$ at all coordinate values at which Θ is finite implies that $H = 0$ in these values

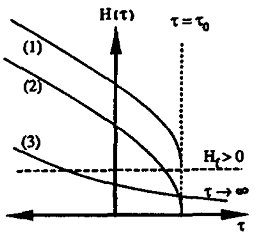


FIG. 4. Plot of H vs proper time τ. The curve (2) depicts $H(\tau) \to 0$ as $\tau \to \tau_0$ and $\Theta \to -\infty$, corresponding to a standard big-bang collapse. The curve (1) indicates the behavior of $H(\tau)$ corresponding to a finite-volume (FV) singularity, which is analogous to that of curve (2), except that $H(\tau_0) = H_f > 0$. The curve (3) illustrates the behavior of $H(\tau)$ along the world line of a comoving observer heading towards the asymptotically delayed (AD) big bang: $H(\tau) \to 0$ as $\tau \to \infty$ with finite Θ.

does not mark a standard FRW big-bang singularity, but rather space-time points in the infinite past or future of the comoving observers. Since the coordinate values satisfying $H = 0$ are finite, this surface simply denotes a sort of coordinate compactification of a singular timelike infinity. However, even if technically speaking $H = 0$ is not a scalar curvature singularity (it does not satisfy the incompleteness criteria), curvature scalars diverge and the proper radii of fluid shells vanish as $H \to 0$. Therefore, apart from incompleteness of causal curves, all other features of a big-bang singularity are present in $H = 0$. And so, it seems, the best characterization of $H = 0$ is that of coordinate values marking a sort of "asymptotically delayed big bang" (AD big bang) which is reached asymptotically in the infinite past or future of the comoving observers (see Fig. 4).

This strange phenomenon follows directly from the fact that Θ is finite along $H = 0$. From Eq. (7), the expansion is defined as $\Theta = d(\ln H^3)/d\tau$, and so for a small increment in proper time, $\tau_f - \tau_i$, along the world line of a comoving observer at $r = r_0$, one has approximately (at first order)

$$\Theta \approx 3 \ln(H_f/H_i)/(\tau_f - \tau_i)\,, \tag{67}$$

where $H_f \equiv H(\tau_f, r_0)$ and $H_i \equiv H(\tau_i, r_0)$. Hence, the standard big bang is characterized by the expansion Θ diverging as the proper volume of fluid shells vanishes [$H_f = 0$ in Eq. (67)] for a finite lapse $|\tau_f - \tau_i| \to 0$. But the combination of finite Θ with vanishing proper volume as $H \to 0$ [$H_f = 0$ in Eq. (67)] can only be reconciled from Eq. (67) if the lapse of τ diverges. The finiteness of the coordinates denoting $H = 0$ merely obscures this situation. The behavior of H as a function of τ near the standard and AD big bangs is illustrated by Fig. 4.

Another important fact is that, regardless of the choice of formal equation of state, the dominant and strong (but not necessarily the weak) energy conditions are violated as $H \to 0$ for $0 \leqslant r < r_0$. This is so because the terms ${}^{(3)}\mathcal{R}$ and $a^{\alpha}{}_{;\alpha}$, as given by Eqs. (16), (20), (21), and (29), do not diverge at the same rate, and so, as $H \to 0$ along the world lines of comoving observers, $\mathcal{A} \to +\infty$ and the state variables p and ρ tend to the following asymptotic limits: $8\pi\rho \to \Theta^2/3 + \mathcal{O}(H^{-3})$ and $8\pi p \to -\Theta^2/3 + \mathcal{O}(H^{-5})$, where $\mathcal{O}(H^{-n})$ means terms of order H^{-n}. Since Θ is finite as $H \to 0$, these asymptotic limits suggest that $|p| \gg \rho > 0$. For neutral solutions which are not conformally flat ($\epsilon = 0$, $\mu \neq 0$), the same situation arises, however, in this case, $\rho \sim H^{-5/2}$ and $p \sim H^{-4}$ as $H \to 0$. However, there might be particular solutions in which the ratio $|p|/\rho$ behaves differently[12] as $H \to 0$ (see Appendix C).

C. "Localized" and "finite-volume" singularities

As mentioned before, in most cases in which $H = H(t,r)$, the combination $|\Theta| \to \infty$ with $H \to 0$ can only happen (if it happens) for a class of comoving observers labeled by $r = r_0$ as they reach the coordinate value $t = t_0$ such that $H(r_0, t_0) = 0$. A sufficient condition allowing for Θ to diverge along $r = r_0$ is a choice of localized equation of state so that the dominant or strong energy conditions hold

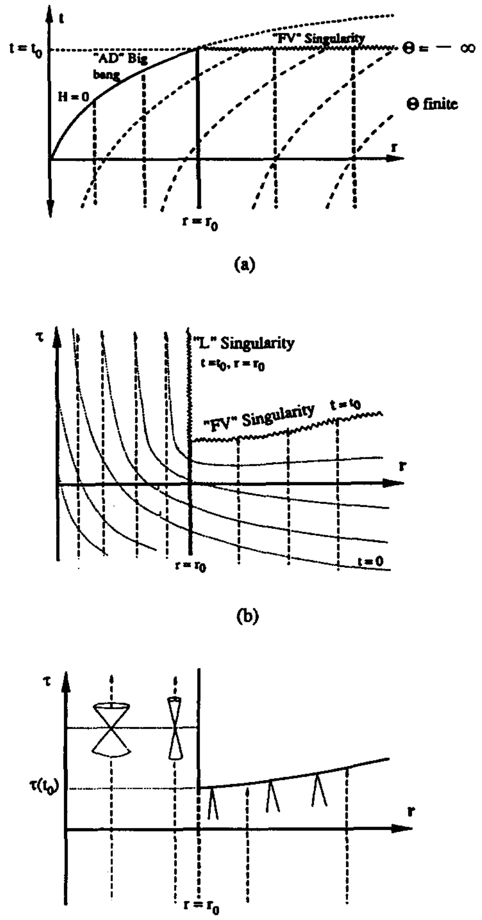


FIG. 5. The boundary $H(t,r) = 0$ when Θ diverges at $t = t_0$. If $\Theta \to -\infty$ as $t \to t_0$, as shown in (a), comoving observers labeled with $0 \leqslant r < r_0$ do not reach $t = t_0$, evolving towards the AD big bang in their infinite future. Comoving observers labeled with $r > r_0$ reach the FV singularity at $t = t_0$, while those observers labeled by $r = r_0$ collapse into the L singularity which appears in (a) as a "point" with coordinates (t_0, r_0). The situation becomes clear in (b) using the coordinate representation (τ, r). This figure shows the surfaces Σ_t (t near t_0) bending towards the AD big bang as $\tau \to \infty$. Observers with $r \neq r_0$, either hit the FV singularity or evolve towards their infinite future, while the "point" (t_0, r_0) unravels into the "line" $\tau > \tau(t_0)$, $r = r_0$, which is the L singularity avoided by observers with $r \neq r_0$. In (c) are displayed null geodesics near the FV and L singularities, showing that the latter are spacelike and null, respectively. Notice how the light cones become "vertical" as $r \to r_0$ along $\tau = \text{const} > \tau(t_0)$.

at this Σ_r surface (see Sec. VII). With such a choice, as illustrated by Fig. 5 for a collapsing configuration [$\Theta<0$, $\Theta(t_0)=-\infty$], the boundary $H=0$ can be divided in three regions, depending on whether Θ is finite and/or H vanishes. From Fig. 5(a), comoving observers labeled by $0\leqslant r<r_0$ reach $H=0$ at coordinate time values $t<t_0$ with Θ negative and finite. Hence, these observers have complete world lines and $H=0$ is an AD big bang in their infinite future ($\tau\to\infty$). Comoving observers with $r>r_0$ do not reach $H=0$, since Θ diverges at $t=t_0$ with $H(t_0,r)>0$ [and $R(t_0,r)>0$], and from the field equations (16) and (20), ρ and p also diverge. This situation, as shown in Fig. 4, corresponds to the theoretical possibility in Eq. (67) of allowing Θ to diverge with $H_f/H_i>0$ and $|\tau_f-\tau_i|\to 0$ as $t\to t_0$, identifying $\tau_f=\tau(t_0)$ and $\tau_i=\tau(t)$. However, for comoving observers at r_0, Θ associated with the ratio of Eq. (67) has the standard FRW behavior mentioned earlier: $H_f/H_i\to 0$ together with $|\tau_f-\tau_i|\to 0$. The full picture for comoving observers labeled by $r\geqslant r_0$, as they head towards $t=t_0$ in Fig. 5(a), is illustrated by Figs. 5(b) and 5(c). Completeness of causal curves at this Σ_t surface is discussed below.

If Θ diverges at $t=t_0$ for $r>r_0$ in Fig. 5(a), even if H and R are not zero, the state variables p and ρ (but not q, since $H>0$) diverge as comoving observers reach this singular Σ_t surface. From Eqs. (7) and (67) and Fig. 4, this behavior of Θ near $t=t_0$ requires that comoving observers reach $t=t_0$ in a finite proper time lapse. This also follows from Eq. (53) using the coordinate choice of Eq. (33), so that $\tau=\int(-g_{tt})^{1/2}\,dT\to 0$ as $\Theta\to\infty$. Since the magnitude of the four-acceleration $\mathscr{A}$ is bounded (Q, H, and/or R do not vanish at $t=t_0$), then the incompleteness criterion given by (52b) is satisfied, and so the world lines of comoving observers terminate at $t=t_0$. For a collapsing configuration (the case of an expanding one is similar) in which $\Theta\to-\infty$ as $t\to t_0$, the integral in (55) with Ω given by (58) converges, and so (future directed) null geodesics also terminate at this Σ_t surface, which marks then a scalar curvature singularity. Since the proper volume of local fluid elements ($\sim H^3$) does not vanish, this singularity, as a sort of a collapse to a nonzero volume, will be referred to as a "finite-volume singularity" (FV singularity).

Since both H and Q are nonzero at $t=t_0$, from Eq. (56), $|(dt/dr)_{\rm null}|>0$ as comoving observers reach this singular Σ_t surface, and so the FV singularity is a spacelike boundary [see Figs. 5(a) and 5(c)]. From the field equations of Secs. III and IV, and due to the fact that curvature terms, such as $\mathscr{A}$ and ${}^{(3)}\mathscr{R}$, remain finite as Θ diverges, the state variables approach the following limiting values:

$$8\pi\rho\to\Theta^2/3\,,\quad 8\pi p\to\frac{-\Theta^2}{3}-2\frac{d}{d\tau}\frac{\Theta}{3}\,.$$

These values, unlike those near the AD big bang or the FD singularity, strongly depend on the choice of equation of state (choice of Θ). Since $(d/d\tau)(\Theta/3)\sim d^2H/d\tau^2$, from Fig. 4, this derivative is negative near the FV singularity, and so $p+\rho>0$ at this limit.

From Eq. (67) and Fig. 4, H vanishes in a finite proper time lapse for comoving observers along $r=r_0$. This also follows from the fact that the strong or dominant energy conditions hold at this surface Σ_r, and as can be shown from the Bianchi identities and the field equations restricted to $r=r_0$, Θ near $H_0=0$ must be constrained by $(H_0)^{-3/2}<\Theta<(H_0)^{-3}$, and thus Θ overtakes the term H_0^{-1} in (65), and the proper time near $H_0=0$ is constrained by

$$\int_t^{t_0}|t-t_0|^{1/2}\,dt\geqslant|\tau-\tau_1|\geqslant\int_t^{t_0}|t-t_0|^2\,dt\,, \tag{68}$$

indicating that τ converges as $H\to 0$ for comoving observers along $r=r_0$. Since Eq. (68) implies that $|t-t_0|^{-3/2}\leqslant|\tau-\tau_0|^{-1}\leqslant|t-t_0|^{-3}$ if the strong energy condition holds along r_0, and since $\mathscr{A}_0\sim H_0^{-1}\sim|t-t_0|^{-1}$ holds for whatever equation of state, the GAP incompleteness condition (62b) holds and the world lines of comoving observers labeled by r_0 are incomplete as $H_0\to 0$.

Since the proper radius of the two-sphere of symmetry corresponding to r_0 vanishes in τ finite, a sort of "localized singularity" (L singularity) must be produced as the end product of the collapse of this two-sphere. However, as observers with $r\neq r_0$ avoid this singularity, evolving towards the AD big bang in their infinite future or hitting the FV singularity, it is difficult to appreciate the properties of the L singularity in the (t,r) coordinates currently used. This situation can be appreciated by comparing Fig. 5(a) with Figs. 5(b) and 5(c).

The L singularity is represented in the (t,r) coordinates as a "point" with coordinates (t_0,r_0) such that $H_0=0$ [see Fig. 5(a)]. If proper time $\tau(t,r)$ defined by Eq. (53) is introduced as a new time coordinate, the "point" marked by (t_0,r_0) unravels in a (τ,r) coordinate diagram into a "line" $(r=r_0,\ \tau>\tau_0)$, where τ_0 is the (finite) proper time value at which $H_0=0$ [see Fig. 5(b)]. Since this locus is marked by the value (t_0,r_0), it can be argued intuitively that it must correspond to a null surface which is the limit of timelike congruences of comoving observers labeled by values of r which are arbitrarily close to r_0 and spacelike surfaces Σ_t labeled by values of t arbitrarily close to t_0. This situation is shown in Fig. 5(b), however, it also follows by examining null geodesics near $r=r_0$ in (τ,r) coordinates. The metric (1) becomes in this representation:

$$\begin{aligned}ds^2=&-d\tau^2+2\Lambda\,d\tau\,dr+[\Lambda^2-H^2]dr^2\\&+R^2[d\theta^2+\sin^2\theta\,d\phi^2]\,,\end{aligned} \tag{69a}$$

where

$$\Lambda\equiv\tau'=fh^2\int\left[-\frac{\mu h}{H}+\frac{\epsilon^2h^2}{H^2}-\frac{LH^2}{h^2}\right]_{r={\rm const}}dt \tag{69b}$$

must be evaluated as a function of τ and r. The null geodesic equation (56) becomes in these coordinates:

$$\left[\frac{d\tau}{dr}\right]_{\rm null}=\Lambda\pm H\,. \tag{70}$$

Since $H=0$ at r_0, the behavior of null geodesics near r_0 depends exclusively on the behavior of the function Λ as r_0 is approached along a surface $\tau=\tau_1>\tau_0$. A tangent vector

$\partial/\partial n$ along such a surface is given by

$$\frac{\partial}{\partial n}\equiv\frac{1}{Q^{1/2}}\frac{\partial}{\partial t}+\frac{\Lambda}{H^2}\frac{\partial}{\partial r}, \tag{71a}$$

so that the rate of change of Λ along $\partial/\partial n$ is

$$\frac{\partial\Lambda}{\partial n}=\mathscr{A}+\frac{2(\Lambda^2)'}{H^2}, \tag{71b}$$

which diverges as $H\to 0$. Since H only vanishes along τ_1 finite if $r\to r_0$, Λ diverges at $r=r_0$. Therefore, from Eq. (70), $(d\tau/dr)_{\rm null}\to\infty$ and the slopes of the null cones become vertical coinciding with $r=r_0$. This situation is illustrated in Fig. 5(c) and will be further discussed in Sec. XI.

For conformally flat solutions (see Sec. IX of Part III), the integral (57) has a finite limit as H vanishes for all comoving observers. In this case, $H=0$ does represent a singularity analogous to the FRW big bang. In Table I the coordinate representation of the boundaries studied in this section is offered for the M- and W-type solutions classified in Tables III and VI of Part I.

XI. GRAVITATIONAL COLLAPSE

Charged Kustaanheimo–Qvist solutions can be continuously matched to the Reissner–Nordstrøm vacuum solution, whose metric is

$$ds^2=-\Phi\,d\ell^2+\Phi^{-1}d\chi^2+\chi^2(d\theta^2+\sin^2\theta\,d\phi^2), \tag{72}$$

with

$$\Phi\equiv 1-2m/\chi+e^2/\chi^2,$$

or to Schwarzschild solution ($e=0$) if the matching ChKQ solution is neutral. Darmois' matching conditions[7,32] are usually applied to characterize invariantly such a matching, which for the case of ChKQ solutions, yields a hybrid space-time consisting of an interior fluid region [assuming that conditions (43) hold at $r=0$] with a metric given by (1) and a vacuum "exterior" whose metric is (72). In other words, a bounded fluid sphere evolving in an asymptotically flat Reissner–Nordstrøm (or Schwarzschild) field.

Since the matching interface is an arbitrary Σ_r surface labeled with $r=r_0>0$, it is useful to choose the time coordinate t in the fluid region as in (32) with r_0 marking the matching interface, i.e., the proper time of observers comoving with the surface of the sphere. However, other choices, such as (31) or (33), can also be used. In either case, it is well known that for spherically symmetric perfect fluid configurations Darmois' matching conditions imply[7,17,18,33]:

$$[\chi(\ell)]_{r=r_0}=R(t,r_0)=R_0, \tag{73a}$$

$$p(t,r_0)=p_0=0, \tag{73b}$$

where subindex 0 indicates evaluation at $r=r_0$. Equation (73a) follows directly from the fact that both matching space-times are spherically symmetric; hence they share the geometric interpretation of $R=(g_{\theta\theta})^{1/2}$ discussed in Sec. II. In fact, the metric coefficient R can be identified throughout the fluid region as a natural continuation of the curvature coordinate χ appearing in (72). On the other hand, (73b) is a particular case of a localized equation of state of the form $p_0=p_0(\rho_0)$ discussed in Sec. VII. Thus, integrating (41) with $p_0=0$ yields ρ_0 proportional to R_0^{-3}, so that in agreement with Darmois' boundary conditions, Eq. (42b) becomes

$$\begin{aligned}[\Theta R_0/3]^2&=(\dot R_0)^2\\&=-\left[1-\frac{2m}{R_0}+\frac{e^2}{R_0^2}\right]+\left[1-ky_0+f_0^2\left(\frac{h_y}{h}\right)_0\right.\\&\qquad\left.-(f_0h_0)^2[Q_0]^{1/2}\right]^2,\end{aligned} \tag{74a}$$

with

$$e=E_0=\pm\epsilon(f_0h_0)^3, \tag{74b}$$

$$m=\tfrac{4}{3}\pi\rho_0R_0^3+J_0, \tag{74c}$$

where $(h_y/h)_0$ is h_y/h evaluated at $r=r_0$. Equation (74a) is the equation of motion of the class of observers comoving with the matching interface, i.e., the "surface" of the fluid sphere. However, the time evolution of this interface, as detected by distant static observers in the Reissner–Nordstrøm (or Schwarzschild) exterior, is given by expressing R_0 in terms of the ℓ coordinate in (72) evaluated at r_0. The relation between the two time coordinates at the interface, which follows immediately from $u_\alpha u^\alpha=-1$, is given by

$$\begin{aligned}\dot\ell_0&=\frac{\partial\ell(t,r_0)}{\partial t}\\&=\frac{1-ky_0+f_0^2(h_y/h)_0-(f_0h_0)^2[Q_0]^{1/2}}{1-2m/R_0+e^2/R_0^2},\end{aligned} \tag{75}$$

so that $\ell_0(R_0)$ follows by integrating $(d\ell_0/dR_0)=\dot\ell_0/\dot R_0$ obtained from (75) and using (74a).

For M- and W-type solutions, the boundary condition (73b) fully determines the field equations as the only t parameter is found through the integration of (74a). Once (74a) is integrated, usually as a quadrature $t=t(R_0)$, the motion of "internal" fluid layers with $0\leqslant r\leqslant r_0$ can be found from expressions like I(30) relating H with H_0 (or R with R_0, see Appendix C). If the t parameter $T(t)$ is used instead of R_0, then if $t(R_0)$ is known, $t(T)$ and $t(R)$ follow from the expressions relating H with T. (See the final paragraph of Appendix C). As mentioned in Sec. VII, the resulting localized type of formal equation of state can be expressed in the form (35) from (16b) and (74a) after eliminating t in terms of H (and thus, N) and prescribing $S(r)$. However, there is no indication that such equations of state will have any physical meaning, other than being mathematically simple, and allowing one to use these fluid solutions as models of collapsing isolated spheres. Most papers dealing with ChKQ solutions found in the literature [category (b) in the Introduction of Part I] study them within this approach. Gravitational collapse of solutions with $L=L(t)$ and solutions which do not satisfy conditions (45) will be discussed in Part III.

The relevant question concerning the gravitational collapse of bounded ChKQ spheres is whether or not a black hole develops. From the discussion in the previous sections, collapsing ($\Theta<0$) bounded spheres corresponding to M- and W-type solutions can have either one of the following two evolutions.

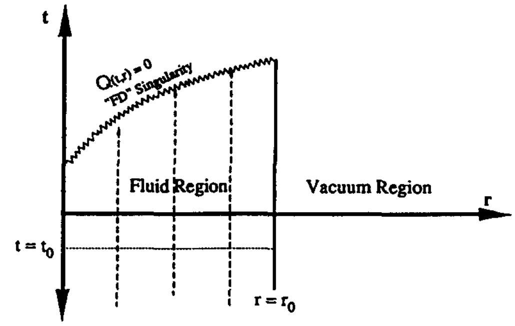


FIG. 6. Collapsing sphere of type (ii). The sphere matches with Schwarzschild or Riessner–Nordström space-times (vacuum region) at $r = r_0$. The evolution of all fluid layers (vertical dotted lines) terminates at the FD singularity. Notice that some surfaces Σ_t, such as $t = t_0$, avoid this singularity.

(1) Fluid layers reach the spacelike FD singularity given by $Q = 0$ [Eq. (49)] at $R > 0$ (for $r > 0$) with Θ negative but finite. This situation arises for solutions belonging to type (ii) in Table I.

(2) Observers comoving along the "surface" layer $r = r_0$, collapse into the null L singularity discussed in the last section. For this class of observers, $\Theta \rightarrow -\infty$ and $R_0 \rightarrow 0$ occur in finite proper time. Interior fluid layers ($0 \leqslant r < r_0$) evolve towards the AD big bang at $R = 0$ [Eq. (48)] in infinite proper time (an asymptotically delayed collapse) with Θ negative but finite. This situation arises in solutions belonging to type (iv) of Table I.

(3) In some solutions of type (i), the fluid sphere evolves between the FD and the FV singularities. At the latter singularity, $\Theta \rightarrow -\infty$, and the proper volume of the "surface" of the sphere vanishes ($R_0 \rightarrow 0$) while interior layers ($0 \leqslant r < r_0$) terminate their evolution with $R > 0$.

Hence, solutions of type (ii) do not form black holes since the evolution of all fluid layers terminates at $Q = 0$ occurring at $R > 0$. This becomes more evident if the constraint $Q = 0$ is expressed as $R(t,r) = fh/A$, where A is a positive root of Q (see Fig. 6). The existence of the FD singularity $Q = 0$ was remarked by Glass and Mashhoon[6] and by Mashhoon and Partovi[7] in their study of collapsing spheres associated with the solutions NMcV($r2$)(X1,2) and ChMcV($r2$)(X1,2), respectively. However, their characterization of the nonsingular fluid region as $r > r_s$, where r is their radial coordinate [see Eq. (74) of Ref. 7], is misleading. As shown in Fig. 6, there are surfaces Σ_t in which the fluid is perfectly regular for all the range of radial comoving coordinate $0 \leqslant r < r_0$. See Appendix C.

Solutions of types (i) and (iv), on the other hand, do form black holes since the localized equation of state (74b) implies that the boundary layer r_0 does collapse ($R_0 \rightarrow 0$) in a finite proper time [$\tau \sim |t - t_0|^{3/2}$, from Eq. (41)]. However, these black holes, which as far as I am aware have not been reported previously, are characterized by the peculiarity that interior layers remain with finite proper volume ($H > 0$). Though, the case (3) above has little interest in the study of gravitational collapse because the collapsing sphere must emerge from the FD singularity [see Fig. (7)], and so it will not be discussed any further. Regarding the case (2), while observers in the vacuum exterior region close to the surface of the sphere would detect how this sphere vanishes into a singularity, just as in the case of the collapse associated

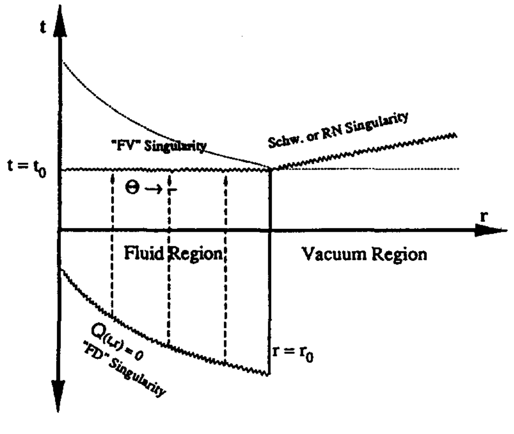


FIG. 7. Collapsing sphere of type (i). As in Fig. 6, the sphere is matched to Riessner–Nordström or Schwarzschild space-time (vacuum region) at $r = r_0$. Fluid layers emerge from the FD singularity and collapse into the FV singularity at $t = t_0$. The surface of the sphere $r = r_0$ collapses into a standard big bang [curve (2) of Fig. 4]. However, for observers with $r < r_0$, R and H behave near $t = t_0$ as curve (1) of Fig. 4.

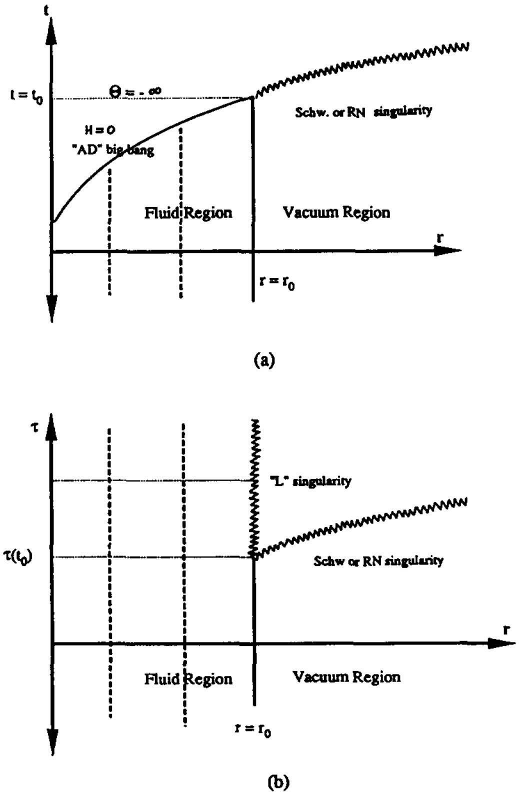


FIG. 8. Collapsing sphere of type (iv). The sphere matches to Schwarzschild or Riessner–Nordström space-times (vacuum region) at $r = r_0$. The surface of the sphere $r = r_0$ collapses into the L singularity shown in (a) as a point and in (b) as a line, in a smaller way as in Figs. 5(a) and 5(b). Dotted curves in (b) are the surfaces Σ_t, which "bend" towards infinite values of τ. Comoving observers in the interior avoid the L singularity, evolving towards the AD big bang in their infinite future, thus surviving the collapse of the surface $r = r_0$.

with a standard big bang, however, observers comoving along internal layers avoid this null singularity (the L singularity) and only collapse asymptotically in the AD big bang in their infinite future [see Figs. 8(a) and 8(b)]. This situation arises because $R_0 = 0$ and $R = 0$ do not imply each other [see Eq. (C4) of Appendix C], and also because $R_0 \to 0$ in (74a) occurs for finite τ if the range of R_0 values for which $|Q_0| > 0$ includes $R_0 = 0$ [i.e., types (i) and (iv) of Table I].

It is difficult to visualize a situation in which the surface of a sphere collapses while interior layers remain with finite proper volume. Such an evolution, which seems to contradict intuition, can be understood by comparing it to the evolution of a 2-D circular disc as seen from 3-D space. This is illustrated in Fig. 9, where an initially flat 2-D circular disc [Fig. 9(a)] is deformed in such a way that its rim is turned into a sort of very narrow "bottleneck" [almost a point with zero area and circumference, see Fig. 9(d)], while its interior circular layers remain quite bent but with finite nonzero area. A succession of 3-D spacelike slices in the interior of the sphere, as seen from space-time [see Fig. (10)], is analogous to the succession of circular disks in this example, as the boundary layer collapses ($R_0 \to 0$) it turns itself into a 3-D "bottleneck" which "closes" as $R_0 = 0$, forming a null singularity in case (2) (the L singularity) and "pinching off" the interior of the sphere from the asymptotically flat region [see Fig. 10(d)]. Once this has happened, the world lines of comoving observers evolve towards the AD big bang avoiding the null L singularity [see Fig. 8(b)]. In case (3), as the bottleneck in Fig. 10(d) closes, the whole sphere (i.e., internal layers) becomes singular (the FV singularity).

The evolution of type (iv) fluid spheres is illustrated in the Penrose diagrams of Figs. 11, 12, 13(b), and 13(c). These diagrams are not rigorous (for an indication of how they could be made rigorous, see Sussman[33]), and as a result of the form of the analytical extension of Reissner–Nordstrøm space-time, it is interesting to examine (for charged spheres) whether the surface layer $r = r_0$ in these spheres collapses into the left- or right-hand sides of the Reissner–Nordstrøm singularity [see Figs. 12, 13(b), and 13(c)]. This question has been studied by de Felice and Maeda,[34] and their methods could be applied to ChKQ solutions through a qualitative analysis of Eqs. (74a) and (75).

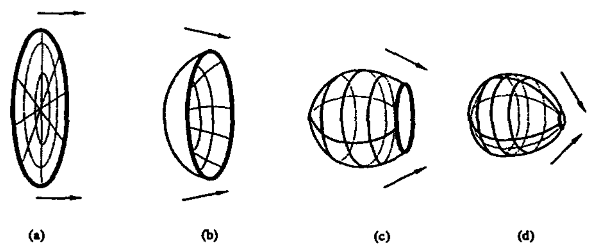


FIG. 9. "Boundary" and "interior" of a disk. This figure is the 2-D analog of the situation described in Sec. IX, in which the surface layer of the sphere collapses while interior layers remain with nonzero volume. In (a), the "boundary" of the disk (thick circle) has a radius greater than any of the interior layers (dotted circles). As the disk is progressively deformed from (b) to (d), the boundary becomes a sort of bottleneck as some of the interior layers will have a greater radius than the boundary. The latter might even become a point if the deformation ends up closing the bottleneck. A similar situation (in three dimensions) occurs in the collapse of type (iv) spheres as seen from 4-D space-time. See Fig. 10.

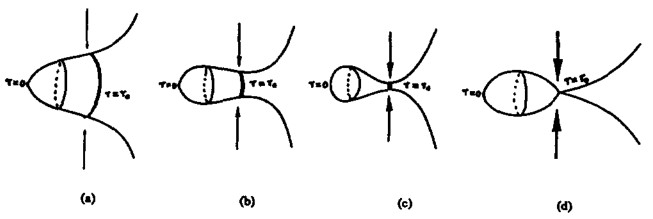


FIG. 10. Spacelike slices near the L singularity. (a)–(d) show a sequence of spacelike hypersurfaces embedded in $\mathbf{R}^4$ (see Sec. II of Part III for a discussion on how these embeddings are defined) as the surface layer ($r = r_0$) approaches the L singularity $t \to t_0$. Such a sequence of spacelike slices, which could be generated by, say, spacelike geodesics is shown in Fig. 11. These hypersurfaces, which experience a deformation analogous to the disk of Fig. 9, are not surfaces Σ_t because the latter (for t near t_0) head towards the AD big bang, and so would not reach the center $r = 0$.

Another important aspect concerning the evolution of fluid spheres in the Reissner–Nordstrøm or Schwarzschild backgrounds is the formation of trapped surfaces within these spheres. As the apparent horizon contains these surfaces, it is necessary to find if such an horizon covers the FD singularity at $Q = 0$ in type (ii) solutions and the L or FV singularities for type (iv) solutions. The expression defining the apparent horizon for spherically symmetric solutions follows from demanding that R does not increase nor decrease with respect to the affine parameter along null geodesics,[7,10,18,35] that is, from the condition $(dR/d\vartheta) = 0$. But since $d/d\vartheta = [\exp \int \Omega\, dv]^{-1}(d/dv)$, with d/dv given by (55b), this condition is for ChKQ solutions:

$$\left[\frac{\Theta R}{3}\right]^2 = \left[\frac{fR'}{R}\right]^2 \Leftrightarrow 1 - \frac{2M}{R} + \frac{E^2}{R^2} = 0\,, \qquad (76a)$$

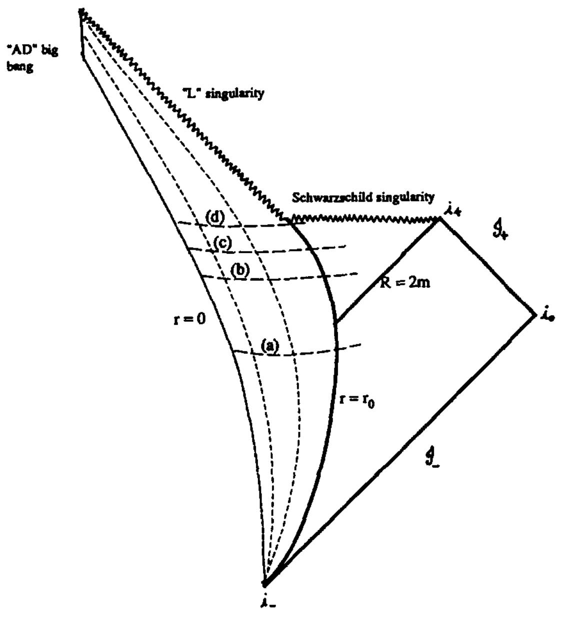


FIG. 11. Qualitative Penrose diagram of a neutral collapsing sphere of type (iv). The world lines of comoving observers (dotted curves) are complete, evolving towards the AD big bang in their infinite future. The surface of the sphere ($r = r_0$) collapses into the null L singularity which joins with the Schwarzschild spacelike singularity. A sequence of four spacelike slices, whose embedding in $\mathbf{R}^4$ is shown in Fig. 10, are displayed as horizontal dotted curves marked (a) to (d).

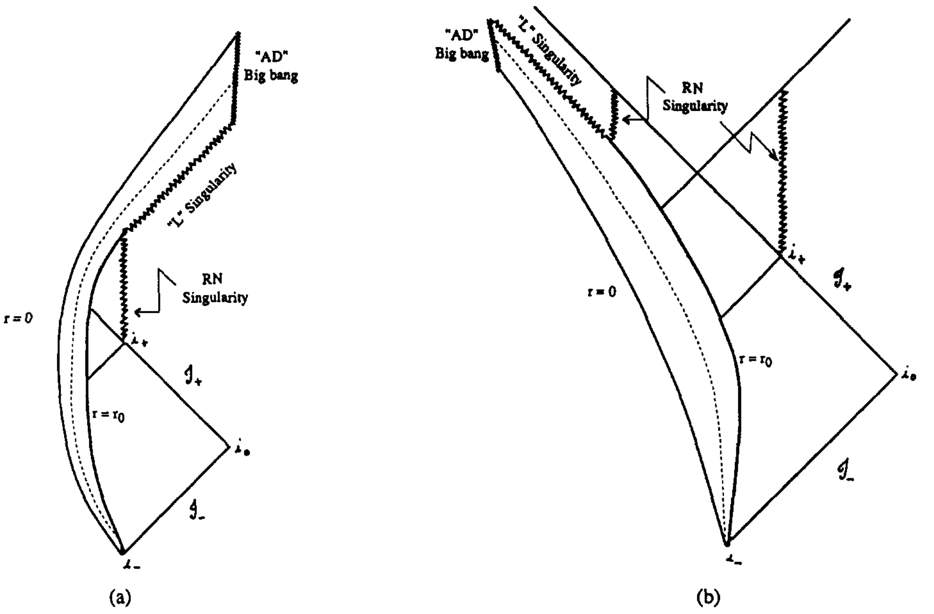


FIG. 12. Qualitative Penrose diagram of a charged collapsing sphere of type (iv). As in Fig. 11, the world lines of comoving observers (dotted curves) are complete, avoid the null L singularity, heading towards the AD big bang in their infinite future. In order to verify whether the surface layer collapses into the right- (a) or left- (b) hand side Riessner–Nordström singularity, the criteriom developed by de Felice and Maeda (Ref. 34) should be applied to Eqs. (74c) and (75).

with M and $(\Theta/3)^2$ given by (23) and (74a), respectively. For M- and W-type solutions, this constraint becomes

$$fH^2\Theta/3 = \pm\,[(1-ky+f^2h_y/h)H-(fh)^2[\Delta H^2 - 2\mu hH+\epsilon^2h^2+Lh^{-2}H^4]^{1/2}]\,, \quad (76b)$$

which, as expected, reduces to $R_0^2-2mR_0+e^2=0$ at $r=r_0$. Since the AD big bang given by $H=0$ in type (iv) solutions occurs in the infinite future of the fluid layers and the L big bang singularity cuts off the interior of the sphere from the external asymptotically flat region, it is unlikely that information from $H=0$ $(0\leqslant r<r_0)$ could ever reach observers in the latter region. As (76b) has no solution for the combination of values $(H=0, r>0)$ and $(H>0, r=0)$, the apparent horizon for type (iv) solutions reaches the center of the sphere at a value of t such that $H(t,0)=0$, such a coordinate value corresponds to a point of the AD big bang at infinite proper time distance. Hence, $H(t,r)=0$ is wholly covered by the apparent horizon (see Fig. 14). The apparent horizon, AD big bang and the black hole formed by the collapse of a neutral sphere of type (iv) is shown in the Penrose diagrams of Fig. 14.

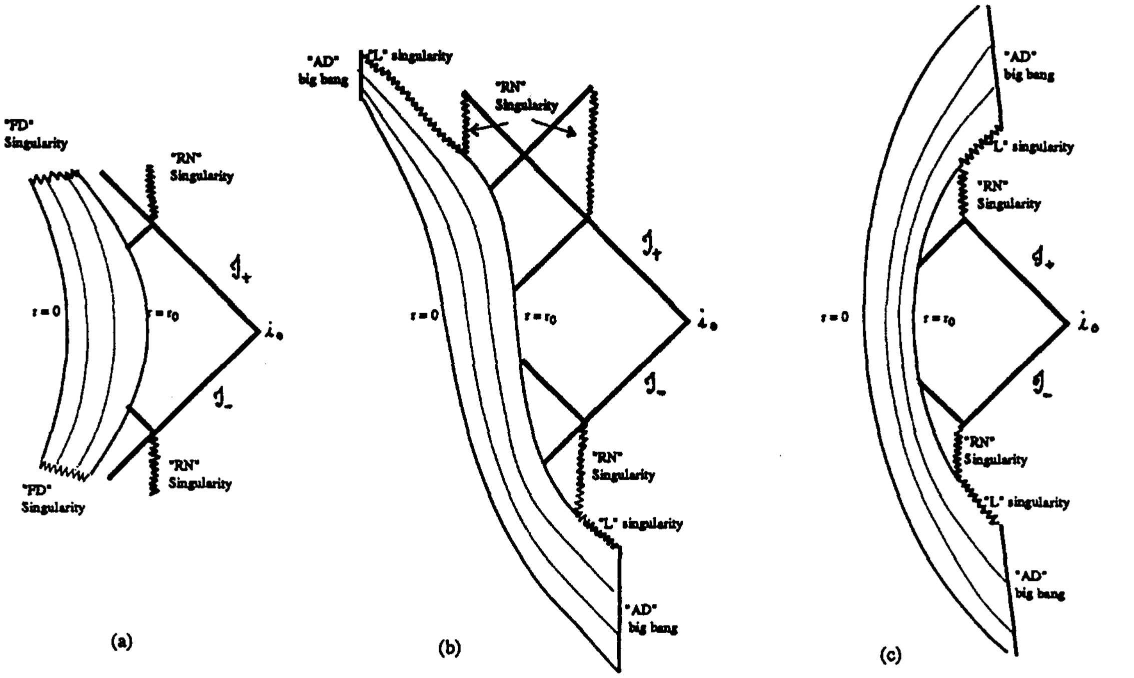
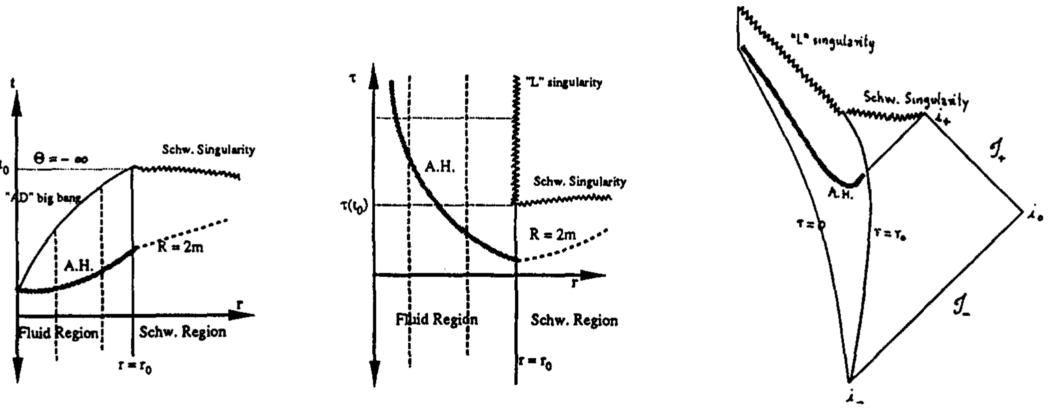


FIG. 13. Qualitative Penrose diagrams of bouncing spheres. The qualitative diagrams of Figs. 11 and 12 assumed that fluid layers collapse from infinity. These figures describe possible bouncing time symmetric evolution patterns. In (a) [sphere of type (ii)], the fluid layers emerge from the FD singularity and bounce back into the latter in the future. In (b) and (c), a type (iv) sphere bounces back into the AD big bang and L singularity. In order to find out whether the layers evolve as in (b) or (c), see Ref. 34.



(a) (b) (c)

FIG. 14. Apparent horizon of a collapsing sphere of type (iv). Since Eq. (76b) is inconsistent with values ($r=0, H>0$) and ($r>0, H=0$), the apparent horizon [curve in (a)] must reach the coordinate value ($r=0$, $H=0$). However, as shown by (b), this coordinate value is in the infinite future of the comoving observers along the center ($r=0$). (c) illustrates the apparent horizon in the qualitative Penrose diagram of the configuration. From this figure, it is clear that light rays from the L singularity of the AD big bang do not reach the Schwarzschild region outside the event horizon $R=2m$. Similar diagrams can be constructed for a charged sphere matched to Riessner–Nordström space-time.

For solutions of type (ii), the apparent horizon intersects the FD singularity $Q=0$ at least in one value (t,r) (with $0<r<r_0$) of the interior of the sphere before reaching the center at $r=0$. This (see Fig. 15) follows from the fact that (76b) has no solution for ($r=0$, $H>0$) and $Q=0$ occurs at points with $H>0$. The coordinates of the intersection are found by inserting $Q=0$ in (76b), leading to the constraint

$$(\Theta/3)fH = \pm[1-ky+f^2h_y/h], \tag{77}$$

which has always at least one solution. However, (77) can have more than one solution, which would mean that the apparent horizon intersects $Q=0$ more than once (see Fig. 15), and so this singularity could be "naked" (see Fig. 16). Since the apparent horizon might have a complicated shape, a sufficient condition to prevent $Q=0$ from being naked follows from the fact that this singularity emerges into the boundary of the sphere at $(R_0)_{(s)} = f_0h_0/A_0$, where $A_0>0$ is a solution of $Q_0=0$. Therefore, $Q=0$ will be censored by the outer event horizon of the Reissner–Nordstrøm exterior region (light rays from $Q=0$ will not reach distant observers in this region) if the constant parameters of the solution are chosen so that

$$f_0h_0/A_0 < m + [m^2-e^2]^{1/2} \tag{78}$$

holds [see Fig. 16(c)]. If the sphere is neutral, the constant: $m+(m^2-e^2)^{1/2}$ becomes $2m$, marking the apparent and event horizons of the Schwarzschild exterior region [see Fig. 16(a)]. Since the FD singularity $Q=0$ does not involve "shell-crossing" effects, nor "null dust," the possibility of having this singularity naked can be thought of as a new example[13] in the study of naked singularities in spherically symmetric collapse. However, the strong and dominant energy conditions are necessarily violated at $Q=0$, and so this case cannot be used as a counterexample of the weak cosmic censorship conjecture.[36]

Finally, for charged M- and W-type solutions, the ratio of charge density to matter energy density at the surface of the sphere can be related to the constant parameters of the solutions and to the ratio e/m by the following expression:

$$q_0/\rho_0 = \pm\frac{f_0^2[(d/dy)\ln(fh)]_0}{1-J_0/m}\frac{e}{m}, \tag{79}$$

showing that in some cases $e/m>1$ might not imply $q_0/\rho_0>1$. Although only the cases $e<m$ has been considered in this section, the cases $e=m$ and $e>m$ can also be examined having in mind the singularity structure of the fluid region; however, these cases will not be investigated in this paper.

ACKNOWLEDGMENTS

I wish to thank Professor M. A. H. MacCallum and Professor C. J. S. Clarke for their encouragement and their interest in this work.

I am indebted to the National University of Mexico (UNAM) for providing financial support.

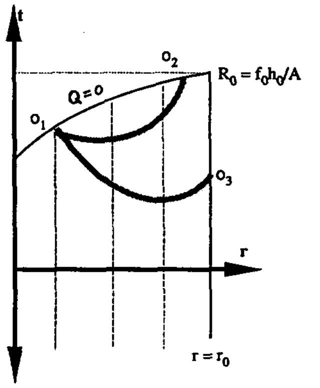


FIG. 15. Apparent horizon in a collapsing sphere of type (ii). Since Eq. (77) has no solution for $r=0$ and $H>0$, the apparent horizon cannot reach the center of the sphere. Depending on whether this equation has one or two solutions, the form of this horizon could be as the curve O_1–O_2 or O_1–O_3. In the former case, the FD singularity would be naked, while it would be censored in the latter. Surfaces near $r=0$ are not trapped, though it seems unlikely that light rays from this region could escape to the vacuum (Schwarzschild or Riessner–Nordström) region.

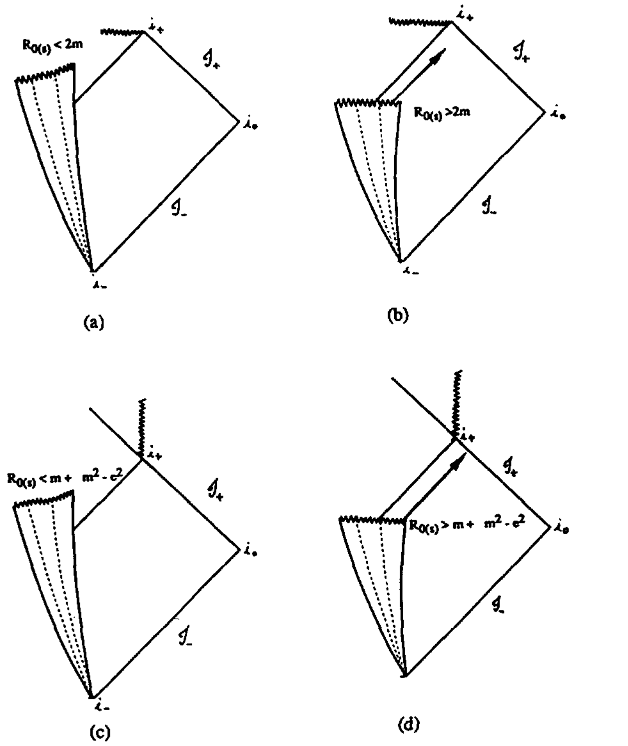


FIG. 16. Censorship of the FD singularity in type (ii) spheres. Depending on whether the parameters of the solution allow for $R_{O(s)} > m + [m^2 - e^2]^{1/2}$ (or $R_{O(s)} > 2m$) to hold, the FD singularity will be naked or censored. This situation is shown for neutral [(a) and (b)] and charged [(c) and (d)] solutions. The arrows in figures (b) and (d) represent light rays emmited from the naked FD singularity reaching the future null infinity of the Schwarzschild and Riessner–Nordström regions.

APPENDIX A: THE WYMAN SOLUTION

As mentioned in the Introduction of Part I, besides FRW solutions, the Wyman solution[37] is the only neutral SSSF solution which satisfies a barotropic equation of state (BES) of the form $p = p(\rho)$. Hence, there is an extensive literature on this solution[11,12] and its charged version,[11,22] and in particular, Collins[12] has examined some aspects of its global structure. By applying results obtained in Secs. VI and X, previous work on local aspects of this solution will be complemented and expanded in this Appendix, while its global aspects will be discussed in Appendix A in Part III. Only the neutral Wyman solution will be considered.

The parameters characterizing the Wyman solution follow directly from imposing $S' = 0$ (the condition leading to a BES) into Eqs. (37). Choosing the time coordinate as in (33), this leads to the constraint

$$J'/f^3 = 4\pi(p+\rho)H^3[h'/h - fh^2 Q^{1/2}], \qquad \text{(A1)}$$

which, from the field equations in Secs. III and IV, implies that for Eq. (A1) to hold one must choose $k = 0$ and $a = b = 0$, $c > 0$, so that h in I(23) becomes $h = c^{-1/2}$. Inserting these restrictions into (A1), leads to

$$\frac{d}{dT}\frac{\Theta^2}{18} + 6L = 0$$

$$\Rightarrow \frac{\Theta}{3} = \begin{cases} [L_0 - 6LT]^{1/2}, & L \neq 0, \quad L_0 \text{ constant}, \qquad \text{(A2a)} \\ \text{const}, & L = 0, \qquad \text{(A2b)} \end{cases}$$

which (save for constant factors and differences of notation) coincides with the values of these parameters found in the literature dealing with this solution. In fact, my expressions (H, Q, L, L_0) correspond (in the same order) to Collins' expressions $(U, Y, -3A/4, -B)$.

From Eqs. (A2), the Wyman solution is in fact a class of solutions characterized by the parameters (μ, L, L_0), however, in order to avoid confusion between W-type solutions and the class of Wyman solutions, the latter will be referred to as the Wyman solution. If $L = 0$, the Wyman solution comprises a particular case of an M-type solution [the solution NMcV$(r3)(X5)$ with $k = 0$ and $\Theta = \text{const}$, see Table VI of Part I] discovered by Faulkes.[38] In this particular case, from (7), the Hubble scale factor can be expressed in terms of the proper time of comoving observers as $H \sim \exp(\tau)$. For $L \neq 0$, the possible forms of H for the Wyman solution have are given by Eqs. I(37), where the signs of L and μ determine the type (i), (ii), or (iv) of domain of regularity of the solutions (see Sec. VII). In this case, the relation between τ and H follows from (7) and (A2a) by eliminating $T = T(H,r)$ from Eq. I(36).

From a thermodynamical point of view, the BES satisfied by the Wyman solution is a formal equation of state, in the sense that it is *obtained* rather than imposed from physical considerations on the fluid model (see Sec. VI). For such a BES, the four-acceleration can be expressed as the gradient of the thermodynamical potential $\ln[(p+\rho)/N]$, which can be identified with the logarithm of the specific enthalpy of the fluid.[27,39] Another thermodynamical consequence of having a BES satisfied is that the temperature $\mathscr{T}$ can be defined by (40) in agreement with the nonstatic generalization of Tolman's law, and so the heat flux vector in (38) can vanish without having to impose the rather strong restriction that the coefficient of thermal conductivity κ vanishes identically [option (39b)]. In such conditions, there is heat exchange between neighboring comoving fluid elements, but in such a way that the entropy production is canceled. However, as commented by authors studying the Wyman solution, the relation between p and ρ that follows from its BES is unphysical, and this situation is obviously related to the formal nature of such a BES.

Depending on the parameters L and μ, which determine whether a Wyman solution is of type (i), (ii), or (iv) of domain of regularity (see Table II for a comparison of these types to those of Fig. 2 of Collins' paper), a given solution may present the spacelike FD singularity at $Q = 0$ [Eq. (49)] or the AD big bang at $H = 0$ [Eq. (48)], or both regularity boundaries (see Sec. X). These features where recorded by Mashhoon and Partovi[11] and Collins,[12] though Mashhoon and Partovi failed to noice that the proper time of comoving observers diverges as $H \to 0$ implying that $H = 0$ is not, technically speaking, the representation of a scalar curvature singularity (see Sec. X). The Wyman solution also presents the regularity boundary $\Pi = 0$ which is related to its asymptotical structure, and so will be discussed in Part III, Appendix A. The (T,r) coordinate representation of the domain of regularity and of the boundaries (48)–(50) together with a comparison of Fig. 2 in Collins' paper are offered in Table II.

TABLE II. Domain of regularity and boundaries of the Wyman solution. The form of H for each of the cases with $L \neq 0$ is given by Eqs. I(36). For the case with $L = 0$, see Table VI of Part I. The last column on the left-hand side provides a comparison between the classification of types (i), (ii), and (iv) with that of Collins' Fig. 2 (see Ref. 12).

Type			Domain of regularity	$Q(T,r)=0$	$H(T,r)=0$	$\Pi(T,r)=0$	Classification in Fig. 2 of Collins
(i)	$\mu<0$	$L<0$	$0<T+r^2/2<\mathrm{cn}^{-1}(0)$	$T=\mathrm{cn}^{-1}(0)-r^2/2$	$T=r^2/2$		2
(ii)	$\mu>0$	$L>0$	$0<T+r^2/2<\mathrm{cn}^{-1}[(\sqrt{3}-1)/(\sqrt{3}+1)]$	$T=-r^2/2$		$T=\mathrm{cn}^{-1}[(\sqrt{3}-1)/(\sqrt{3}+1)]-r^2/2$	1 and 4
(iv)	$\mu<0$	$L>0$	$0<T+r^2/2<\mathrm{cn}^{-1}(0)$		$T=-r^2/2$	$T=\mathrm{cn}^{-1}(0)$	3
	$\mu<0$	$L=0$	$0<[T+r^2/2]^2$		$T=-r^2/2$		3

From the point of view of its singularity structure, the Wyman solution is qualitatively analogous to other ChKQ solutions having the same types of domain of regularity and $\Theta(T)$ qualitatively analogous to (A2). These solutions do not obey a BES and, depending on the values of their parameters, may even satisfy extremely pathological equations of state. This situation occurs because, as mentioned earlier, the specific behavior of the regularity boundaries follows from the existence of curvature terms ${}^{(3)}\mathscr{R}$ and $a^{\alpha}{}_{;\alpha}$ product of the imposition of zero shear, which might diverge independently of the choice of a formal equation of state [i.e., choice of $\Theta(T)$]. In fact, other nonstatic ChKQ solutions not obeying a BES, such as the conformally flat subclass, are probably more physically acceptable than the Wyman solution. (See Sec. IX of Part III.)

APPENDIX B: PROOF OF THE CRITERION OF GENERALIZED AFFINE PARAMETER COMPLETENESS

In Sec. IX, the proposition due to Clarke (private communication) leading to Eq. (52a) establishes a criterion for an accelerated timelike curve to be incomplete, in the sense of having finite generalized affine parameter (GAP) length, as defined on p. 259 of Hawking and Ellis.[30] If the space-time manifold is inextendible, then such curves end at a singularity.[29] The above-mentioned proposition is proved below.

Let $\{\mathbf{e}_{(i)}(\tau)\}$, $(i)=0,1,2,3$, be a Lorentz frame along $\gamma(\tau)$ chosen to be parallely propagated as τ varies, that is,

$$\frac{D}{\partial\tau}\mathbf{e}_{(i)}(\tau)=0. \tag{B1}$$

If $\mathbf{V}(\tau)=V^i(\tau)\mathbf{e}_{(i)}(\tau)$ is the tangent vector to γ at τ, then the GAP length from τ_2 to τ is

$$\zeta(\tau_2,\tau)=\int_{\tau_2}^{\tau}\left[\sum_{i=0}^{3}(V^i)^2\right]^{1/2}d\tau, \tag{B2}$$

which is defined up to a constant multiple, depending on the choice of $\mathbf{e}_{(i)}$. The curve is incomplete if $\zeta(\tau_2,\tau)$ remains bounded as $\tau\to\tau_2$, and so its completeness does not depend on the choice of $\mathbf{e}_{(i)}$. A suitable choice is then

$$\mathbf{e}_{(0)}(\tau_2)=\mathbf{V}(\tau_2), \tag{B3}$$

so that from Eqs. (52a) and (B1) one has

$$a^{\alpha}=\frac{dV^i}{d\tau}e^{\alpha}_{(i)}$$

and

$$[\mathscr{A}(\tau)]^2=\eta_{ij}\frac{dV^i}{d\tau}\frac{dV^j}{d\tau}, \tag{B4}$$

where η_{ij} is the Minkowski metric.

Since τ is proper time, it follows that

$$(V^0)^2=1+\sum_{i=1}^{3}(V^i)^2, \tag{B5}$$

so that it is possible to define $\alpha\equiv\cosh^{-1}(V^0)$, $\alpha>0$, and set $n^i=V^i/\sinh\alpha$, whence (B5) gives $\Sigma\, n^i=1$. Then, Eq. (B4) becomes

$$[\mathscr{A}(\tau)]^2=\sinh^2\alpha\left(\frac{d\alpha}{d\tau}\right)^2+\sum_{i=1}^{3}\left[n^i\cosh\alpha\frac{d\alpha}{d\tau}+\sinh\alpha\frac{d\alpha}{d\tau}\right]^2\geqslant\left(\frac{d\alpha}{d\tau}\right)^2, \tag{B6}$$

and from (B2), the GAP length is

$$\zeta(\tau_2,\tau)=\int_{\tau_2}^{\tau}[\cosh^2\alpha+\sinh^2\alpha]^{1/2}d\tau<\sqrt{2}\int_{\tau_2}^{\tau}\cosh\alpha\, d\tau'\leqslant\sqrt{2}\int_{\tau_2}^{\tau}\cosh\left(\int_{\tau_2}^{\tau'}\mathscr{A}(\tau'')\,d\tau''\right)d\tau$$

[from (B3) and (B6)], and so

$$\zeta(\tau_2,\tau)<-\frac{\tau_1-\tau_2}{2}\left[\left(\frac{\tau_1-\tau}{\tau_1-\tau_2}\right)^{1-l}+\left(\frac{\tau_1-\tau}{\tau_1-\tau_2}\right)^{1+l}\right]+\frac{\tau_1-\tau_2}{1-l^2}.$$

From Eq. (52b), as $\tau+\tau_1$, $\zeta\to(\tau_1-\tau_2)/(1-l^2)$, and thus $\zeta(\tau_2,\tau)$ is bounded on $[\tau_2,\tau_1)$, proving GAP length incompleteness as required.

APPENDIX C: A SIMPLE COLLAPSING SOLUTION

It is very useful to use a simple example in order to illustrate various properties which are common to M- and W-type solutions in general. The simple subclass of solutions labeled as ChMcV$(r2,r2)(X2)$ in the classification scheme of Part I (see Tables III and VIII of Part I) will be used as such an example in this Appendix. This subclass is a particular case of the class discovered by Vaidya and Shah[40] (see

Appendix D of Part I), and also follows from ChMcV$(r2)(X2)$ by taking $\mu^2 = \Delta\epsilon^2$. It has been discussed previously (see Table IV of Part I) by Banerjee, Chakravorty, and Duttachoudhuri[41] and by Mashhoon and Partovi[7] (their particular case, $v = 0, \gamma = 0$, and $\delta = 1$) as models of collapsing spheres. However, the former authors did not examine its singularity structure, while the latter did (see their Appendix D) in a nonrigorous manner, and thus some of their conclusions are wrong.

ChMcV$(r2,r2)(X2)$ solutions are characterized by the following parameter values in Eqs. I(21b), I(21c), I(23), and (25a): $L = 0$, $a = 0$, $\Delta = (\mu/\epsilon)^2 = b^2 > 0$. If the solutions are regular at the center $r = 0$, conditions (45) must hold, and this requires one to set $c > 0$. This specific combination of parameters leads to

$$Q = (\epsilon h/H - b)^2, \tag{C1a}$$

$$h = [2by + c]^{-1/2}, \tag{C1b}$$

$$fR'/R = 1 - ky - \epsilon f^2 h^3/H = 1 - ky - E/R, \tag{C1c}$$

where $E = \pm\epsilon(fh)^3$, and f and y are defined by Eqs. I(16b) and I(17). From Eqs. (C1), (14a), (16), (20), and (21), the state variables are

$$4\pi q = 3\epsilon h^5[c - ky(by + c)]/H^3, \tag{C2a}$$

$$\frac{8}{3}\pi\rho = \frac{\Theta^2}{9} + \frac{1}{H^2} \times \left[k - \frac{2f^3h^3[\mu(fh)^2 - (1-ky)\epsilon]}{H}\right], \tag{C2b}$$

$$8\pi p = -\frac{\Theta^2}{3} - \frac{2\epsilon H(\Theta/3)}{\epsilon^2 h - \mu H}\frac{\Theta}{3} - \frac{k}{H^2}, \tag{C2c}$$

from which one can identify the regularity boundaries $H = 0$ (q,ρ diverge, p diverges if $k = \pm 1$) and $Q = 0$ (only p diverges), the latter given explicitly as $H = (\mu/\epsilon^2)h$. Since H is given in Table III of Part I as $H = b^{-1}[T(t) + \mu h]$, the regularity boundaries $H = 0$ and $Q = 0$ can be expressed as $T = -\epsilon h$ and $T = 0$, respectively. Hence, in agreement with Table I, if $\mu > 0$ the solutions can be of type (i) $(0 < R < A)$; type (ii) $(R > A)$ with $A \equiv (\mu/\epsilon^2)fh$; or type (iv) with $R > 0 (r > 0)$ if $\mu < 0$.

In order to use a ChMcV$(r2,r2)(X2)$ solution as a model of a sphere collapsing in a Reissner–Nordstrøm spacetime, the time coordinate can be chosen as in Eqs. (32) and the matching condition (73b) leads to Eqs. (74), which for this particular subclass take the simple form

$$\left(\frac{\Theta R_0}{3}\right)^2 = (\dot{R}_0)^2 = \frac{2[m - (1 - ky_0)e]}{R_0} - kf_0^2, \tag{C3}$$

with e and m given by (74b) and (74c). Since R can be expressed as a function of R_0 [replacing $T(t)$ as t parameter] and r as

$$R(R_0,r) = (f/f_0)R_0 + (\epsilon^2/\mu)f(h - h_0) \tag{C4}$$

once Eq. (C3) is integrated, the time evolution of all "internal" layers [$R(t,r)$ for $0 \leqslant r < r_0$] can be found by inserting $R_0(t)$ into (C4). In particular, taking the case $m > e > 0$, $c > 0$, and $k = 0$, so that $f = r$, $y = r^2/2$ and $h = [(\mu/\epsilon)r^2 + c]^{-1/2}$, and thus

$$R_0(t) = (\tfrac{3}{2}\sqrt{2(m-e)}\, t)^{2/3}. \tag{C5}$$

Hence, for $\mu > 0$ [types (i) and (ii)], the FD singularity $Q = 0$ occurs at $R_0 = R_{0(s)} = (\epsilon^2/\mu)r_0 h_0$, which corresponds to the coordinate time value

$$\tfrac{3}{2}t_{(s)} = \epsilon^3(r_0h_0)^{3/2}/[2\mu^3(m-e)]^{1/2}, \tag{C6}$$

or, using $T(t)$ as the t parameter instead of R_0, at $T(t) = 0$. Since in this particular case this singularity coincides with a (singular) surface Σ_t, it is clear that it is spacelike, though this can be verified also through Eq. (56) and Fig. 2. Mashhoon and Partovi[7] erroneously concluded [see the paragraph following Eq. (90) in Sec. V of their paper] that this singularity is null, and that it occurs at a surface of constant comoving radial coordinate $r_{\rm MP} = 0$, where $I_{\rm MP}$ is the radial coordinate defined by their Eq. (74). (Reference to quantities defined by Mashhoon and Partovi are denoted with a subscript MP.) Also, in their Appendix C they failed to notice that the metric coefficient $A_{\rm MP}$ [Eq. (D1)] can vanish, and thus $|p| \to \infty$, for whatever choice of function $U_{\rm MP}$ (equivalent to h in this paper) if the t parameter $f_{\rm MP}$ [equivalent to $T(t)$] defined by their equation (35) vanishes. Their commentaries on possible singularities if $U_{\rm MP}$ diverges describe a different situation concerning solutions in which conditions (45) do not hold. This situation will be discussed in Secs. V and VI of Part III.

The proper time of comoving observers in the "interior" layers can be computed from Eqs. (53) and (C1a), (C4), and (C5), leading to

$$(2\alpha)^{1/2}\mu\tau(t,r) = \frac{2}{3}R_0^{3/2} + \frac{\epsilon^2}{\mu}r_0(h - h_0)R_0^{1/2} + \frac{\epsilon}{\mu^{1/2}}[r_0(h - h_0)]\wp^{-1}\left(\frac{\mu R_0}{\epsilon^2(h - h_0)}\right)^{1/2}, \tag{C7a}$$

where $\alpha \equiv m - e$, and

$$\wp^{-1} \equiv \begin{cases} \tan^{-1}, & \text{if } \mu > 0, \quad \text{(C7b)} \\ \tanh^{-1}, & \text{if } \mu < 0. \quad \text{(C7c)} \end{cases}$$

Since $\mathscr{A} = \epsilon r h^3/H$ is finite at $R_{0(s)}$ (the locus of $Q = 0$), and τ from (C7) with $\mu > 0$ is also finite, the product $(\tau - \tau_1)\mathscr{A}$ always converges and so, the GAP completeness criterion (52a) holds. Also it can be verified by a qualitative analysis of Eqs. (55) to (58), that the affine parameter of null geodesics approaching the Σ_t surface (C6) is finite.

The condition (77) for the FD singularity $Q = 0$ to be censored can be given in terms of the ratio $m/e > 1$ as

$$\frac{\epsilon}{\mu}\frac{c}{r_0^2} < \frac{m}{e} - 1 + \sqrt{\frac{m^2}{e^2} - 1}, \tag{C8}$$

which will hold if the parameter ϵ, μ, c, and r_0 are conveniently chosen. In particular, if $\rho_0 = 0$ ("gaseous sphere") $(m/e) = (\mu/\epsilon)(r_0h_0)^2$ and the condition (C8) always holds. If (C8) holds, the evolution of the sphere in the analytical extension of Reissner–Nordstrøm is qualitatively analogous to that shown in Fig. 16(c).

However, a "bad" combination of parameters, such as $c = r_0^2$ and $\mu/\epsilon \ll 1$ when $e \gtrsim m$, could have condition (C8) violated leaving the singularity $Q = 0$ naked, and so one would have a situation qualitatively similar to that of Fig.

16(d). Though, from Eq. (C2c), $p \to -\infty$ as $Q \to 0$, and so energy conditions are necessarily violated near the FD singularity. Hence if this singularity is naked it cannot be used as a counterexample of the weak cosmic censorship conjecture.[35] If the case $k = 1$ were considered in (C3) instead of $k = 0$, then $f_0 = \sin r_0$, and the sphere as seen from the Reissner–Nordstrøm exterior would start at the finite-density singularity at $R_{0(s)} = (\epsilon^2/\mu)\cos r_0 h_0$, bouncing at $R_{0(\max)} = 2(m - \cos r_0 e)/\sin^2 r_0$ and collapsing back into $R_{0(s)}$. This evolution is illustrated in Fig. 13(a).

If $\mu < 0$, the ChMcV$(r2,r2)(X2)$ solutions are of type (iv), a situation overlooked by Mashhoon and Partovi[7] (it corresponds to negative values of their parameer λ_0). In this case $h = [c - (|\mu|/\epsilon)r^2]^{-1/2}$ and so, in order to keep this function bounded for all $r \geqslant 0$, it is necessary to choose, either $\epsilon < 0$, which implies from (2a) having a negatively charged sphere, or $r_0 > [c|\mu|/\epsilon]^{1/2}$ if $\epsilon > 0$. The constraint $Q = 0$ implies $R_{O(s)} < 0$, and so it does not occur during the evolution of the fluid. On the other hand, the regularity boundary $H = 0$ does occur, however, $R = 0$ $(r < r_0)$ and $R_0 = 0$ do not imply each other. Hence, $H = 0$ is marked by $R_0^* = (\epsilon^2 r_0/|\mu|)(h - h_0)$ corresponding to

$$\frac{3}{2}t^* = \frac{1}{[2\alpha]^{1/2}}\left[\frac{\epsilon^2 r_0}{|\mu|}(h - h_0)\right]^{3/2}, \tag{C9}$$

where $\alpha \equiv m - e$. From Eqs. (C2) one can see that if $k = 0$, ρ and q diverge as $H \to 0$, while p remains finite and negative. This situation is rather exceptional, and does not occur in more general solutions. If $k = -1$, then both p and ρ become infinite positive [$\mu < 0$ in (C2b)], and since $p \sim H^{-2}$ while $\rho \sim H^{-3}$, one has $\rho \gg p$, and this case seems to provide the less unphysical situation in collapsing ChMcV$(r2,r2)(X2)$ solutions.

For $r = r_0$, Eq. (C9) reduces to $t = 0$ (or $R_0 = 0$), indicating that the "surface" of the sphere collapses in finite proper time. Observers in the Reissner–Nordstrøm exterior comoving near this surface, inside the internal horizon, would detect how the radius of the sphere vanishes. However, from Eqs. (C7a) and (C7c), τ diverges for comoving observers approaching (C9) if $0 \leqslant r < r_0$, indicating that the world lines of these observers are GAP complete. Since $\mathscr{A}$ is finite, affine parameter completeness of null geodesics approaching coordinate values given by (9) can also be readily proved from Eqs. (55)–(58). Hence, the "interior" layers continue their evolution collapsing in the AD big bang in their infinite future, as indicated by Figs. 8(b) and 12. As discussed in Secs. X and XI, the surface of the sphere produces a null singularity (the L singularity), which is avoided by the internal fluid layers and joins to the timelike singularity of Reissner–Nordstrøm space-time as shown in Figs. 8(b) and 12. If $k = 1$ in Eq. (C3), the two possible time symmetric evolutions depicted in Figs. 13(b) or 13(c) could occur. As mentioned in Sec. XI and in the Introduction (Sec. I), the collapsing picture occuring in solutions of type (iv) (whether $k = 0$ or ± 1) has not been discussed before, and seems to be the first example in which the collapse of a sphere could be "survived" by observers in its interior.

A third possible evolution corresponds to a type (i) of domain of regularity (see Fig. 7). In this case, the evolution of the sphere is constrained to $0 < t < t_{(s)}$, where $t(s)$ is given by Eq. (C6); that is, between the spacelike FD singularity and the FV big-bang singularity which occurs at the Σ_t surface $t = 0$ where Θ diverges with $R_0 \to 0$ but $R > 0$ for $0 < r < r_0$. At the latter singularity, ρ and p diverge while q remains finite. If there is no matching with Reissner–Nordstrøm at $r = r_0$, the fluid extends along $0 \leqslant r < \infty$, leading to an unbounded configuration. If $\mu < 0$ and $\epsilon < 0$ [a negatively charged type (iv) solution with h given by (C1b)], and a localized equation of state satisfying the strong energy condition is chosen at $r = r_0$, one has a collapsing picture qualitatively similar to that described in Fig. 5. In particular, if $p(t,r_0) = 0$, leading to Eq. (C3), p calculated from Eq. (C2c) would be negative for all $r > r_0$.

The ChMcV$(r2,r2)(X2)$ solutions presented in this Appendix are probably the mathematically simplest ChKQ solutions. More complicated M- and W-type solutions of types (ii) or (iv) describing collapsing spheres, such as the cases ChMcV$(r2)(X1,2)$ discussed by Mashhoon and Partovi,[7] contain the same FD spacelike singularity if type (ii), the null L singularity with the interior layers collapsing in the infinite future if type (iv) or, in some cases, the FV singularity if the solution is type (i). As far as the singularity structure associated to their gravitational collapse is concerned, these spheres behave qualitatively in an analogous manner to the simple case discussed here; however, the time evolution equation (C3) and other expressions, such as the integral (7), might require numerical integration even for relatively simple cases.

[1] R. A. Sussman, J. Math. Phys. **28**, 1118 (1987). This paper is referred to as Part I, while Part III corresponds to R. A. Sussman, to be published in J. Math. Phys.
[2] G. E. Tauber and J. W. Weinberg, Phys. Rev. **122**, 1342 (1962).
[3] J. Ehlers, P. Geren, and R. K. Sachs, J. Math. Phys. **9**, 1344 (1968).
[4] C. B. Collins and G. F. R. Ellis, Phys. Rep. **56**, 1976 (1979).
[5] G. C. McVittie, Astrophys. J. **143**, 682 (1966); Ann. Inst. H. Poincaré, VI, 1 (1967).
[6] E. N. Glass and B. Mashhoon, Astrophys. J. **205**, 570 (1976).
[7] B. Mashhoon and H. Partovi, Phys. Rev. D **20**, 2455 (1979).
[8] H. Knutsen, J. Math. Phys. **24**, 2188 (1983).
[9] A. Krasinski, Gen. Relativ. Gravit. **15**, 673 (1983).
[10] H. Knutsen, Phys. Scr. **30**, 289 (1984).
[11] B. Mashhoon and H. Partovi, Phys. Rev. D **30**, 1839 (1984).
[12] C. B. Collins, J. Math. Phys. **26**, 2009 (1985).
[13] M. Demianski and J. P. Lasota, Astrophys. Lett. **1**, 255 (1968); P. Yodzis, H. J. Seifert, and H. Müller zum Hagen, Commun. Math. Phys. **34**, 135 (1973); **37**, 29 (1974); D. Christodoulou, *ibid.* **93**, 171 (1974). For a review of papers dealing in this topic, see R. P. Newman, *Asymptotic Properties and Global Structure of Spacetime*, NATO Series B (Plenum, New York, 1986), Vol. 137.
[14] G. F. R. Ellis, in *General Relativity and Cosmology, XLVII Enrico Fermi Summer School Proceedings*, edited by R. K. Sachs (Academic, New York, 1971).
[15] D. Kramer, H. Stephani, M. A. H. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).
[16] G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969).
[17] J. Ehlers, in *Relativity, Astrophysics and Cosmology* (Reidel, Dordrecht, Holland, 1973), pp. 114–116; see also C. W. Misner and D. H. Sharp, Phys. Rev. **136**, B571 (1964).
[18] J. D. Beckenstein, Phys. Rev. D. **4**, 2185 (1971).
[19] E. N. Glass, J. Math. Phys. **20**, 1508 (1979).
[20] B. Mashhoon and H. Partovi, Ann. Phys. (NY) **130**, 99 (1982).
[21] P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists* (Springer, Berlin, 1971), 2nd ed.
[22] D. C. Srivastava and S. Prasad, Gen. Relativ. Gravit. **15**, 65 (1983).

[23]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[24]P. T. Landsberg, J. Phys. A **7**, 859 (1974).
[25]J. M. Stewart, *Non-equilibrium Relativistic Kinetic Theory* (Springer, Berlin, 1971).
[26]J. Ehlers, *General Relativity and Cosmology*, LXVII Enrico Fermi Summer School Proceedings, edited by R. K. Sachs (Academic, New York, 1971).
[27]R. Treciokas and G. F. R. Ellis, Commun. Math. Phys. **23**, 1 (1971).
[28]R. C. Tolman, Phys. Rev. **35**, 896 (1930).
[29]G. F. R. Ellis and B. G. Schmidt, Gen. Relativ. Gravit. **8**, 915 (1977).
[30]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U.P., Cambridge, 1973).
[31]S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford U.P., Oxford, 1978), p. 20, Eqs. (125) and (126).
[32]P. Vickers, Ph.D. thesis, University of London, 1973.
[33]R. A. Sussman, Gen. Relativ. Gravit. **17**, 251 (1985).
[34]F. de Felice and K. Maeda, Prog. Theor. Phys. **68**, 1967 (1982).
[35]W. C. Hernandez and C. W. Misner, Astrophys. J. **143**, 452 (1966).
[36]R. Penrose, Nuovo Cimento **1**, 252 (1969).
[37]M. Wyman, Phys. Rev. **70**, 396 (1946).
[38]M. C. Faulkes, Prog. Theor. Phys. **42**, 1139 (1969).
[39]J. Ehlers, *Beiträge zur Relativistischen Mechanik Kontinuierlicher Medien* (Verlag der Akademie der Wissenschaften und der Literatur, Mainz, 1961).
[40]P. C. Vaidya and Y. Shah, Ann. Inst. H. Poincaré **6**, 219 (1967).
[41]A. Bannerjee, N. Chakravorty, and S. B. Duttachoudhuri, Nuovo Cimento B **29**, 357 (1975).

The separability of Maxwell's equations in type-D backgrounds

G. F. Torres del Castillo
Departamento de Física Matemática, Instituto de Ciencias de la Universidad Autónoma de Puebla, Puebla, Mexico and Mathematical Institute, University of Oxford, Oxford OX1 3LB, England

(Received 4 June 1987; accepted for publication 11 November 1987)

It is shown that in a type-D space-time that admits a two-index Killing spinor a differential operator can be constructed that maps a solution of the Maxwell equations into another solution. By considering as a background the Plebański–Demiański metric, which includes all the vacuum type-D metrics, this operator is used to obtain all the components of the electromagnetic field and the vector potential. The separated functions appearing in the solutions are shown to obey identities of the Teukolsky–Starobinsky type and the separable solutions are shown to be eigensolutions of a certain differential operator with the "Starobinsky constant" as the eigenvalue.

I. INTRODUCTION

The study of the behavior of various kinds of fields on curved space-times has received considerable interest in recent years and there exist by now many results in this area. Especially important in this connection are the type-D space-times, which include the Kerr solution, because of their properties. In fact, most of the results obtained so far are concerned with type-D space-times partly because in these backgrounds some of the relevant equations can be solved by separation of variables.

As a result of its intrinsic interest, the Kerr solution has been the subject of many of the studies in this area and the successful results achieved in this case have stimulated similar investigations with other backgrounds, and the search for the reasons behind these successes. For example, Walker and Penrose[1] showed that the existence of Carter's "fourth constant," which is related to the separability of the Hamilton–Jacobi equation in the Kerr background, is associated with a two-index Killing spinor, whose existence follows from the Bianchi identities in all type-D vacuum space-times (see also Ref. 2). Similarly, Carter and McLenaghan[3] found that the separability of the Dirac equation in the Kerr geometry, obtained by Chandrasekhar,[4] is related to the fact that the skew-symmetric tensor corresponding to the two-index Killing spinor admitted by the Kerr metric is a Killing–Yano tensor. More specifically, from this Killing–Yano tensor a differential operator can be constructed, which commutes with the Dirac operator, and the separable solutions of the Dirac equation are eigensolutions of this operator. It has also been shown, in an explicit way, that the Dirac equation admits separable solutions in a restricted class of the type-D vacuum metrics.[5]

In the cases of the massless free fields with spins $\frac{1}{2}$ and 1 and of the gravitational perturbations, Teukolsky[6] found that, in the Kerr geometry, certain components of the fields satisfy decoupled equations that admit separable solutions. This result was extended by Dudley and Finley[7] to the Plebański–Demiański[8] metric, which includes all the type-D solutions of the Einstein vacuum field equations. However, when the spin of the field is greater than $\frac{1}{2}$, these decoupled components do not constitute all the components of the field. In the case of the Kerr metric, by integrating the corresponding field equations, Chandrasekhar[9,10] obtained the expressions for all the components of the fields, including those of the vector potential for the electromagnetic (i.e., the spin-1) field (see also Ref. 11).

Kamran and McLenaghan[12] have shown that the spin-$\frac{1}{2}$ massless field equations are separable in a wide class of type-D backgrounds and that there is a differential operator, constructed from the two-index Killing spinor that these backgrounds admit, which maps a solution of the spin-$\frac{1}{2}$ massless field equations into another solution, and the separable solutions are eigensolutions of this differential operator. The implications of the existence of a Killing spinor, in relation to the separability of the massless field equations, have been also investigated by Jeffryes[13] who obtained the form of the type-D metrics that admit such a spinor field, without imposing explicit restrictions on the Ricci tensor.

Another remarkable feature found in the study of the perturbations of the Kerr metric is the existence of the Teukolsky–Starobinsky identities (see, e.g., Ref. 14) that relate the separated functions corresponding to the components of the extreme helicities (see also Ref. 13).

In this paper, following Ref. 15, we show explicitly that in a type-D background that admits a two-index Killing spinor one can construct a differential operator that maps a solution of the Maxwell equations into another solution and by taking the Plebański–Demiański metric as background, this operator is used to construct the complete solutions to the Maxwell equations, including the corresponding vector potential, in terms of separated functions. We also find, in this general background, that these separated functions satisfy identities analogous to the Teukolsky–Starobinsky identities, previously found in the case of the Kerr background. We show that the separable solutions of the Maxwell equations are eigensolutions of a certain differential operator with the constant appearing in the Teukolsky–Starobinsky identities as an eigenvalue.

In most of this paper we make use of the notation of Newman and Penrose[16] (see also Refs. 14 and 17). For the sake of completeness and in order to show its analogy with the spin-1 case, the case of the spin-$\frac{1}{2}$ massless fields is presented in the Appendix.

 0022-2488/88/040971-07$02.50

II. PRELIMINARIES

A two-index Killing spinor[1] is a symmetric spinor field L_{AB} that satisfies

$$\nabla_{A'(B}L_{CD)} = 0, \tag{1}$$

where the round brackets denote symmetrization on the indices enclosed. If L_{AB} is algebraically general (which is equivalent to $L_{AB}L^{AB} \neq 0$) then it can be expressed in the form

$$L_{AB} = -2\phi^{-1} o_{(A}\iota_{B)}, \tag{2}$$

where o_A, ι_A form a spin-frame,[17] with $o_A\iota^A = 1$, and ϕ is a complex function. Substituting (2) into (1) one obtains, using the standard Newman–Penrose[16] notation,

$$\kappa = \sigma = \lambda = \nu = 0, \tag{3a}$$

$$\rho = D\ln\phi, \quad \tau = \delta\ln\phi, \quad \pi = -\bar{\delta}\ln\phi, \quad \mu = -\Delta\ln\phi. \tag{3b}$$

The integrability conditions of Eqs. (3) imply that only Ψ_2 may be different from zero; hence the space-time must be of type D or conformally flat. In the case of a vacuum type-D space-time, from the Bianchi identities one can easily see that $\phi = (\Psi_2)^{1/3}$ satisfies Eqs. (3b).[1] From Eq. (2) one obtains the expression

$$\phi^{-2} = -\tfrac{1}{2}L_{AB}L^{AB}. \tag{4}$$

The (source-free) Maxwell equations expressed in terms of the electromagnetic spinor φ_{AB}, $\nabla^{AC'}\varphi_{AB} = 0$, amount to

$$\begin{aligned}
&(\bar{\delta} - 2\alpha + \pi)\varphi_0 - (D - 2\rho)\varphi_1 - \kappa\varphi_2 = 0,\\
&(\Delta - 2\gamma + \mu)\varphi_0 - (\delta - 2\tau)\varphi_1 - \sigma\varphi_2 = 0,\\
&(\bar{\delta} + 2\pi)\varphi_1 - (D + 2\epsilon - \rho)\varphi_2 - \lambda\varphi_0 = 0,\\
&(\Delta + 2\mu)\varphi_1 - (\delta + 2\beta - \tau)\varphi_2 - \nu\varphi_0 = 0,
\end{aligned} \tag{5}$$

and the expression for the electromagnetic spinor in terms of the vector potential $\Phi_{AA'}$, $\varphi_{AB} = \nabla_{A'(A}\Phi^{A'}_{B)}$, is given explicitly by

$$\begin{aligned}
\varphi_0 &= (D - \epsilon + \bar{\epsilon} - \bar{\rho})\Phi_{01'} - (\delta - \beta - \bar{\alpha} + \bar{\pi})\Phi_{00'}\\
&\quad + \kappa\Phi_{11'} - \sigma\Phi_{10'},\\
2\varphi_1 &= (D + \epsilon + \bar{\epsilon} + \rho - \bar{\rho})\Phi_{11'}\\
&\quad + (\bar{\delta} - \alpha + \bar{\beta} - \pi - \bar{\tau})\Phi_{01'}\\
&\quad - (\delta + \beta - \bar{\alpha} + \tau + \bar{\pi})\Phi_{10'}\\
&\quad - (\Delta - \gamma - \bar{\gamma} - \mu + \bar{\mu})\Phi_{00'},\\
\varphi_2 &= (\bar{\delta} + \alpha + \bar{\beta} - \bar{\tau})\Phi_{11'} - (\Delta + \gamma - \bar{\gamma} + \bar{\mu})\Phi_{10'}\\
&\quad + \nu\Phi_{00'} - \lambda\Phi_{01'}.
\end{aligned} \tag{6}$$

Whenever κ, σ, Ψ_0, and Ψ_1 vanish, by making use of the commutation relation[6]

$$\begin{aligned}
&(D + (p-1)\epsilon + \bar{\epsilon} + q\rho - \bar{\rho})(\delta + p\beta + q\tau)\\
&\quad = (\delta + (p-1)\beta - \bar{\alpha} + q\tau + \bar{\pi})(D + p\epsilon + q\rho),
\end{aligned} \tag{7}$$

where p and q are constants, from Eqs. (5) a decoupled equation for φ_0 can be derived, namely,

$$\begin{aligned}
&[(D - \epsilon + \bar{\epsilon} - 2\rho - \bar{\rho})(\Delta - 2\gamma + \mu)\\
&\quad - (\delta - \beta - \bar{\alpha} - 2\tau + \bar{\pi})(\bar{\delta} - 2\alpha + \pi)]\varphi_0 = 0.
\end{aligned} \tag{8}$$

Analogously, if λ, ν, Ψ_3, and Ψ_4 are equal to zero, one gets the following decoupled equation for φ_2:

$$\begin{aligned}
&[(\Delta + \gamma - \bar{\gamma} + 2\mu + \bar{\mu})(D + 2\epsilon - \rho)\\
&\quad - (\bar{\delta} + \alpha + \bar{\beta} + 2\pi - \bar{\tau})(\delta + 2\beta - \tau)]\varphi_2 = 0.
\end{aligned} \tag{9}$$

A relevant fact is that Eqs. (8) and (9) can be solved by separation of variables in the Kerr background[6,14] and, more generally, in the Plebański–Demiański background.[7]

III. GENERATION OF SOLUTIONS

In the forthcoming we shall assume that the space-time admits a two-index Killing spinor of the form (2). As mentioned above, all the type-D vacuum space-times admit such a spinor field.

In Ref. 15 it was stated without proof that if φ_{AB} is a solution to Maxwell's equations then

$$\chi_{R'S'} \equiv \tfrac{1}{2}\nabla^{R}{}_{(R'}\phi^{-2}\nabla^{S}{}_{S')}\phi^2 L_{AR}L_{BS}\varphi^{AB} \tag{10}$$

also satisfies Maxwell's equations,

$$\nabla^{CR'}\chi_{R'S'} = 0. \tag{11}$$

In fact, the following stronger claim can be made. The (complex) vector field

$$X_{RS'} \equiv -\tfrac{1}{2}\phi^{-2}\nabla^{S}{}_{S'}(\phi^2 L_{AR}L_{BS}\varphi^{AB}) \tag{12}$$

is the vector potential of a self-dual electromagnetic field, i.e.,

$$\nabla_{A'(A}X^{A'}_{B)} = 0, \tag{13}$$

from which Eq. (11) follows. [Actually, also in the case of a self-dual or anti-self-dual Yang–Mills field, the (source-free) field equations are automatically satisfied (see, e.g., Ref. 17).]

We shall prove (13) by a direct evaluation that will be useful later. The components $X_{RS'}$ defined in Eq. (12) are given explicitly by

$$\begin{aligned}
X_{00'} &= -\tfrac{1}{2}\phi^{-2}\{(D - 2\rho)\varphi_1 + (\bar{\delta} - 2\alpha + \pi)\varphi_0\},\\
X_{01'} &= -\tfrac{1}{2}\phi^{-2}\{(\delta - 2\tau)\varphi_1 + (\Delta - 2\gamma + \mu)\varphi_0\},\\
X_{10'} &= \tfrac{1}{2}\phi^{-2}\{(D + 2\epsilon - \rho)\varphi_2 + (\bar{\delta} + 2\pi)\varphi_1\},\\
X_{11'} &= \tfrac{1}{2}\phi^{-2}\{(\delta + 2\beta - \tau)\varphi_2 + (\Delta + 2\mu)\varphi_1\}
\end{aligned} \tag{14}$$

[cf. Eqs. (5)], where we have made use of Eqs. (2) and (3a). By using Maxwell's equations (5) and Eqs. (3), these expressions can be rewritten in many equivalent forms, e.g.,

$$\begin{aligned}
X_{00'} &= -\phi^{-2}(\bar{\delta} - 2\alpha + \pi)\varphi_0,\\
X_{01'} &= -\phi^{-2}(\Delta - 2\gamma + \mu)\varphi_0,\\
X_{10'} &= \phi^{-2}(D + 2\epsilon - \rho)\varphi_2 = (D + 2\epsilon + \rho)(\phi^{-2}\varphi_2),\\
X_{11'} &= \phi^{-2}(\delta + 2\beta - \tau)\varphi_2 = (\delta + 2\beta + \tau)(\phi^{-2}\varphi_2).
\end{aligned} \tag{15}$$

Then, for instance,

$$\nabla_{A'0}X^{A'}_{0} = (D - \epsilon + \bar{\epsilon} - \bar{\rho})X_{01'} - (\delta - \beta - \bar{\alpha} + \bar{\pi})X_{00'}$$

[cf. the first equation in (6)], which, using Eqs. (3b) and (15), amounts to

$$\begin{aligned}
&-\phi^{-2}(D - \epsilon + \bar{\epsilon} - 2\rho - \bar{\rho})(\Delta - 2\gamma + \mu)\varphi_0\\
&\quad + \phi^{-2}(\delta - \beta - \bar{\alpha} - 2\tau + \bar{\pi})(\bar{\delta} - 2\alpha + \pi)\varphi_0,
\end{aligned}$$

which is zero according to the decoupled equation (8). In an analogous way one finds that the equation $\nabla_{A'1}X_1^{A'}=0$ is equivalent to the decoupled equation (9) and that the equation $\nabla_{A'(0}X_{1)}^{A'}=0$ is also satisfied as a consequence of the Maxwell equations (5).

Thus $\chi_{R'S'}=\nabla_{C(R'}X_{S')}^{C}$ satisfies Maxwell's equations (11) and, moreover, due to (13), $\chi_{R'S'}=\nabla_{C(R'}[X_{S')}^{C}+\overline{X}_{S')}^{C}]$, where $\overline{X}_{AB'}(=\overline{X_{BA'}})$ is the Hermitian conjugate of $X_{AB'}$. The combination $X_{AB'}+\overline{X}_{AB'}$ is a real vector potential that also generates the field $\chi_{R'S'}$. The fact that the field $\chi_{R'S'}$ given in Eq. (10) by means of a differential operator applied to a solution of the Maxwell equations is also a solution of Maxwell's equations can be expressed by saying that the differential operator in Eq. (10) commutes with the Maxwell operator modulo the Maxwell operator itself. The differential operator in (10) is analogous to that considered by Kamran and McLenaghan[12] in the case of the spin-$\frac{1}{2}$ massless fields (see also the Appendix).

The components $\chi_{R'S'}$ can be written in various equivalent ways by using Eqs. (3), (5), (7), and (14). For instance [cf. (6)],

$$\begin{aligned}\overline{\chi}_0&=(D+\epsilon-\bar{\epsilon}-\rho)X_{10'}-(\bar{\delta}-\alpha-\bar{\beta}+\pi)X_{00'}\\&=\phi^{-2}\{(D+\epsilon-\bar{\epsilon}-3\rho)(D+2\epsilon-\rho)\varphi_2\\&\quad+(\bar{\delta}-\alpha-\bar{\beta}+3\pi)(D-2\rho)\varphi_1\}\\&=\phi^{-2}(D+\epsilon-\bar{\epsilon}-3\rho)[(D+2\epsilon-\rho)\varphi_2\\&\quad+(\bar{\delta}+2\pi)\varphi_1]\\&=2\phi^{-2}(D+\epsilon-\bar{\epsilon}-3\rho)(D+2\epsilon-\rho)\varphi_2,\end{aligned}\tag{16}$$

and, eliminating φ_2 and φ_1 in favor of φ_0,

$$\begin{aligned}\overline{\chi}_0&=\phi^{-2}\{(D+\epsilon-\bar{\epsilon}-3\rho)(\bar{\delta}+2\pi)\varphi_1\\&\quad+(\bar{\delta}-\alpha-\bar{\beta}+3\pi)(\bar{\delta}-2\alpha+\pi)\varphi_0\}\\&=2\phi^{-2}(\bar{\delta}-\alpha-\bar{\beta}+3\pi)(\bar{\delta}-2\alpha+\pi)\varphi_0.\end{aligned}\tag{17}$$

This implies, in particular, the identity

$$\begin{aligned}&(D+\epsilon-\bar{\epsilon}-3\rho)(D+2\epsilon-\rho)\varphi_2\\&\quad=(\bar{\delta}-\alpha-\bar{\beta}+3\pi)(\bar{\delta}-2\alpha+\pi)\varphi_0,\end{aligned}\tag{18}$$

from which the Teukolsky–Starobinsky identities can be obtained. [The validity of this relation can be proven directly from the Maxwell equations and Eqs. (3a) and (7) once one considers the appropriate operators, which arise naturally in the present context.]

In a similar way one gets

$$\begin{aligned}\overline{\chi}_2&=2\phi^{-2}(\Delta-\gamma+\bar{\gamma}+3\mu)(\Delta-2\gamma+\mu)\varphi_0\\&=2\phi^{-2}(\delta+\beta+\bar{\alpha}-3\tau)(\delta+2\beta-\tau)\varphi_2\end{aligned}\tag{19}$$

and

$$\begin{aligned}\overline{\chi}_1&=\phi^{-2}[(D+\epsilon+\bar{\epsilon}-3\rho+\bar{\rho})(\delta+2\beta-\tau)\\&\quad+(\delta+\beta-\bar{\alpha}-3\tau-\bar{\pi})(D+2\epsilon-\rho)]\varphi_2\\&=\phi^{-2}[(\Delta-\gamma-\bar{\gamma}+3\mu-\bar{\mu})(\bar{\delta}-2\alpha+\pi)\\&\quad+(\bar{\delta}-\alpha+\bar{\beta}+3\pi+\bar{\tau})(\Delta-2\gamma+\mu)]\varphi_0.\end{aligned}\tag{20}$$

[Equations (19) and (20) provide additional relations similar to (18).]

In the following section these general results will be applied to the specific case of the type-D solutions of the Einstein field equations given in the form found by Plebański and Demiański.

IV. SOLUTION OF MAXWELL'S EQUATIONS IN THE PLEBAŃSKI–DEMIAŃSKI BACKGROUND

The Plebański–Demiański metric is given by

$$\begin{aligned}g=(1-pq)^{-2}\Big\{&\frac{\mathcal{Q}}{p^2+q^2}(du-p^2\,dv)^2-\frac{p^2+q^2}{\mathcal{Q}}dq^2\\&-\frac{\mathcal{P}}{p^2+q^2}(du+q^2\,dv)^2-\frac{p^2+q^2}{\mathcal{P}}dp^2\Big\},\end{aligned}\tag{21}$$

where p,q,u,v are real coordinates and $\mathcal{P}$ and $\mathcal{Q}$ are functions of p and q, respectively. In the notation of Ref. 8 (see also Ref. 7),

$$\begin{aligned}\mathcal{P}&=-(\lambda/6+g^2-\gamma)+2np-\epsilon p^2+2mp^3\\&\quad-(\lambda/6+e^2+\gamma)p^4,\\\mathcal{Q}&=-(\lambda/6-e^2-\gamma)-2mq+\epsilon q^2-2nq^3\\&\quad-(\lambda/6-g^2+\gamma)q^4.\end{aligned}\tag{22}$$

The parameters m, n, e, g, and λ correspond to mass, NUT parameter, electric and magnetic charge, and cosmological constant, respectively, and the parameters ϵ and γ are related to the angular momentum per unit mass, a, and the acceleration parameter b by

$$\epsilon=-\frac{1}{ab}\left(\frac{a^2-b^2}{a^2+b^2}\right)\left(1-\frac{\lambda}{3}(a^2+b^2)\right)^{1/2},$$

$$\gamma=(a^2+b^2)^{-1}-\frac{\lambda}{6}.$$

These parameters should not be confused with the spin coefficients; actually, the explicit expressions (22) will not be required in what follows. When $e=0=g$, the metric (21) is a solution of the Einstein vacuum field equations.

The tangent vectors

$$\begin{aligned}D&=\partial_q+(1/\mathcal{Q})(\partial_v-q^2\,\partial_u),\\\Delta&=\tfrac{1}{2}\phi\bar{\phi}\mathcal{Q}(-\partial_q+(1/\mathcal{Q})(\partial_v-q^2\,\partial_u)),\\\delta&=\left(\frac{\mathcal{P}}{2}\right)^{1/2}\bar{\phi}\left(\partial_p+\frac{i}{\mathcal{P}}(\partial_v+p^2\,\partial_u)\right),\\\bar{\delta}&=\left(\frac{\mathcal{P}}{2}\right)^{1/2}\phi\left(\partial_p-\frac{i}{\mathcal{P}}(\partial_v+p^2\,\partial_u)\right),\end{aligned}\tag{23}$$

where

$$\phi\equiv(1-pq)/(q+ip),\tag{24}$$

form a null tetrad such that D and Δ are double principal null directions of the conformal curvature. The spin coefficients are given by

$$\begin{aligned}&\kappa=\sigma=\lambda=\nu=0,\quad\epsilon=D\ln(\mathcal{P}^{1/4}/(1-pq)),\\&\beta=\delta\ln(\mathcal{P}^{1/4}/(1-pq)),\quad\rho=D\ln\phi,\\&\tau=\delta\ln\phi,\quad\pi=-\bar{\delta}\ln\phi,\quad\mu=-\Delta\ln\phi,\\&\alpha=-\bar{\delta}\ln(\mathcal{P}^{1/4}\mathcal{Q}^{1/2}/(q+ip)),\\&\gamma=-\Delta\ln(\mathcal{P}^{1/4}\mathcal{Q}^{1/2}/(q+ip)).\end{aligned}\tag{25}$$

Comparison with Eqs. (3) shows that the metric (21) admits a two-index Killing spinor of the form (2) with ϕ given by (24).

Contrary to some statements in the literature, in order to solve Eqs. (8) and (9) by separation of variables, it is not necessary to use a specific tetrad provided it satisfies Eqs. (3a). The relation between any two such tetrads corresponds to a rescaling of o_A and ι_A (the spin frame), which only rescales φ_0 and φ_2 (see, e.g., Ref. 18). The tetrad (23) is different from those used in Ref. 7; the present choice is such that D takes a relatively simple form and $\alpha = \pi - \bar{\beta}$. The representation of the spin coefficients in the form (25) does not depend on the tetrad chosen here and is very useful in the forthcoming computations.

A. The separated functions

Assuming that the components of the fields have a dependence on the ignorable coordinates u and v of the form $e^{i(ku+lv)}$, the tetrad vectors (23) can be replaced according to

$$\begin{aligned} &D \to \mathscr{D}_0, \quad \Delta \to -\tfrac{1}{2}\phi\bar{\phi}\mathscr{Q}\,\mathscr{D}_0^\dagger, \\ &\delta \to (1/\sqrt{2})\bar{\phi}\mathscr{L}_0^\dagger, \quad \bar{\delta} \to (1/\sqrt{2})\phi\mathscr{L}_0, \end{aligned} \tag{26}$$

where

$$\begin{aligned} \mathscr{D}_n &\equiv \partial_q + (i/\mathscr{Q})(l-kq^2) + (n/\mathscr{Q})\partial_q\mathscr{Q} = \mathscr{Q}^{-n}\mathscr{D}_0\mathscr{Q}^n, \\ \mathscr{D}_n^\dagger &\equiv \partial_q - (i/\mathscr{Q})(l-kq^2) + (n/\mathscr{Q})\partial_q\mathscr{Q} \\ &= \mathscr{Q}^{-n}\mathscr{D}_0^\dagger\mathscr{Q}^n, \\ \mathscr{L}_n &\equiv \sqrt{\mathscr{P}}(\partial_p + (1/\mathscr{P})(l+kp^2) + (n/2\mathscr{P})\partial_p\mathscr{P}) \\ &= \mathscr{P}^{-n/2}\mathscr{L}_0\mathscr{P}^{n/2}, \\ \mathscr{L}_n^\dagger &\equiv \sqrt{\mathscr{P}}(\partial_p - (1/\mathscr{P})(l+kp^2) + (n/2\mathscr{P})\partial_p\mathscr{P}) \\ &= \mathscr{P}^{-n/2}\mathscr{L}_0^\dagger\mathscr{P}^{n/2}. \end{aligned} \tag{27}$$

Notice that these operators depend parametrically on the separation constants k and l, and that

$$\begin{aligned} \mathscr{D}_n(k,l) &= \overline{\mathscr{D}_n(-k,-l)} \\ &= \overline{\mathscr{D}_n^\dagger(k,l)} = \mathscr{D}_n^\dagger(-k,-l), \\ \mathscr{L}_n(k,l) &= \mathscr{L}_n^\dagger(-k,-l). \end{aligned} \tag{28}$$

In what follows $\mathscr{D}_n$ will denote $\mathscr{D}_n(k,l)$ and so on.

Then, using Eqs. (25) and (26), the decoupled equation (8) takes the form

$$[\mathscr{D}_0\mathscr{D}_0^\dagger\mathscr{Q} + 2ikq + \mathscr{L}_0^\dagger\mathscr{L}_1 + 2kp](\varphi_0/(1-pq)) = 0, \tag{29}$$

hence

$$\varphi_0 = (1-pq)e^{i(ku+lv)}R_{+1}(q)S_{+1}(p), \tag{30}$$

with

$$\begin{aligned} &(\mathscr{Q}\mathscr{D}_0\mathscr{D}_0^\dagger + 2ikq)\mathscr{Q}R_{+1}(q) = A_1\mathscr{Q}R_{+1}(q), \\ &(\mathscr{L}_0^\dagger\mathscr{L}_1 + 2kp)S_{+1}(p) = -A_1S_{+1}(p), \end{aligned} \tag{31}$$

and A_1 is a separation constant. Similarly, from Eq. (9) one gets

$$\begin{aligned} &[\mathscr{Q}\mathscr{D}_0^\dagger\mathscr{D}_0 - 2ikq + \mathscr{L}_0\mathscr{L}_1^\dagger - 2kp] \\ &\quad\times(\phi^{-2}\varphi_2/(1-pq)) = 0; \end{aligned} \tag{32}$$

therefore

$$\varphi_2 = (1-pq)\phi^2e^{i(ku+lv)}R_{-1}(q)S_{-1}(p), \tag{33}$$

with

$$\begin{aligned} &(\mathscr{Q}\mathscr{D}_0^\dagger\mathscr{D}_0 - 2ikq)R_{-1}(q) = A_2R_{-1}(q), \\ &(\mathscr{L}_0\mathscr{L}_1^\dagger - 2kp)S_{-1}(p) = -A_2S_{-1}(p), \end{aligned} \tag{34}$$

where A_2 is a separation constant.

The components $\bar{\chi}_0$ and $\bar{\chi}_2$ of the self-dual part of the electromagnetic field satisfy the complex conjugates of Eqs. (8) and (9), respectively. Therefore $\bar{\chi}_0$ and $\bar{\chi}_2$ could be expressed by the complex conjugates of Eqs. (30) and (33). However, in order to have a dependence of the form $e^{i(ku+lv)}$, the complex conjugation must be accompanied by the substitution of (k,l) by $(-k,-l)$. Thus

$$\bar{\chi}_0 = (1-pq)e^{i(ku+lv)}\widetilde{R}_{+1}(q)\widetilde{S}_{+1}(p) \tag{35}$$

and

$$\bar{\chi}_2 = (1-pq)\bar{\phi}^2e^{i(ku+lv)}\widetilde{R}_{-1}(q)\widetilde{S}_{-1}(p), \tag{36}$$

where, using Eqs. (28), (29), and (32), the functions $\widetilde{R}_{\pm 1}$, $\widetilde{S}_{\pm 1}$ must satisfy

$$\begin{aligned} &(\mathscr{Q}\mathscr{D}_0\mathscr{D}_0^\dagger + 2ikq)\mathscr{Q}\widetilde{R}_{+1}(q) = A_3\mathscr{Q}\widetilde{R}_{+1}(q), \\ &(\mathscr{L}_0\mathscr{L}_1^\dagger - 2kp)\widetilde{S}_{+1}(p) = -A_3\widetilde{S}_{+1}(p), \end{aligned} \tag{37}$$

and

$$\begin{aligned} &(\mathscr{Q}\mathscr{D}_0^\dagger\mathscr{D}_0 - 2ikq)\widetilde{R}_{-1}(q) = A_4\widetilde{R}_{-1}(q), \\ &(\mathscr{L}_0^\dagger\mathscr{L}_1 + 2kp)\widetilde{S}_{-1}(p) = -A_4\widetilde{S}_{-1}(p), \end{aligned} \tag{38}$$

with A_3 and A_4 being separation constants.

By comparing Eqs. (37) and (38) with (31) and (34) we see that we can take $\widetilde{R}_{\pm 1} = R_{\pm 1}$, $\widetilde{S}_{+1} = S_{-1}$, $\widetilde{S}_{-1} = S_{+1}$, and that the separation constants are related by

$$A_1 = A_2 = A_3 = A_4 \equiv A. \tag{39}$$

Furthermore, since R_{-1} and $\mathscr{Q}R_{+1}$ satisfy complex-conjugate equations, it follows that A is real,

$$A = \bar{A}. \tag{40}$$

[This fact can also be obtained independently from Eq. (47), below.]

B. The Teukolsky–Starobinsky identitites

From Eqs. (3b), (16), (25), and (26) we have

$$\begin{aligned} \bar{\chi}_0 &= 2(D + \epsilon - \bar{\epsilon} - \rho)(D + 2\epsilon + \rho)(\phi^{-2}\varphi_2) \\ &= 2(1-pq)\mathscr{D}_0\mathscr{D}_0(\phi^{-2}\varphi_2/(1-pq)) \end{aligned} \tag{41}$$

and, at the same time, from Eqs. (17), (25), and (26),

$$\bar{\chi}_0 = (1-pq)\mathscr{L}_0\mathscr{L}_1(\varphi_0/(1-pq)). \tag{42}$$

Analogously, from Eqs. (19), (25), and (26),

$$\bar{\chi}_2 = \tfrac{1}{2}(1-pq)\bar{\phi}^2\mathscr{Q}\mathscr{D}_0^\dagger\mathscr{D}_0^\dagger(\mathscr{Q}[\varphi_0/(1-pq)]) \tag{43}$$

$$\begin{aligned} &= 2(\delta + \beta + \bar{\alpha} - \tau)(\delta + 2\beta + \tau)(\phi^{-2}\varphi_2) \\ &= (1-pq)\bar{\phi}^2\mathscr{L}_0^\dagger\mathscr{L}_1^\dagger(\phi^{-2}\varphi_2/(1-pq)). \end{aligned} \tag{44}$$

Substituting now the separable solutions (30) and (33) into Eqs. (41)–(44) and comparing with (35) and (36) one finds that, by normalizing appropriately, the separated functions satisfy the relations

$$\begin{aligned}&\mathscr{D}_0\mathscr{D}_0R_{-1}(q)=BR_{+1}(q),\\&\mathscr{L}_0\mathscr{L}_1S_{+1}(p)=BS_{-1}(p),\\&\mathscr{Q}\mathscr{D}_0^\dagger\mathscr{D}_0^\dagger\mathscr{Q}R_{+1}(q)=BR_{-1}(q),\\&\mathscr{L}_0^\dagger\mathscr{L}_1^\dagger S_{-1}(p)=BS_{+1}(p),\end{aligned}\tag{45}$$

where B is a real constant [taking into account that $\mathscr{Q}R_{+1}$ and R_{-1} satisfy complex-conjugate equations and that $\bar{\chi}_0$ and $\bar{\chi}_2$ must satisfy the complex conjugate of (18)].

In what follows we shall assume that the separated functions are normalized in such a way that Eqs. (45) hold. Then the components of the electromagnetic field, with the correct relative normalization, are given by

$$\begin{aligned}&\varphi_0=(1-pq)e^{i(ku+lv)}R_{+1}(q)S_{+1}(p),\\&\varphi_2=\tfrac{1}{2}(1-pq)\phi^2e^{i(ku+lv)}R_{-1}(q)S_{-1}(p).\end{aligned}\tag{46}$$

The value of the constant B can be obtained by combining the first and the third equations in (45), which gives $\mathscr{Q}\mathscr{D}_0^\dagger\mathscr{D}_0^\dagger\mathscr{Q}\mathscr{D}_0\mathscr{D}_0R_{-1}=B^2R_{-1}$. Then, by commuting the differential operators and using Eq. (34), one obtains

$$B^2=A^2+4kl.\tag{47}$$

Following Ref. 14, B can be called the Starobinsky constant and the relations (45) will be called the Teukolsky–Starobinsky identities.

C. The complete solution and the vector potential

Using Eqs. (3b), (25), (26), (45), and (46), from Eqs. (20) a straightforward computation gives

$$\begin{aligned}\bar{\chi}_1=&(B/\sqrt{2})\bar{\phi}^2e^{i(ku+lv)}\\&\times[g_{+1}(q)\mathscr{L}_1^\dagger S_{-1}-if_{+1}(p)\mathscr{D}_0R_{-1}],\end{aligned}\tag{48}$$

where

$$\begin{aligned}&g_{+1}(q)\equiv(1/B)(q\mathscr{D}_0R_{-1}-R_{-1}),\\&f_{+1}(p)\equiv(1/B)(p\mathscr{L}_1^\dagger S_{-1}-\sqrt{\mathscr{P}}S_{-1})\end{aligned}\tag{49}$$

and, also,

$$\begin{aligned}\bar{\chi}_1=&-(B/\sqrt{2})\bar{\phi}^2e^{i(ku+lv)}\\&\times[g_{-1}(q)\mathscr{L}_1S_{+1}-if_{-1}(p)\mathscr{D}_0^\dagger(\mathscr{Q}R_{+1})],\end{aligned}\tag{50}$$

where

$$\begin{aligned}&g_{-1}(q)\equiv(1/B)(q\mathscr{D}_0^\dagger(\mathscr{Q}R_{+1})-\mathscr{Q}R_{+1}),\\&f_{-1}(p)\equiv(1/B)(p\mathscr{L}_1S_{+1}-\sqrt{\mathscr{P}}S_{+1}).\end{aligned}\tag{51}$$

[In the Kerr limit the functions $f_{\pm1}$ differ by a factor $(-a)$ from the expressions defined by Chandrasekhar[9,14] (see below).] The remaining components $\chi_{A'B'}$ (with the correct relative normalization) are obtained by substituting (46) into Eqs. (42) and (43). Using the Teukolsky–Starobinsky identities (45), this gives

$$\begin{aligned}&\bar{\chi}_0=B(1-pq)e^{i(ku+lv)}R_{+1}S_{-1},\\&\bar{\chi}_2=\tfrac{1}{2}B(1-pq)\bar{\phi}^2e^{i(ku+lv)}R_{-1}S_{+1}.\end{aligned}\tag{52}$$

Therefore, the component φ_1, missing in (46), must be given by the following two equivalent expressions:

$$\begin{aligned}\varphi_1=&(1/\sqrt{2})\phi^2e^{i(ku+lv)}\\&\times[g_{+1}(q)\mathscr{L}_1S_{+1}+if_{-1}(p)\mathscr{D}_0R_{-1}]\\=&-(1/\sqrt{2})\phi^2e^{i(ku+lv)}\\&\times[g_{-1}(q)\mathscr{L}_1^\dagger S_{-1}+if_{+1}(p)\mathscr{D}_0^\dagger(R_{+1})].\end{aligned}\tag{53}$$

From Eqs. (15), using (25), (26), (46), (49), and (51), one immediately obtains the following expressions for the components of the complex vector potential $X_{AB'}$, which corresponds to the field $\chi_{A'B'}$:

$$\begin{aligned}X_{00'}&=-(1/\sqrt{2})e^{i(ku+lv)}[iBf_{-1}R_{+1}+qR_{+1}\mathscr{L}_1S_{+1}],\\X_{01'}&=\tfrac{1}{2}\bar{\phi}e^{i(ku+lv)}[Bg_{-1}S_{+1}+ipS_{+1}\mathscr{D}_0^\dagger(R_{+1})],\\X_{10'}&=\tfrac{1}{2}\phi e^{i(ku+lv)}[Bg_{+1}S_{-1}+ipS_{-1}\mathscr{D}_0R_{-1}],\\X_{11'}&=(1/2\sqrt{2})\phi\bar{\phi}e^{i(ku+lv)}[iBf_{+1}R_{-1}+qR_{-1}\mathscr{L}_1^\dagger S_{-1}].\end{aligned}\tag{54}$$

The (complex) potential given by

$$\begin{aligned}\Phi_{00'}&=(1/\sqrt{2})e^{i(ku+lv)}[if_{+1}R_{+1}-B^{-1}qR_{+1}\mathscr{L}_1^\dagger S_{-1}],\\\Phi_{01'}&=\tfrac{1}{2}\bar{\phi}e^{i(ku+lv)}[g_{+1}S_{+1}-iB^{-1}pS_{+1}\mathscr{D}_0R_{-1}],\\\Phi_{10'}&=\tfrac{1}{2}\phi e^{i(ku+lv)}[g_{-1}S_{-1}-iB^{-1}pS_{-1}\mathscr{D}_0^\dagger(\mathscr{Q}R_{+1})],\\\Phi_{11'}&=-(1/2\sqrt{2})\phi\bar{\phi}e^{i(ku+lv)}\\&\quad\times[if_{-1}R_{-1}-B^{-1}qR_{-1}\mathscr{L}_1S_{+1}]\end{aligned}\tag{55}$$

generates the field φ_{AB}. Since, by analogy with (13), $\Phi_{AB'}$ satisfies $\nabla_{A(A'}\Phi^A_{\ B')}=0$, the vector potential (55) plus its Hermitian conjugate is a real vector potential that also generates the field φ_{AB}. Any other real vector potential with this property differs (locally) from this real potential by a gauge transformation. (The vector potential found in the case of the Kerr metric in Refs. 9 and 14 is not real and no explicit procedure is given to get a real vector potential.)

D. The Kerr limit

As indicated in Refs. 7 and 8, the Kerr metric can be exhibited as a limiting case of the Plebański–Demiański metric by making the substitutions

$$\begin{aligned}&q\to r,\quad p\to -a\cos\theta,\quad u\to -t+a\varphi,\quad v\to\varphi/a,\\&\mathscr{P}\to a^2\sin^2\theta,\quad \mathscr{Q}\to r^2-2Mr+a^2,\quad 1-pq\to 1,\end{aligned}$$

thus

$$\begin{aligned}&\phi\to(r-ia\cos\theta)^{-1},\quad \partial_q\to\partial_r,\\&\partial_p\to(a\sin\theta)^{-1}\partial_\theta,\quad \partial_u\to-\partial_t,\\&\partial_v\to a\,\partial_\varphi+a^2\,\partial_t,\quad k\to-\omega,\quad l\to a(m+a\omega).\end{aligned}$$

Therefore the Starobinsky constant is given in this case by

$$B^2=A^2-4\omega^2(a^2+am/\omega).$$

(Compare with Ref. 14.)

E. Characterization of the separable solutions

Form the derivation presented above it follows that if one considers a pair of fields φ_{AB} and $\tilde{\varphi}_{A'B'}$ that satisfies the (source-free) Maxwell equations (though these fields need not be related to one another), then the pair of fields $\tilde{\chi}_{AB}$ and $\chi_{A'B'}$, given by

$$\begin{bmatrix} \tilde{\chi}_{AB} \\ \chi_{A'B'} \end{bmatrix} = \mathscr{K} \begin{bmatrix} \varphi_{AB} \\ \tilde{\varphi}_{A'B'} \end{bmatrix}$$

$$\equiv \frac{1}{2} \begin{bmatrix} \nabla^{R'}_{(A}\bar{\phi}^{-2}\nabla^{S'}_{B)}\bar{\phi}^{2}\bar{L}_{R'C'}\bar{L}_{S'D'}\tilde{\varphi}^{C'D'} \\ \nabla^{R}_{(A'}\phi^{-2}\nabla^{S}_{B')}\phi^{2}L_{RC}L_{SD}\varphi^{CD} \end{bmatrix}, \tag{56}$$

is also a solution of the Maxwell equations; and the separable solutions in the Plebański–Demiański background, with a dependence in the variables u and v of the form $e^{i(ku+lv)}$, can be normalized in such a way that they satisfy the eigenvalue equation

$$\mathscr{K} \begin{bmatrix} \varphi_{AB} \\ \tilde{\varphi}_{A'B'} \end{bmatrix} = B \begin{bmatrix} \varphi_{AB} \\ \tilde{\varphi}_{A'B'} \end{bmatrix}, \tag{57}$$

where B is the Starobinsky constant (47). [Compare with the Appendix.] [The two equations contained in (56) are redundant as one is the complex conjugate of the other; however, by using the field φ_{AB} only, an eigenvalue equation similar to (57) would require a fourth-order differential operator or the use of complex conjugation that would modify the dependence in the variables u and v. This pairing arises more naturally in the case of spin-$\frac{1}{2}$ massless fields using Dirac's notation.]

V. CONCLUDING REMARKS

In the case of (conformally) flat space-time, Eq. (1) is known as the twistor equation (see, e.g., Ref. 19) and the operations given by Eqs. (56) and (A5) correspond to the spin-raising and spin-lowering operations induced by twistors (cf., also, Ref. 20). Since in a type-D space-time the one-index Killing spinors are not allowed, these shifts of the spin of the fields must be of at least one unit. However, for a solution φ_{AB} of Maxwell's equations, the combination $L_{AB}\varphi^{AB}$, which would satisfy the conformally invariant spin-0 massless field equation in (conformally) flat space-time, does not satisfy this equation in a type-D background.

Using Eqs. (3) one finds that $\varphi_0 = 0 = \varphi_2$, $\varphi_1 = \phi^2$ is a solution of the Maxwell equations (5), which is not of the form given by Eqs. (46) and (53). Nevertheless, this solution, together with that given by $\tilde{\varphi}_0 = 0 = \tilde{\varphi}_2$, $\tilde{\varphi}_1 = \bar{\phi}^2$, satisfies Eq. (57) with $B = 0$. This particular solution plays a relevant role because the electromagnetic field of the type-D solutions of the Einstein–Maxwell equations such that the principal null directions of the electromagnetic field coincide with those of the conformal curvature (such as the Kerr–Newman solution) has this form.

The results presented here are not applicable to the case of the solutions of the Einstein–Maxwell equations since the background electromagnetic field couples the electromagnetic and the gravitational perturbations. At least in the case of the perturbations of the Reissner–Nordström solution, which can be analyzed by separation of variables, there exist relations similar to the Teukolsky–Starobinsky identities that relate four radial functions[21] (the angular functions correspond to the spin-weighted spherical harmonics that also satisfy similar relations).

ACKNOWLEDGMENTS

The author would like to thank Professor Roger Penrose for the kind hospitality provided at The Mathematical Institute, University of Oxford.

The author acknowledges the financial support received from the Foreign and Commonwealth Office through The British Council and from the Sistema Nacional de Investigadores (Mexico).

APPENDIX: THE SOLUTION OF THE SPIN-$\frac{1}{2}$ MASSLESS FIELD EQUATIONS

Using the notation and the null tetrad of Sec. IV one easily finds that the spin-$\frac{1}{2}$ massless field equations $\nabla^{AB'}\eta_A = 0$, which amount to

$$\begin{aligned} &(D + \epsilon - \rho)\eta_1 - (\bar{\delta} - \alpha + \pi)\eta_0 = 0, \\ &(\delta + \beta - \tau)\eta_1 - (\Delta - \gamma + \mu)\eta_0 = 0 \end{aligned} \tag{A1}$$

in the Plebański–Demiański background, admit the separable solution

$$\begin{aligned} \eta_0 &= -(1 - pq)e^{i(ku+lv)}R_{+1/2}(q)S_{+1/2}(p), \\ \eta_1 &= (1/\sqrt{2})(1 - pq)\phi e^{i(ku+lv)}R_{-1/2}(q)S_{-1/2}(p), \end{aligned} \tag{A2}$$

with

$$\begin{aligned} &\mathscr{D}_0 R_{-1/2} = CR_{+1/2}, \quad \mathscr{L}_{1/2}S_{+1/2} = -CS_{-1/2} \\ &\mathscr{Q}^{1/2}\mathscr{D}_0^\dagger \mathscr{Q}^{1/2}R_{+1/2} = CR_{-1/2}, \quad \mathscr{L}^\dagger_{1/2}S_{-1/2} = CS_{+1/2}, \end{aligned} \tag{A3}$$

where C is a real separation constant (the fact that C is real can be derived as in Sec. IV, using that $R_{-1/2}$ and $\mathscr{Q}^{1/2}R_{+1/2}$ satisfy complex-conjugate equations). These equations are the analog of the Teukolsky–Starobinsky identities (45) and the first of Eqs. (A1) is, in this sense, analogous to Eq. (18). The separable solutions of the equations $\nabla^{BA'}\tilde{\eta}_{A'} = 0$ are given by

$$\begin{aligned} \tilde{\eta}_{0'} &= (1 - pq)e^{i(ku+lv)}R_{+1/2}S_{-1/2}, \\ \tilde{\eta}_{1'} &= (1/\sqrt{2})(1 - pq)\bar{\phi}e^{i(ku+lv)}R_{-1/2}S_{+1/2}. \end{aligned} \tag{A4}$$

By combining Eqs. (A3) one obtains a decoupled second-order equation for each separated function with C^2 appearing as eigenvalue. Hence, in this case, a combination like (47) does not arise.

One can verify, in general, that the spinor fields $\tilde{\xi}_A$ and $\xi_{A'}$ defined by

$$\begin{bmatrix} \tilde{\xi}_A \\ \xi_{A'} \end{bmatrix} = \mathscr{K} \begin{bmatrix} \eta_A \\ \tilde{\eta}_{A'} \end{bmatrix} \equiv \frac{1}{\sqrt{2}} \begin{bmatrix} \bar{\phi}^{-1}\nabla^{B'}_A(\bar{\phi}\bar{L}_{B'C'}\tilde{\eta}^{C'}) \\ -\phi^{-1}\nabla^{B}_{A'}(\phi L_{BC}\eta^C) \end{bmatrix} \tag{A5}$$

satisfy the spin-$\frac{1}{2}$ massless field equations provided η_A and $\tilde{\eta}_{A'}$ are solutions of the spin-$\frac{1}{2}$ massless field equations and L_{AB} is a two-index Killing spinor.[15] Then, with η_A and $\tilde{\eta}_{A'}$ given by (A2) and (A4), Eq. (A3) is equivalent to the eigenvalue equation

$$\mathscr{K} \begin{bmatrix} \eta_A \\ \tilde{\eta}_{A'} \end{bmatrix} = C \begin{bmatrix} \eta_A \\ \tilde{\eta}_{A'} \end{bmatrix}. \tag{A6}$$

[The opposite signs appearing in the definition of $\mathscr{K}$ (A5) are essential. Using Dirac's γ matrices these signs can be taken into account by means of γ_5 (cf. Refs. 3 and 12).]

[1]M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970).
[2]L. P. Hughston, R. Penrose, P. Sommers, and M. Walker, Commun. Math. Phys. **27**, 303 (1972).
[3]B. Carter and R. G. McLenaghan, Phys. Rev. D **19**, 1093 (1979).
[4]S. Chandrasekhar, Proc. R. Soc. London Ser. A **349**, 571 (1976).
[5]R. Güven, Proc. R. Soc. London Ser. A **356**, 465 (1977).
[6]S. A. Teukolsky, Astrophys. J. **185**, 635 (1973).
[7]A. L. Dudley and J. D. Finley, III, J. Math. Phys. **20**, 311 (1979).
[8]J. F. Plebański and M. Demiański, Ann. Phys. (NY) **98**, 98 (1976).
[9]S. Chandrasekhar, Proc. R. Soc. London Ser. A **349**, 1 (1976).
[10]S. Chandrasekhar, Proc. R. Soc. London Ser. A **358**, 421 441 (1978).
[11]P. L. Chrzanowski, Phys. Rev. D **11**, 2042 (1975).
[12]N. Kamran and R. G. McLenaghan, Lett. Math. Phys. **7**, 381 (1983); J. Math. Phys. **25**, 1019 (1984).
[13]B. P. Jeffryes, Proc. R. Soc. London Ser. A **392**, 323 (1984).
[14]S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Clarendon, Oxford, 1983).
[15]G. F. Torres del Castillo, Proc. R. Soc. London Ser. A **400**, 119 (1985).
[16]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962); **4**, 998 (1963).
[17]R. Penrose and W. Rindler, *Spinors and Space-Time* (Cambridge U.P., Cambridge, 1984), Vol. 1.
[18]R. Geroch, A. Held, and R. Penrose, J. Math. Phys. **14**, 874 (1973).
[19]R. Penrose and W. Rindler, *Spinors and Space-Time* (Cambridge U.P., Cambridge, 1986), Vol. 1.
[20]G. F. Torres del Castillo, J. Math. Phys. 27, 1583 (1986).
[21]G. F. Torres del Castillo, "Perturbations of solutions of the Einstein equations with sources," Class. Quantum Gravit. (in press).

The hyperspin structure of unitary groups

Christian Holm
School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 26 October 1987; accepted for publication 25 November 1987)

A sequence of homogeneous, isotropic, and compact solutions to the empty space hypergravity equations [D. Finkelstein, S. R. Finkelstein, and C. Holm, Phys. Rev. Lett. **59**, 1265 (1987); S. R. Finkelstein, Ph.D. thesis, Georgia Institute of Technology, 1987 (unpublished)] is studied. These solutions are hyperspin manifolds constructed from unitary groups $\mathbf{U}_N := \mathrm{U}(N,C)$ and describe static compact Einstein universes of the Kaluza–Klein type with Finslerian geometry. By taking the universal covering group of $\mathbf{U}_N$ an N^2-dimensional manifold with topology $\mathbb{R}\times \mathbf{SU}_N$ is obtained, where the Abelian group $\mathbb{R}$ is the time axis and $\mathbf{SU}_N$ provides a compact homogeneous spatial part. To describe the Finslerian geometry of these manifolds the Cartan–Penrose exterior calculus is extended from spinors to hyperspinors and the Maurer–Cartan equations are applied to obtain the hyperspin structure of $\mathbf{U}_N$. Hyperspin geometry seems to be a consistent alternative to the usual Riemannian geometry for spaces with internal dimensions, which deserves further study.

I. INTRODUCTION

In 1986 Finkelstein[1] proposed unifying two seemingly contradictory principles, the *hyper*space (i.e., Kaluza–Klein models) and the *spin*or principle, into what he called *hyperspin*. The apparent contradiction rested upon the fact that the underlying spinor space was built from two-component spinors,[2] leading to a time-space of dimension $2^2 = 4$, whereas the hyperspace principle makes us believe that the world has dimension $d > 4$. The solution was to enlarge the number of components of the spinors to N (hence, *hyper*spinors), which leads to causal time spaces of dimension $n = N^2$. The transformation group of the hyperspinors is taken to be $\mathbf{SL}_N := \mathrm{SL}(N,C)$ and not the (pseudo-) orthogonal groups of ordinary spinors. The manifolds with a hyperspin structure are denoted by $\mathbf{B}_N$ and called Bergmann manifolds after Bergmann[3] who first took the spin manifold (the case $N=2$) as fundamental and showed that general relativity emerges from it. The geometry of Bergmann manifolds is in general Finslerian and coincides only for $N=2$ with Riemannian geometry. The dimensionality of $\mathbf{B}_N$ is restricted to be a square of an integer, four being the first interesting manifold after the case of pure time ($N=1$). The first two components (in a suitable frame) of a general hyperspinor are supposed to give rise to the (external) four-dimensional time-space we live in, whereas the other components are free to build internal spaces, from which gauge groups could arise.

In this paper we study a series of hyperspin manifolds based on the unitary groups $\mathbf{U}_N := \mathrm{U}(N,C)$, which describe static, homogeneous, isotropic, and compact Kaluza–Klein cosmologies. The reason for studying a group manifold is its great symmetry, which makes calculations easier, and also gives rise to unitary symmetry groups acting on the internal dimensions. In the course of this work it also turned out that the unitary groups are solutions to the hypergravity field equations.[4]

The paper is organized as follows. In Sec. II we recapitulate the elementary ingredients for the hyperspin theory.

In Sec. III we determine the hyperspin structure of $\mathbf{U}_N$. We do so by performing calculations on the complexified $\mathbf{U}_N$, which is $\mathbf{GL}_N := \mathrm{GL}(N,C)$. We introduce the idea of a complexified spin structure. As a simple example we treat $\mathbf{U}_2$ in detail. At the end of this section we prove that $\mathbf{U}_N$ satisfies the hypergravitational field equations.

The sometimes cumbersome calculations lead us to generalize Cartan's calculus of differential forms to hyperspinors in Sec. IV. We invent a method that allows us to read off the hyperspin structure of suitable Lie groups from the Mauer–Cartan equations.

In Sec. V we conclude the work with a summary and discussion of the results.

II. THE HYPERSPIN CONCEPT

The notation we use is designed to provide as much information as possible with a minimum of "index gymnastics." We use capital Greek letters to denote spinor indices, $\Sigma, \Omega,... = 1,...,N$. A dot over a spinor index means that it belongs to an antispinor and transforms according to the complex conjugate representation, where $\dot{\Sigma}$ and Σ are treated as independent variables. Lowercase Greek letters denote the time-space manifold and run from $1,...,N^2$.

We also introduce collective indices for subscripts and superscripts. To cut down on the number of indices we write $(\Sigma) := \Sigma_1,...,\Sigma_N$ as standing for N spinor indices. The kind of parentheses used indicates the symmetry: $[\Sigma]$ for N antisymmetrized indices, $\{\Sigma\}$ for symmetrized ones, and (Σ) meaning that there is no special symmetry. The number of primes on a collective index gives the number of omitted indices. The collective tensor product for a tensor $\sigma(\alpha,\Sigma,\dot{\Sigma})$ is defined by

$$\Pi\sigma(\alpha,\Sigma,\dot{\Sigma}) := \sigma(\alpha_1,\Sigma_1,\dot{\Sigma}_1),...,\sigma(\alpha_N,\Sigma_N,\dot{\Sigma}_N).$$

The comma and semicolon before an index mean the ordinary and covariant derivative, respectively. We also make use of the Einstein summation convention throughout.

A. Hyperspace geometry

Hyperspinors ψ^{Σ} are defined on a complex linear space $\mathbf{C}^{\Sigma}$ and transform according to the defining representation of

$\mathbf{SL}_N$. Antispinors $\overline{\psi}^{\dot\Sigma}$ belong to the complex conjugate space $\overline{\mathbf{C}}^{\dot\Sigma}$ and transform under the complex conjugate representation. Let $\mathbf{R}^\sigma$ be the real linear subspace of $\mathbf{C}^\Sigma \otimes \overline{\mathbf{C}}^{\dot\Sigma}$ whose elements are Hermitian and called *sesquispinors*. Due to Hermiticity the dimensionality of $\mathbf{R}^\sigma$ is N^2. A Bergmann manifold $\mathbf{B}_N$ is defined as a differentiable manifold that is locally a $\mathbf{R}^\sigma$ and possesses a differentiable spin map $\sigma(u)$ at each point u of $\mathbf{B}_N$.

Each point u carries a spin vector $\sigma^\alpha{}_{\dot\Sigma\Sigma}(u)$ that is a sufficiently differentiable local isomorphism from $\mathbf{R}^\sigma$ to $d\,\mathbf{B}_N(u)$, which denotes the tangent space of $\mathbf{B}_N$ at u. For $N=2$ σ is also known as the soldering form or the Infeld–van der Waerden symbol. The inverse map $\tilde\sigma$ is called the spin form and has the index structure $\tilde\sigma_\alpha{}^{\Sigma\dot\Sigma}(u)$. It obeys the inverse relations

$$\sigma^\alpha{}_{\dot\Sigma\Sigma}\tilde\sigma_\alpha{}^{\Omega\dot\Omega} = \delta^\Omega{}_\Sigma \delta^{\dot\Omega}{}_{\dot\Sigma} \quad\text{and}\quad \sigma^\alpha{}_{\dot\Sigma\Sigma}\tilde\sigma_\beta{}^{\Sigma\dot\Sigma} = \delta^\alpha{}_\beta. \tag{2.1}$$

The spin vector is the basic dynamical variable in the theory. In $\mathbf{B}_2$ it is the square root of Einstein's metric tensor g, whereas for $N>2$ the extra dimensions it gives rise to could account for gauge fields of the other known forces. In flat space σ is globally defined and can be taken to be the identity map. In the overcomplete non-Hermitian basis $\epsilon_\Omega \otimes \epsilon_{\dot\Omega}$ for the tangent space this spin vector may then be written as $\delta^\Omega{}_\Sigma \delta^{\dot\Omega}{}_{\dot\Sigma}$. This means that the action of σ on a sesquispinor $h^{\Sigma\dot\Sigma}$ is

$$\sigma:\ h^{\Sigma\dot\Sigma} \mapsto \delta^{\dot\Omega}{}_{\dot\Sigma}\delta^\Omega{}_\Sigma h^{\Sigma\dot\Sigma} = h^{\Omega\dot\Omega}.$$

The result $h^{\Omega\dot\Omega}$ is a Hermitian matrix, which is regarded as a tangent vector. The use of an overcomplete basis should not confuse the reader and will be used throughout this work, as it greatly simplifies the formulas. The spin form defines also the chronometric density ρ via

$$\rho = \det(\tilde\sigma_{\alpha\beta}) = K\det(\tilde\sigma^{\Omega\dot\Omega\Sigma\dot\Sigma}),$$

where α is the sesquispinor index and β is the time-space index of an Hermitian basis $\epsilon_\beta{}^{\Omega\dot\Omega}$. The constant K on the right-hand side can be used to normalize the Hermitian bases $\epsilon_\alpha, \epsilon_\beta$. This ρ is the density used to construct invariant actions.

The action of the covariant derivative D on tensors is defined in the usual fashion and induces the vector connection. The spin connection is defined by the following action of D on spinors:

$$D_\alpha(\psi^\Sigma + \overline{\chi}^{\dot\Sigma}) = \partial_\alpha(\psi^\Sigma + \overline{\chi}^{\dot\Sigma}) + \Gamma_\alpha{}^\Sigma{}_{\Sigma'}\psi^{\Sigma'} + \overline{\Gamma}_\alpha{}^{\dot\Sigma}{}_{\dot\Sigma'}\overline{\chi}^{\dot\Sigma'}. \tag{2.2}$$

The extension to cospinors and spinors of arbitrary valence (polyspinors) can be paralleled exactly to the treatment of Ref. 2 and was outlined already in Ref. 5.

Here, $D\sigma$ is called the torque tensor and $D\sigma=0$ is called torque-freeness. This is an important assumption, because it enables us to express the vector connection C uniquely in terms of Γ, σ and derivatives of σ:

$$\begin{aligned} D_\alpha\sigma^\beta{}_{\dot\Sigma\Sigma} &= \partial_\alpha\sigma^\beta{}_{\dot\Sigma\Sigma} - \sigma^\beta{}_{\dot\Sigma\Sigma'}\Gamma_\alpha{}^{\Sigma'}{}_\Sigma - \Gamma^H{}_{\alpha\dot\Sigma}{}^{\dot\Sigma'}\sigma^\beta{}_{\dot\Sigma'\Sigma} \\ &\quad + C^\beta{}_{\alpha\gamma}\sigma^\gamma{}_{\dot\Sigma\Sigma} = 0 \\ \Rightarrow C^\beta{}_{\alpha\gamma} &= -Tp((\partial_\alpha\sigma^\beta)\cdot\tilde\sigma_\gamma) \\ &\quad + Tp(\sigma^\beta\cdot\Gamma_\alpha\cdot\tilde\sigma_\gamma) + Tp(\tilde\sigma_\gamma\cdot\Gamma^H{}_\alpha\cdot\sigma^\beta), \end{aligned} \tag{2.3}$$

where Tp stands for the trace operation over the spinor indices. We will make the assumption of torque-freeness throughout.

As another requirement we demand that D annihilates the Levi-Civita ϵ spinor, which has the index structure $\epsilon_{[\Sigma]}$ and is completely defined by $\epsilon_{12\ldots N}=1$. This will make the spin connection traceless in the spinor indices:

$$D_\alpha\epsilon_{[\Sigma]} = \partial_\alpha\epsilon_{[\Sigma]} - \sum \Gamma_\alpha{}^\Omega{}_\Sigma\epsilon_{[\Omega]} = 0 \Rightarrow \Gamma_\alpha{}^\Sigma{}_\Sigma = 0.$$

The chronometric norm of a tangent vector du is defined as

$$\begin{aligned} \|du\| &:= g_{\{\alpha\}}\,du^{\{\alpha\}} \\ &:= \frac{1}{(N-1)!}\epsilon_{[\Sigma]}\epsilon_{[\dot\Sigma]}\prod\tilde\sigma^{(\Sigma)(\dot\Sigma)}{}_{(\alpha)}\,du^{(\alpha)} \\ &= N\det(\tilde\sigma_\alpha\,du^\alpha). \end{aligned} \tag{2.4}$$

This equation serves also to define the N-index totally symmetric chronometric form $g_{\{\alpha\}}$, and its dual is defined according to

$$g^{\{\alpha\}} := \frac{1}{(N-1)!}\epsilon^{[\Sigma]}\epsilon^{[\dot\Sigma]}\prod\sigma^{(\alpha)}{}_{(\dot\Sigma)(\Sigma)} \tag{2.5}$$

in order to preserve the usual inverse relation

$$g^{\{\alpha'\}\beta}g_{\{\alpha'\}\gamma} = \delta^\beta{}_\gamma. \tag{2.6}$$

The Nth root of the norm defines the proper time

$$d\tau^N = \|du\|. \tag{2.7}$$

Future vectors are defined to be the σ image of positive definite sesquispinors, thus defining a causal structure of all $\mathbf{B}_N$. The lack of a quadratic metric for $N>2$ makes $\mathbf{B}_N$ a Finsler space. Generalized metrics with N indices have been considered before,[6] but to our knowledge nobody ever used the determinantal form. For a recent comprehensive work on Finsler spaces see Ref. 7.

The departure from Riemannian geometry leads to several new features. For example, now the natural orthogonality relation (not found in earlier works on Finsler spaces) is a relation between N vectors rather than two vectors. The raising and lowering of indices with g is modified as well. The "dual" D to a (co)vector u is now a $N-1$ index covector (vector) and of order $N-1$ in u:

$$(u^D)_{\{\alpha'\}} := u^\beta g_{\beta\{\alpha'\}}.$$

In Riemannian geometry there exists a unique vector connection C, the Levi-Civita connection, defined by the absence of torsion and the metricity condition $D_\alpha g_{\beta\gamma}=0$, which leads to the familiar Christoffel formula $C^\gamma{}_{\alpha\beta} = \frac{1}{2}g^{\gamma\delta}(g_{\delta\alpha,\beta} + g_{\delta\beta,\alpha} - g_{\alpha\beta,\delta})$.

Surprisingly, the same two requirements provide the hyperspin geometry as well with a unique vector connection[8] C. The metricity condition

$$D_\alpha g_{\{\beta\}} = 0 \tag{2.8}$$

can no longer be solved by cyclic permutation. The hyper-Christoffel C greatly resembles the Levi-Civita connection with 2 replaced by N:

$$C^\gamma{}_{\alpha\beta} = (1/N)g^{\gamma\lambda\{\delta''\}} \times(\partial_\alpha g_{\beta\lambda\{\delta''\}} + \partial_\beta g_{\alpha\lambda\{\delta''\}} - \partial_\lambda g_{\alpha\beta\{\delta''\}}). \tag{2.9}$$

The verification of the above formula is by direct substitution in (2.8).

Torque-freeness (and therefore metricity) and torsion-freeness determine a unique spin connection Γ. This is easy to see, because (2.9) gives us a unique vector connection C and solving (2.3) for Γ determines thus the spin connection uniquely,

$$\Gamma_\alpha{}^\Sigma{}_{\Sigma'} = (1/N)\{C^\delta{}_{\alpha\beta}\sigma^\beta{}_{\dot\Sigma\Sigma'}\tilde\sigma_\delta{}^{\Sigma\dot\Sigma} - (\partial_\alpha\tilde\sigma_\beta{}^{\Sigma\dot\Sigma})\sigma^\beta{}_{\dot\Sigma\Sigma'}\}. \tag{2.10}$$

A lengthy calculation involving the hyper Christoffel formula (2.9) and the definition of the chronometric (2.4) gives Γ entirely in terms of σ and its derivative $\partial\sigma$:

$$\begin{aligned}\Gamma_\alpha{}^\Sigma{}_{\Sigma'} = {} & (-1/N)[(\partial_\alpha\tilde\sigma_\beta)\cdot\sigma^\beta]^\Sigma{}_{\Sigma'} \\ & - [\delta^\Sigma{}_{\Sigma'}/N^2(N-1)^2]\,T\rho[N\sigma^\delta\cdot\partial_\delta\tilde\sigma_\alpha \\ & - N(N-2)\sigma^\delta\cdot\partial_\alpha\tilde\sigma_\delta - \sigma^\delta\cdot\tilde\sigma_\alpha\cdot\sigma^\beta\cdot\partial_\delta\tilde\sigma_\beta \\ & - \sigma^\beta\cdot\tilde\sigma_\alpha\cdot\sigma^\delta\cdot\partial_\delta\tilde\sigma_\beta] - [1/N(N-1)^2] \\ & \times[(\partial_\delta\tilde\sigma_\beta)\cdot\sigma^\beta\cdot\tilde\sigma_\alpha\cdot\sigma^\delta + \tilde\sigma_\alpha\cdot\sigma^\beta\cdot(\partial_\delta\tilde\sigma_\beta)\cdot\sigma^\delta \\ & - N(\partial_\delta\tilde\sigma_\alpha)\cdot\sigma^\delta - T\rho(\sigma^\beta\cdot\partial_\delta\tilde\sigma_\beta)\tilde\sigma_\alpha\cdot\sigma^\delta]^\Sigma{}_{\Sigma'}.\end{aligned} \tag{2.11}$$

In this notation matrix multiplication is used to suppress the spinor indices.

B. Hypergravity

Treating the spin vector σ as a dynamical variable will lead to a field theory called hypergravity.[1] In Ref. 4 a one-parameter family of actions is discussed, which reproduces in the case $N=2$ Einstein's theory of gravity. The spin vector does not appear explicitly in any invariant action, but only through the chronometric g.

The most natural action scalar is obtained by covariantly differentiating the Ricci tensor $N-2$ times and contracting the result with g,

$$R := R_{\alpha\beta;(\gamma'')}\,g^{\alpha\beta\{\gamma''\}}.$$

Note that for $N=2$ R is the usual curvature scalar. A variation of this action leads to the following equations of motion:

$$R^\alpha{}_\beta - (1/N)\delta^\alpha{}_\beta R = T^\alpha{}_\beta, \tag{2.12}$$

where $R^\alpha{}_\beta$ is the mixed Ricci tensor defined by

$$R^\alpha{}_\beta := g^{\alpha\gamma\{\delta''\}}R_{\beta(\gamma;\delta'')},$$

$T^\alpha{}_\beta$ is the conserved energy-momentum tensor defined by

$$T^\alpha{}_\beta := -(1/N\rho)\sigma^\alpha{}_{\dot\Sigma\Sigma}[\delta(\rho L')/\delta\sigma^\beta{}_{\dot\Sigma\Sigma}],$$

and L' is the Lagrangian of the matter fields present. The formal similarity to Einstein's equations is very remarkable, indeed.

III. THE HYPERSPIN STRUCTURE OF $\mathbf{U}_N$

In this section we will build a toy model for a Kaluza–Klein time-space based on the $\mathbf{SL}_N$ tangent space group using the hyperspin construction. One reason to consider the unitary groups $U(N,C) =: \mathbf{U}_N$ as group manifolds is the fact that the dimension works out right. $\mathbf{U}_N$ has N^2 real dimensions, exactly what we need for a manifold built from N-component hyperspinors. Another reason is that a group manifold has a great deal of symmetry, which allows analytic computations.

A. Construction

The unitary groups describe homogeneous, isotropic, compact spaces. The group structure of $\mathbf{U}_N$ is $(\mathbf{U}_1\times\mathbf{SU}_N)/\mathbf{Z}_N$, where $\mathbf{Z}_N$ denotes the cyclic group of order N. The $\mathbf{U}_1$ part is the natural candidate for the time axis, and to avoid the cyclicity in the time direction we look specifically at the universal covering group of $\mathbf{U}_N$, denoted by Cov $\mathbf{U}_N$, which is the unique simply connected covering group of $\mathbf{U}_N$. Cov $\mathbf{U}_N$ is topologically $\mathbf{R}\times\mathbf{SU}_N$. Because R commutes with the spatial part $\mathbf{SU}_N$, the dynamics is that of a static universe. We call our $\mathbf{U}_N$ manifolds Einstein universes, because they are generalizations of the special case $N=2$, which gives the static Einstein universe $\mathbf{R}\times S^3$.

A third reason for studying $\mathbf{U}_N$ is that they are cosmological solutions to the hypergravitational field equations (2.12), although the equations were obtained well after this study of $\mathbf{U}_N$ started. The details of this aspect are treated at the end of this section.

One disadvantage of $\mathbf{U}_N$ is the fact that all space dimensions are treated on the same footing, so that the internal dimensions are of the order of the external ones in size. A dimensional reduction is necessary to give the model an actual physical meaning.

To define the unitary group we will use an embedding of $\mathbf{U}_N$ in $\mathbf{C}^\Sigma{}_{\Sigma'}$, $\Sigma,\Sigma' = 1,\dots,N$, which is the complex linear space of $N\times N$ matrices. As a convenient basis in $\mathbf{C}^\Sigma{}_{\Sigma'}$ we use the matrix units $E^{\Sigma'}{}_\Sigma$, which are $N\times N$ matrices with a one in row Σ and column Σ' and 0 otherwise. We write the general element $z\in\mathbf{C}^\Sigma{}_{\Sigma'}$ as $z = z^\Sigma{}_{\Sigma'}E^{\Sigma'}{}_\Sigma$ and the unitary element $u\in\mathbf{U}_N$ as $u = u^\Sigma{}_{\Sigma'}E^{\Sigma'}{}_\Sigma$. We provide $\mathbf{C}^\Sigma{}_{\Sigma'}$ with a fixed Hermitian, positive definite metric $\mu\in\mathbf{R}_\sigma$, where $\mathbf{R}_\sigma$ denotes the dual space to $\mathbf{R}^\sigma$. The metric has the index structure $\mu_{\dot\Sigma\Sigma}$ and corresponds to a global future timelike vector field, obtained through the spin map.

We use the metric to define unitary operators $u\in\mathbf{U}_N$ via the constraint equation

$$u^H\mu u = \mu.$$

The tangent vectors du at u therefore obey the constraints

$$du^H\,\mu u = -u^H\mu\,du. \tag{3.1}$$

To specify the spin structure we have to give the spin map $\sigma(u,du)$, which is a local isomorphism from the space of sesquispinors $\mathbf{R}^\sigma$ to the vector space of tangent vectors. We make the spin map satisfy two requirements:

(1) it has to be Hermitian in its spinor indices, i.e., $\sigma(u,du) = \sigma^H$;

(2) it has to be left invariant.

(1) follows from the definition of a spin map, while (2) will result in a homogeneous time space. Right invariance would also accomplish this purpose.

We start by defining not σ but the spin form $\tilde\sigma := \sigma^{-1}$, which is the inverse of the spin map: $d\,\mathbf{U}_N(u)\to\mathbf{R}^\sigma$. The spin form maps each tangent vector $du(u)$ at a point u into a sesquispinor $d\psi\in\mathbf{R}^\sigma$.

From (3.1) we see that left multiplication of du by $u^H\mu$, denoted by $\urcorner(u^H\mu)$, produces an anti-Hermitian matrix. To obtain the correct index structure we need to raise the lower indices, which can be done by right multiplication with $\tilde{\mu}$, denoted by $\ulcorner(\tilde{\mu})$, and an additional left multiplier of $\tilde{\mu}$. The expression $\tilde{\mu}u^H\mu$ simplifies further to u^{-1}, which we denote by $\tilde{u}$. Therefore we take the definition of $\tilde{\sigma}$ to be

$$\tilde{\sigma}(u) := -i\urcorner(\tilde{\mu}u^H\mu)\ulcorner(\tilde{\mu}) = -i\urcorner(\tilde{u})\ulcorner(\tilde{\mu}). \tag{3.2}$$

The spin vector is obtained by inverting this relation,

$$\sigma(u) := \tilde{\sigma}(u)^{-1} = i\urcorner(u)\ulcorner(\mu). \tag{3.3}$$

The spin map is left invariant, i.e., a group transformation of left multiplication by $u_0 \in \mathbf{U}$, $\urcorner u_0 : u \mapsto u' = u_0 u$, leaves the spin map unchanged:

$$\tilde{\sigma}'(u') \cdot du' = -i(u_0 u)^{-1}(u_0\, du)\tilde{\mu} = \tilde{\sigma}(u) \cdot du.$$

The spin map is an operator on matrices, and whenever it is necessary we will use the complex matrix elements as redundant coordinates, which means we use the matrix units as the basis for the tangent space. Let $e_\alpha{}^\Sigma{}_{\Sigma'}(u)$ be a basis for $d\,\mathbf{U}_N(u)$, i.e., a set of anti-Hermitian $N \times N$ matrices and let $\tilde{e}^{\beta\Omega'}{}_\Omega$ be the dual basis defined by

$$\tilde{e}^{\beta\Omega'}{}_\Omega e_\alpha{}^\Omega{}_{\Omega'} = \delta^\beta{}_\alpha.$$

The matrix unit coordinates of a tangent vector $v(u)$ are simply given by the matrix elements of e_α,

$$v^\alpha e_\alpha{}^\Sigma{}_{\Sigma'}(u) = v^\Sigma{}_{\Sigma'}.$$

The $v^\Sigma{}_{\Sigma'}\tilde{\mu}^{\Sigma'\dot{\Sigma}} =: v^{\Sigma\dot{\Sigma}}$ are then anti-Hermitian in the sense $v^{\Sigma\dot{\Sigma}} = \overline{v}^{\dot{\Sigma}\Sigma}$. The spin map written out in matrix unit coordinates looks like

$$\begin{aligned} e_\beta \sigma^\beta{}_{\dot{\Sigma}\Sigma}(u) &=: \sigma^\Omega{}_{\Omega'\dot{\Sigma}\Sigma} = iu^\Omega{}_\Sigma \mu_{\dot{\Sigma}\Omega'}, \\ \tilde{e}^\beta \tilde{\sigma}_\beta{}^{\Sigma\dot{\Sigma}}(u) &=: \tilde{\sigma}^{\Omega'}{}_\Omega{}^{\Sigma\dot{\Sigma}} = -i\tilde{u}^\Sigma{}_\Omega \tilde{\mu}^{\Omega'\dot{\Sigma}}. \end{aligned} \tag{3.4}$$

When it is clear from the context we will use the convention that the index β stands for a pair of indices $\Omega, \Omega' = 1,\ldots,N$, without inserting the matrix $e_\beta{}^\Omega{}_{\Omega'}$ explicitly. As manifold indices they are in capital Greek letters. The position of the primed index shows if it is a covariant or contravariant index. For the following we will work completely with left invariant quantities, but for completeness we include also the definition for the right invariant spin map,

$$\tilde{\sigma}(u) = -i\ulcorner(\tilde{\mu}u^H) \quad \text{and} \quad \sigma(u) = i\ulcorner(\mu u)$$

and in index notation

$$\sigma^\Omega{}_{\Omega'\dot{\Sigma}\Sigma} = i\delta^\Omega{}_\Sigma(\mu u)_{\dot{\Sigma}\Omega'} \quad \text{and} \quad \tilde{\sigma}^{\Omega'}{}_\Omega{}^{\Sigma\dot{\Sigma}} = -i\delta^\Sigma{}_\Omega(\tilde{\mu}u^H)^{\Omega'\dot{\Sigma}}.$$

The physics of the geometry is of course independent of our decision to work with left invariant quantities instead of right invariant ones.

To specify the geometry of $\mathbf{U}_N$ completely, we have to calculate the spin connection form Γ, which fixes the covariant derivative D of spinors. We assume Γ to be torque-free and torsion-free, so that Γ is uniquely determined by the spin map through (2.11). Γ then induces a unique vector connection C and covariant derivative on vectors.

The spin connection form is a Lie algebra valued covector, mapping $d\,\mathbf{U}_N$ into $d\,\mathbf{S}_N$, the Lie algebra of the structural group[9] $\mathbf{S}_N := \mathrm{SL}(N,C)$. To determine Γ we demand the following requirements.

(1) $\Gamma_\beta{}^\Sigma{}_{\Sigma'}$ has to be traceless in its spinor indices, i.e., $\Gamma_\beta{}^\Sigma{}_\Sigma = 0$.

(2) Γ has to be left invariant in order to obtain a homogeneous space.

(3) Γ has to be anti-Hermitian in the sense that $\Gamma_\beta{}^\Sigma{}_{\Sigma'} = -(\tilde{\mu}\Gamma_\beta{}^H\mu)^\Sigma{}_{\Sigma'}$. This follows from the requirement that the connection form of a scalar is zero:

$$\begin{aligned} &D(\overline{\psi}\mu\psi) = \partial(\overline{\psi}\mu\psi) \\ &\quad\Rightarrow (\overline{\Gamma}^{\dot{\Sigma}'}{}_{\dot{\Sigma}}\mu_{\dot{\Sigma}'\Sigma} + \mu_{\dot{\Sigma}\Sigma'}\Gamma^{\Sigma'}{}_\Sigma)\overline{\psi}^{\dot{\Sigma}}\psi^\Sigma = 0 \\ &\quad\Rightarrow \Gamma = -\tilde{\mu}\Gamma^H\mu. \qquad \blacksquare \end{aligned}$$

Here Γ has to be a traceless anti-Hermitian spin operator on $d\,\mathbf{S}_N$. By our previous construction $\tilde{\sigma}$ is a Hermitian spin matrix. The metric μ relates such spin matrices to spin operators. We therefore make the assumption that Γ on $\mathbf{U}_N$ is given by

$$\Gamma = ik(\tilde{\sigma}\mu - (1/N)\mathbf{1}Tp(\tilde{\sigma}\mu)), \tag{3.5}$$

where k is for the moment an arbitrary real constant which we will fix later. In index notation this implies

$$\begin{aligned} \tilde{e}^{\beta\Omega'}{}_\Omega \Gamma_\beta{}^\Sigma{}_{\Sigma'}(u) &=: \Gamma^{\Omega'}{}_\Omega{}^\Sigma{}_{\Sigma'} \\ &= k(\tilde{u}^\Sigma{}_\Omega \delta^{\Omega'}{}_{\Sigma'} - (1/N)\delta^\Sigma{}_{\Sigma'}\tilde{u}^{\Omega'}{}_\Omega). \end{aligned}$$

The spin connection obviously satisfies requirements (1) and (2). To prove requirement (3) we apply Γ to a tangent vector du,

$$\begin{aligned} \{\Gamma_\beta{}^\Sigma{}_{\Sigma'} \cdot du^\beta\}^H &= k\{(\tilde{u}\,du)^\Sigma{}_{\Sigma'} - (1/N)\delta^\Sigma{}_{\Sigma'}Tp(\tilde{u}\,du)\}^H \\ &= -k\{(\mu\tilde{u}\,du\,\tilde{\mu})_{\dot{\Sigma}'}{}^{\dot{\Sigma}} \\ &\qquad - (1/N)\delta_{\dot{\Sigma}'}{}^{\dot{\Sigma}}Tp(\tilde{u}\,du)\} \\ &= -(\mu\Gamma_\beta\tilde{\mu})_{\dot{\Sigma}'}{}^{\dot{\Sigma}} \cdot du^\beta. \qquad \blacksquare \end{aligned}$$

In order to simplify the calculations, we introduce the idea of a complexified "spin structure." Following the previous notation we analytically continue $u \in \mathbf{U}_N$ to $z \in \mathbf{GL}_N$. Let $e_\alpha(u)$ be a basis for $d\,\mathbf{U}_N(u)$. Then

$$\{u^\alpha e_\alpha \mid u^\alpha \in \mathbf{R}\} = d\,\mathbf{U}_N(u)$$

and

$$\{z^\alpha e_\alpha \mid z^\alpha \in \mathbf{C}\} = d\,\mathbf{GL}_N(z).$$

Let $\varphi(u): \mathbf{U}_N \to \mathbf{R}$ be a test function on $\mathbf{U}_N$, $\hat{\varphi} \in C^\omega(u)$. There exists an extension $\hat{\varphi}(z): \mathbf{C}^\Sigma{}_{\Sigma'} \to \mathbf{C}$ that is unique and analytic in a neighborhood of $\mathbf{U}_N$. Then the real derivative

$$\frac{d}{du^\alpha}\varphi(u) = \left.\frac{d\hat{\varphi}(z)}{dz^\alpha}\right| \mathbf{U}.$$

We call the structure obtained by analytic extending the spin map off the unitary group a "complex spin structure." We simply replace u and $\tilde{u}$ in (3.2) and (3.3) by z and $\tilde{z}$, respectively. We then have

$$\hat{\sigma}(z) | \mathbf{U} = \sigma(u).$$

The complex spin map is no longer Hermitian but still left invariant.

We also will make use of the ordinary derivative operator $\tilde{e}^{\beta\Omega'}{}_\Omega \partial_\beta := \partial^{\Omega'}{}_\Omega := \partial/\partial_z{}^\Omega{}_{\Omega'}$ which is a derivation with respect to the basis of matrix units. Sometimes we also write ∂_z to indicate that we differentiate with respect to $z \in \mathbf{C}^\Sigma{}_{\Sigma'}$. Simple examples of using this operator on matrices are

$$\partial^{\Omega'}{}_{\Omega} z^{\Sigma}{}_{\Sigma'} = \delta^{\Omega'}{}_{\Sigma'} \delta^{\Sigma}{}_{\Omega},$$
$$\partial^{\Omega'}{}_{\Omega} \tilde{z}^{\Sigma}{}_{\Sigma'} = -\tilde{z}^{\Sigma}{}_{v} (\partial^{\Omega'}{}_{\Omega} z^{v}{}_{v'}) \tilde{z}^{v'}{}_{\Sigma'} = -\tilde{z}^{v'}{}_{\Omega'} \tilde{z}^{\Omega'}{}_{\Sigma'}.$$

The last identity is obtained by applying $\partial^{\Omega'}{}_{\Omega}$ to $\tilde{z}\cdot z$ and using the Leibniz rule.

When we perform a differentiation we have to be careful when we go from one basis to the other, especially when we deal with a moving basis. As an example we treat the ordinary differentiation of a vector v,

$$\begin{aligned}\partial_\alpha v^\beta &= e_\alpha{}^{\Omega}{}_{\Omega'}(z) \partial^{\Omega'}{}_{\Omega} \tilde{e}^{\beta\Sigma'}{}_{\Sigma}(z) v^{\Sigma}{}_{\Sigma'} \\ &= e_\alpha{}^{\Omega}{}_{\Omega'}(z) \tilde{e}^{\beta\Sigma'}{}_{\Sigma}(z) \partial^{\Omega'}{}_{\Omega} v^{\Sigma}{}_{\Sigma'} \\ &\quad + v^{\Sigma}{}_{\Sigma'} e_\alpha{}^{\Omega}{}_{\Omega'}(z) \partial^{\Omega'}{}_{\Omega} \tilde{e}^{\beta\Sigma'}{}_{\Sigma}(z).\end{aligned}$$

The second term arises from the moving basis, as is familiar for nonholonomic systems, i.e., if the tangent vectors do not commute. To calculate this term we make use of the fact that the group acts freely and transitively on itself. Let λ be the basis at the identity (id), i.e., $e_\beta(\mathrm{id}) = :\lambda_\beta$ and $\tilde{e}^\beta(\mathrm{id}) = :\tilde{\lambda}^\beta$. The basis at any other point z is then simply obtained by left translation of the group,

$$e_\beta{}^{\Omega}{}_{\Omega'}(z) = (z\cdot\lambda_\beta)^{\Omega}{}_{\Omega'}, \text{ and } \tilde{e}^{\beta\Omega'}{}_{\Omega}(z) = (\tilde{\lambda}^\beta\cdot\tilde{z})^{\Omega'}{}_{\Omega}. \tag{3.6}$$

The second term reduces now to

$$\begin{aligned} v^{\Sigma}{}_{\Sigma'} e_\alpha{}^{\Omega}{}_{\Omega'}(z) \partial^{\Omega'}{}_{\Omega} \tilde{e}^{\beta\Sigma'}{}_{\Sigma}(z) &= v^{\Sigma}{}_{\Sigma'} e_\alpha{}^{\Omega}{}_{\Omega'}(z) \tilde{e}^{\beta\Sigma'}{}_{\Omega}(z) \tilde{z}^{\Omega'}{}_{\Sigma} \\ &= -\operatorname{Tr}(v\cdot\tilde{e}^\beta\cdot e_\alpha\cdot\tilde{z}).\end{aligned}$$

The vector connection $C^{\gamma}{}_{\alpha\beta}$ is uniquely defined by the spin connection, if we demand that the spin map is torque-free (2.3). We will first show that the ordinary derivative of $\hat{\sigma}$ is zero,

$$\begin{aligned} Tp(\tilde{\sigma}_\lambda\cdot(\partial_\alpha\hat{\sigma}^\beta)) &= e_\alpha{}^{\Delta}{}_{\Delta'} \tilde{e}^{\beta\Omega'}{}_{\Omega} e_\lambda{}^{\Lambda}{}_{\Lambda'} (\partial^{\Delta'}{}_{\Delta} \hat{\sigma}^{\Omega}{}_{\Omega'\dot{\Sigma}\Sigma}) \tilde{\hat{\sigma}}^{\Lambda'}{}_{\Lambda}{}^{\Sigma\dot{\Sigma}} \\ &\quad + e_\alpha{}^{\Delta}{}_{\Delta'} e_\lambda{}^{\Omega}{}_{\Omega'} (\partial^{\Delta'}{}_{\Delta} \tilde{e}^{\beta\Omega'}{}_{\Omega}) \\ &= \operatorname{Tr}(\tilde{e}^\beta\cdot e_\alpha\cdot\tilde{z}\cdot e_\lambda) - \operatorname{Tr}(\tilde{e}^\beta\cdot e_\alpha\cdot\tilde{z}\cdot e_\lambda) \\ &= 0,\end{aligned}$$

where we used the definition of $\hat{\sigma}$ (3.4). Because we can solve for $\partial_\alpha \hat{\sigma}^\beta{}_{\dot{\Sigma}\Sigma}$ by transvecting with $\hat{\sigma}^\lambda$, we showed that $\partial_\alpha\hat{\sigma}^\beta$ is zero. This result is true everywhere in the group, and it also implies that the ordinary derivative of g is zero everywhere. This does of course not mean that the space is flat. Our coordinate system is such that the metric is everywhere constant, a condition that is familiar from Cartan's "repère mobile." Because all the ordinary derivatives of g are zero, the hyper Christoffel formula (2.9) tells us that the symmetric part of C is zero. We therefore expect to find an antisymmetric expression for C.

Using definition (3.4) of the analytically extended spin connection and (2.3) we obtain

$$\begin{aligned} \hat{C}^{\Omega}{}_{\Omega'}{}^{\Delta'}{}_{\Delta}{}^{\Lambda'}{}_{\Lambda} &:= e_\beta{}^{\Omega}{}_{\Omega'} \tilde{e}^{\alpha\Delta'}{}_{\Delta} \tilde{e}^{\gamma\Lambda'}{}_{\Lambda} \hat{C}^{\beta}{}_{\alpha\gamma} \\ &= kz^{\Omega}{}_{\Sigma'} \mu_{\dot{\Sigma}\Omega'} \hat{\Gamma}_\alpha{}^{\Sigma'}{}_{\Sigma} \tilde{z}^{\Sigma}{}_{\Lambda} \tilde{\mu}^{\Lambda'\dot{\Sigma}} \\ &\quad - k\tilde{z}^{\Sigma}{}_{\Lambda} \tilde{\mu}^{\Lambda'\dot{\Sigma}} z^{\Omega}{}_{\Sigma} \mu_{\dot{\Sigma}'\Omega'} \mu_{\dot{\Sigma}\Sigma'} \hat{\Gamma}_\alpha{}^{\Sigma'}{}_{\Sigma'} \tilde{\mu}^{\Sigma'\dot{\Sigma}} \\ &= k(\delta^{\Omega}{}_{\Delta} \delta^{\Lambda'}{}_{\Omega'} \tilde{z}^{\Delta'}{}_{\Lambda} - \delta^{\Omega}{}_{\Lambda} \delta^{\Delta'}{}_{\Omega'} \tilde{z}^{\Lambda'}{}_{\Delta}).\end{aligned}$$

This expression can be brought into a nicer form by rewriting it with the help of (3.6). We find

$$\hat{C}^{\beta}{}_{\alpha\gamma}(z) = k\operatorname{Tr}(\tilde{\lambda}^\beta[\lambda_\alpha,\lambda_\gamma]) = kc^{\beta}{}_{\alpha\gamma}.$$

The $c^{\beta}{}_{\alpha\gamma}$ are the structure constants of the group defined by $[\lambda_\alpha,\lambda_\gamma] = c^{\beta}{}_{\alpha\gamma}\lambda_\beta$. From the assumed form of Γ it is therefore not possible to determine the value of the constant k. The exact value of k is fixed, however, if we demand that the vector connection is also torsion-free, which we wanted to happen in the first place. The torsion tensor[10] $T^{\beta}{}_{\alpha\gamma}$ in an anholonomic frame is defined as

$$\hat{T}^{\beta}{}_{\alpha\gamma}(z) := \hat{C}^{\beta}{}_{\alpha\gamma} - \hat{C}^{\beta}{}_{\gamma\alpha} - c^{\beta}{}_{\alpha\gamma} = (2k-1)c^{\beta}{}_{\alpha\gamma}.$$

It is immediate that the torsion vanishes if and only if $k = \frac{1}{2}$. The unique torsion-free vector connection is therefore given by

$$\hat{C}^{\beta}{}_{\alpha\gamma} = \tfrac{1}{2}c^{\beta}{}_{\alpha\gamma}. \tag{3.7}$$

This expression is familiar in the differential geometry of Lie groups. It is exactly the connection obtained by using the left invariant Mauer–Cartan forms of the group, and is also called Cartan's (0)-connection.[11] We will introduce the Cartan calculus in Sec. IV, but continue for completeness with our calculations.

We note also one identity: The left invariant vector connection in the hyperspin geometry and the Levi-Civita connection on the unitary troup coincide for all N. For $N=2$ this was expected because the spin structure gives rise to a Riemannian geometry. At first it is surprising that it holds also true for the Finsler geometries arising from the higher values of N, since left invariance does not uniquely determine the connection. However, the torsion-free condition is the same for both the hyperspin and the Riemannian geometries, and thus fixes the connection.

For completeness we show here also the spin connection Γ which we can write from (3.5) as

$$\Gamma_\alpha{}^{\Sigma}{}_{\Sigma'} = \tfrac{1}{2}(\lambda_\alpha{}^{\Sigma}{}_{\Sigma'} - (1/N)\delta^{\Sigma}{}_{\Sigma'} \operatorname{Tr}(\lambda_\alpha)). \tag{3.8}$$

We see that Γ has values only in $d\,\mathbf{SU}_N$, because the imaginary identity matrix as the $\mathbf{U}_1$ generator gives a zero connection, whereas the $\mathbf{SU}_N$ generators are already traceless.

The next thing to do is to compute the left invariant curvature. We leave for the moment the constant k undetermined, to obtain greater generality. The definition of the curvature tensor in a nonholonomic frame is

$$\begin{aligned}\hat{R}^{\alpha}{}_{\beta[\gamma\delta]} &= \partial_\gamma \hat{C}^{\alpha}{}_{\beta\delta} - \partial_\delta \hat{C}^{\alpha}{}_{\beta\gamma} + \hat{C}^{\alpha}{}_{\lambda\gamma}\hat{C}^{\lambda}{}_{\beta\delta} \\ &\quad - \hat{C}^{\alpha}{}_{\lambda\delta}\hat{C}^{\lambda}{}_{\beta\gamma} - \hat{C}^{\alpha}{}_{\lambda\beta}c^{\lambda}{}_{\gamma\delta} \\ &= -k^2 c^{\alpha}{}_{\lambda\gamma}c^{\lambda}{}_{\delta\beta} - k^2 c^{\alpha}{}_{\lambda\delta}c^{\lambda}{}_{\beta\gamma} - kc^{\alpha}{}_{\lambda\beta}c^{\lambda}{}_{\gamma\delta} \\ &= k(k-1)c^{\alpha}{}_{\lambda\beta}c^{\lambda}{}_{\gamma\delta}.\end{aligned} \tag{3.9}$$

The last step involves the Jacobi identity of the structure constants,

$$c^{\alpha}{}_{\lambda\beta}c^{\lambda}{}_{\gamma\delta} + c^{\alpha}{}_{\lambda\delta}c^{\lambda}{}_{\beta\gamma} + c^{\alpha}{}_{\lambda\gamma}c^{\lambda}{}_{\delta\beta} = 0. \tag{3.10}$$

We see that the curvature tensor reduces to a simple form. It manifestly has all the standard symmetries of a curvature tensor in the absence of a Riemann metric, i.e., is antisymmetric in γ and δ, and satisfies the differential (second) Bianchi identity.

It also satisfies the cyclic first Bianchi identity for all values of k, which again is a direct consequence of the Jacobi identity. This is somewhat surprising, because the existence of this symmetry is normally only proven in the absence of torsion, but we have to keep in mind that we used a special torsion arising from a left invariant spin connection.

For $k = 1$ or $k = 0$ we observe that the curvature tensor vanishes identically as though the curvature contributions of the metric and the torsion exactly cancel and we call the corresponding connections "flat." The vector connections give rise to torsion of equal magnitude but opposite sign. They correspond to Cartan's (+) and (−) connections on arbitrary simple and semisimple groups.[11] It is interesting to note that torsion can arise even with zero spin connection ($k = 0$).

The Ricci tensor is found by contracting the curvature tensor

$$\widehat{R}^{\gamma}{}_{\alpha\gamma\beta} = \widehat{R}_{\alpha\beta} = k(k-1)c^{\gamma}{}_{\lambda\alpha}c^{\lambda}{}_{\gamma\beta}. \tag{3.11}$$

The Ricci tensor is symmetric for all values of k. This expression is proportional to the Killing metric of the group,[10] which is bi-invariant and negative definite, because we are dealing with a semisimple group (i.e., $\mathbf{SU}_N$).

We have now determined the geometry of $\mathbf{U}_N$ completely. We note that the time dimension does not contribute to the curvature, because all structure constants involving the unit matrix are zero. This was to be expected, because the manifold is static.

Looking at hypergravity we can state the following Lemma.

Lemma: The unitary groups form a one-parameter family of solutions for the hypergravity equations (2.12). For $N > 2$ they obey the empty space equations.

Proof: This is a simple consequence of the fact that the Ricci tensor is covariantly constant.

$$D_\gamma R_{\alpha\beta} = \partial_\gamma R_{\alpha\beta} - C^{\lambda}{}_{\gamma\alpha}R_{\lambda\beta} - C^{\lambda}{}_{\gamma\beta}R_{\alpha\lambda} = 0.$$

The vanishing of the two connection terms follows because the Ricci tensor is proportional to the Killing metric of the group and the trace of the structure constants of $\mathbf{U}_N$ is zero. R and $R^{\alpha}{}_{\beta}$ are zero because they involve covariant derivatives of Ricci, and (2.12) is trivially satisfied in the absence of matter. ∎

Here, $\mathbf{U}_2$ is a solution only in the presence of matter and with the inclusion of a cosmological term λ, which is linked to the radius ρ of the universe and the mean matter density. This was Einstein's result[12] which led him to introduce the infamous λ term.

B. Application: The geometry of $\mathbf{U}_2$ as example

In order to make our calculations more explicit we apply our formulas to Cov $\mathbf{U}_2$, which is the time space $\mathbf{R}\times S^3$. In this case our hyperspinors are the ordinary $\mathbf{SL}_2$ Weyl spinors and the geometry they give rise to is Riemannian. Only for this case can we form the completely contracted Riemann curvature scalar R because it requires the existence of a quadratic metric. For the following we assume zero torsion ($k = \frac{1}{2}$) and without loss of generality we specialize the metric μ to be the unit matrix, so that $\delta^{\Sigma\dot{\Sigma}} = \tilde{\mu}^{\Sigma\dot{\Sigma}}$. As a convenient basis for $d\,\mathbf{U}_2(\mathrm{id})$ we choose the anti-Hermitian matrices

$$\lambda_0 := \frac{i}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\quad \lambda_1 := \frac{i}{\sqrt{2}}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},$$

$$\lambda_2 := \frac{i}{\sqrt{2}}\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix},\quad \lambda_3 := \frac{i}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

Due to our choice the dual basis $\tilde{\lambda}^{\alpha}$ has the same matrix representation, only i is replaced by $-i$. The λ_α obey the commutation relation

$$[\lambda_i,\lambda_j] = \sqrt{2}\epsilon^{k}{}_{ij}\lambda_k,\quad \text{with } i,j,k = 1,\ldots,3,$$

where $\epsilon^{k}{}_{ij}$ is the three-dimensional Levi-Civita symbol with $\epsilon^{3}{}_{12} = 1$. From (3.4) and (3.6) we have

$$\sigma^{\alpha}{}_{\dot{\Sigma}\Sigma}(u) = i\mu_{\dot{\Sigma}\Omega}\tilde{\lambda}^{\alpha\Omega}{}_{\Sigma} =: i\tilde{\lambda}^{\alpha}{}_{\dot{\Sigma}\Sigma}.$$

The cometric $g^{\alpha\beta}$ is defined by

$$\begin{aligned} g^{\alpha\beta} &= \epsilon^{\Sigma\Omega}\epsilon^{\dot{\Sigma}\dot{\Omega}}\sigma^{\alpha}{}_{\Sigma\dot{\Sigma}}\sigma^{\beta}{}_{\Omega\dot{\Omega}} \\ &= (\delta^{\Sigma\dot{\Sigma}}\delta^{\Omega\dot{\Omega}} - \delta^{\Sigma\dot{\Omega}}\delta^{\Omega\dot{\Sigma}})(-\tilde{\lambda}^{\alpha}{}_{\dot{\Sigma}\Sigma}\tilde{\lambda}^{\beta}{}_{\dot{\Omega}\Omega}) \\ &= \mathrm{Tr}(\tilde{\lambda}^{\alpha}\cdot\tilde{\lambda}^{\beta}) - \mathrm{Tr}(\tilde{\lambda}^{\alpha})\mathrm{tr}(\tilde{\lambda}^{\beta}). \end{aligned}$$

This leads to $\mathrm{diag}(g^{\alpha\beta}) = (1,-1,-1,-1) = \mathrm{diag}(g_{\alpha\beta})$. The curvature tensor according to (3.9) is

$$R^{\alpha}{}_{\beta\gamma\delta} =: R^{i}{}_{jkl} = \tfrac{1}{2}\epsilon^{i}{}_{jm}\epsilon^{m}{}_{kl} = \tfrac{1}{2}(\delta^{i}{}_{k}\delta_{jl} - \delta^{i}{}_{l}\delta_{jk}),$$

where we used the convention $\alpha,\beta,\gamma,\delta = 0,\ldots,3$ and $i,j,k,l = 1,\ldots,3$. The equality has to be understood to mean that whenever an index is zero the whole expression is zero. The Ricci tensor is obtained by an approprite contraction

$$R_{\beta\delta} := R_{jl} = \delta_{jl}.$$

The curvature scalar is then obtained by contracting the Ricci tensor with the cometric,

$$R = g^{\beta\delta}R_{\beta\delta} = -\delta^{jl}\delta_{jl} = 3.$$

The curvature scalar is a kind of averaged curvature. Because the time dimension does not contribute to the curvature the result shows that each spatial dimension contributes -1 on the average to the curvature.

IV. EXTERIOR CALCULUS WITH HYPERSPINORS

Cartan's calculus of differential forms compactifies the tensorial calculus of pseudo-Riemannian manifolds and renders symmetries more obvious. In this section we generalize the formalism to include Bergmann manifolds as well. The notation of this section is as follows: small Greek letters stand for the manifold index, $\mu = 1,\ldots,N^2$. The index $A\dot{A}$ is a pair of spinor *n-ad* indices ($n = N^2$), where $A = 1,\ldots,N$. Small Latin indices stand for the vector *n-ad* indices. $A\dot{A}$ can also be used as a composite vector index and vice versa and we will switch freely between both notations.

The basis one forms are $\tilde{\sigma}^{a} := \tilde{\sigma}^{a}{}_{\mu}\,dx^{\mu}$ together with their duals $\sigma_a := \sigma_a{}^{\mu}\partial_\mu$. They obey as usual the inverse relations

$$\tilde{\sigma}_{\mu}{}^{a}\sigma^{\mu}{}_{b} = \delta^{a}{}_{b}$$

and

$$\tilde{\sigma}_{\mu}{}^{a}\sigma^{\nu}{}_{a} = \delta^{\nu}{}_{\mu}.$$

Next we introduce the exterior derivative, which is defined as

$$d := \partial_\mu \wedge dx^{\mu}$$

and acts by exterior multiplication. The chronometric is defined as

$$g_{\alpha\beta\cdots\nu} = \eta_{ab\cdots n}\tilde{\sigma}^a{}_\alpha \ldots \tilde{\sigma}^n{}_\nu,$$

where $\eta_{ab\cdots n}$ is the constant frame chronometric and has determinantal form,

$$\eta_{ab\cdots n} = [1/(N-1)!]\epsilon_{AB\cdots N}\epsilon_{\dot{A}\dot{B}\cdots\dot{N}}.$$

The inverse chronometric is defined as

$$\eta^{ab\cdots n} = [1/(N-1)!]\epsilon^{AB\cdots N}\epsilon^{\dot{A}\dot{B}\cdots\dot{N}}$$

and obeys

$$\eta^{sb\cdots n}\eta_{b\cdots nt} = \delta^s{}_t.$$

Cartan's structure equations in the vector notation are not altered. They are

$$d\tilde{\sigma}^a + w^a{}_b \wedge \tilde{\sigma}^b = T^a = \tfrac{1}{2}T^a{}_{bc}\tilde{\sigma}^b \wedge \tilde{\sigma}^c, \tag{4.1a}$$

$$R^a{}_b = dw^a{}_b + w^a{}_s \wedge w^s{}_b = \tfrac{1}{2}R^a{}_{bcd}\tilde{\sigma}^c \wedge \tilde{\sigma}^d. \tag{4.1b}$$

Here T^a is the torsion two-form and $R^a{}_b$ is the curvature two-form. They are related in the usual way to the torsion and curvature tensor. The $w^a{}_b$ are called "affine spin connection" in the usual literature. This name is now misleading, because it has nothing to do with the spin substructure we assume here, and is therefore replaced by "vector connection one-forms." Taking the exterior derivative of (4.1a) gives a consistency condition

$$dT^a + w^a{}_b \wedge T^b = R^a{}_b \wedge \tilde{\sigma}^b,$$

while differentiating (4.1b) gives the Bianchi identities

$$dR^a{}_b + w^a{}_c \wedge R^c{}_b - R^a{}_c \wedge w^c{}_b = 0.$$

Imposing two conditions on the structure equation will lead to a uniquely defined vector connection,[8] analogous to the Levi-Civita connection for Riemannian manifolds. These two conditions are the following.

(1) Metricity: $Dg_{\alpha\beta\cdots\eta} = 0$.

This is equivalent to $w^a{}_a = 0$. (4.2)

Proof: $d\eta_{ab\cdots n} = 0 = d(\sigma^\alpha{}_a\sigma^\beta{}_b \ldots \sigma^\eta{}_n)g_{\alpha\beta\cdots\eta} = \eta_{b\cdots ns} \times w^s{}_a + \cdots + \eta_{ab\cdots s}w^s{}_n$. Transvecting with $\eta^{ab\cdots n}$ gives the desired result. ■

(2) Zero torsion: $T^a = 0$.

This means according to (4.1),

$$d\tilde{\sigma}^a = -w^a{}_b \wedge \tilde{\sigma}^b = -w^a{}_{cb}\tilde{\sigma}^c \wedge \tilde{\sigma}^b,$$

where $w^a{}_{cb}$ are the coefficients of the connetion one-form in the *n-ad* basis. Assuming a complete basis of one-forms, the last equation can be rewritten in terms of the expansion coefficients $c^a{}_{[bc]}$ which are antisymmetric in b and c,

$$d\tilde{\sigma}^a = \tfrac{1}{2}c^a{}_{bc}\tilde{\sigma}^b \wedge \tilde{\sigma}^c. \tag{4.3}$$

For the following lemma it will be useful to define the contraction tensor $h^{sc}{}_{s'd}$ as

$$h^{sc}{}_{s'b} := \eta^{scd\cdots n}\eta_{d\cdots ns'b}. \tag{4.4}$$

Lemma: The unique metric and torsion-free hyper-Christoffel connection is determined by

$$w^s{}_{ab} = \tfrac{1}{2}(c^s{}_{ab} - c^{s'}{}_{ac}h^{sc}{}_{s'b} - c^{s'}{}_{bc}h^{sc}{}_{s'a}). \tag{4.5}$$

Proof: From (4.3) it is found that

$$w^a{}_{bc} - w^a{}_{cb} = c^a{}_{bc}. \tag{4.6}$$

An easy calculation shows that (4.5) satisfies (4.6) and (4.2). Because the solution is necessarily unique, this proves the lemma. ■

We also note that for $N = 2$, h is a product of two Riemannian metrics, $h^{sc}{}_{s'b} = \eta^{sc}\eta_{s'b}$, and (4.5) reduces to the standard expression of Riemannian geometry

$$w_{cab} = \tfrac{1}{2}(c_{cab} - c_{bac} - c_{abc}),$$

where $c_{cab} := \eta_{sc}c^s{}_{ab}$.

As in the Riemannian case the $w^a{}_{bc}$ are related to the manifold Christoffel symbols $C^\nu{}_{\mu\lambda}$ via[13]

$$w^a{}_{bc} = -\tilde{\sigma}^a{}_\nu\sigma^\mu{}_c(\partial_\mu\sigma^\nu{}_b + C^\nu{}_{\mu\lambda}\sigma^\lambda{}_b).$$

We now extend the Cartan calculus to include $\mathbf{SL}_N$ hyperspinors as well. We call it the hyper-Cartan–Penrose calculus, because to our knowledge Penrose was the first to give a detailed exposition of this subject for $N = 2$, although a remark of Nester in Ref. 14 shows that Trautmann was well aware of the methods, too.

We will rewrite the structure equations with explicit spinor indices. The basis one forms are $\tilde{\sigma}^{A\dot{A}} := \tilde{\sigma}_\mu{}^{A\dot{A}}\,dx^\mu$ together with their duals $\sigma_{\dot{A}A} = \sigma^\mu{}_{\dot{A}A}\partial_\mu$. They obey, of course,

$$\tilde{\sigma}_\mu{}^{A\dot{A}}\sigma^\mu{}_{\dot{B}B} = \delta^A{}_B\delta^{\dot{A}}{}_{\dot{B}}$$

and

$$\sigma^\mu{}_{\dot{A}A}\tilde{\sigma}_\nu{}^{A\dot{A}} = \delta^\mu{}_\nu.$$

Indeed, the $\tilde{\sigma}_\mu{}^{A\dot{A}}$ are nothing but the components of the spin form, as the notational similarity already suggests. Cartan's n-ad has acquired a real physical meaning, because *the n-ad is σ*, which is the basic variable in the hyperspin theory.

For simplicity we set the torsion to be zero for the following. The structure equations become now

$$d\tilde{\sigma}^{A\dot{A}} = w^{A\dot{A}}{}_{\dot{B}B} \wedge \tilde{\sigma}^{B\dot{B}} = (-w^A{}_B\delta^{\dot{A}}{}_{\dot{B}} - \overline{w}^{\dot{A}}{}_{\dot{B}}\delta^A{}_B) \wedge \tilde{\sigma}^{B\dot{B}} \tag{4.7}$$

and

$$R^{A\dot{A}}{}_{\dot{B}B} = dw^{A\dot{A}}{}_{\dot{B}B} + w^{A\dot{A}}{}_{\dot{C}C} \wedge w^{C\dot{C}}{}_{\dot{B}B}$$
$$= R^A{}_B\delta^{\dot{A}}{}_{\dot{B}} + \overline{R}^{\dot{A}}{}_{\dot{B}}\delta^A{}_B,$$

where $R^A{}_B$ and $w^A{}_B$ are the spin curvature two-form and the spin-connection one-form, respectively. They take values in $d\,\mathbf{SL}_N$, the Lie algebra of the structural group and are therefore traceless. The relation to the vector quantities is as follows:

$$N\cdot R^A{}_B = R^{A\dot{A}}{}_{\dot{A}B}, \quad N\cdot w^A{}_B = w^{A\dot{A}}{}_{\dot{A}B}.$$

The contraction tensor h (4.4) can be written out explicitly with spinor indices due to the special algebraic structure of η. We find

$$h^{S\dot{S}C\dot{C}}{}_{\dot{T}T\dot{B}B}$$
$$= [1/(N-1)^2](\delta^S{}_T\delta^C{}_B\delta^{\dot{S}}{}_{\dot{T}}\delta^{\dot{C}}{}_{\dot{B}} + \delta^S{}_B\delta^C{}_T\delta^{\dot{S}}{}_{\dot{B}}\delta^{\dot{C}}{}_{\dot{T}}$$
$$- \delta^S{}_T\delta^C{}_B\delta^{\dot{S}}{}_{\dot{B}}\delta^{\dot{C}}{}_{\dot{T}} - \delta^S{}_B\delta^C{}_T\delta^{\dot{S}}{}_{\dot{T}}\delta^{\dot{C}}{}_{\dot{B}}).$$

The unique torsion-free hyperspin connection of (4.5) in spinor coordinates has the form

$$w^{S\dot{S}}{}_{\dot{A}A\dot{B}B} = \tfrac{1}{2}c^{S\dot{S}}{}_{\dot{A}A\dot{B}B} + [1/2(N-1)^2](c^{S\dot{C}}{}_{\dot{A}AB\dot{C}}\delta^{\dot{S}}{}_{\dot{B}}$$
$$+ c^{C\dot{S}}{}_{\dot{A}A\dot{B}C}\delta^S{}_B - c^{C\dot{C}}{}_{\dot{A}A\dot{C}C}\delta^S{}_B\delta^{\dot{S}}{}_{\dot{B}}$$
$$+ c^{S\dot{C}}{}_{\dot{B}BA\dot{C}}\delta^{\dot{S}}{}_{\dot{A}} + c^{C\dot{S}}{}_{\dot{B}B\dot{A}C}\delta^S{}_A$$
$$- c^{C\dot{C}}{}_{\dot{B}B\dot{C}C}\delta^S{}_A\delta^S{}_A\delta^{\dot{S}}{}_{\dot{A}}). \tag{4.8}$$

The spin connection coefficients Γ are the expansion coefficients of the spin connection one-form, i.e.,

$$\Gamma_{A\dot{A}}{}^{B}{}_{C}\tilde{\sigma}^{A\dot{A}} = w^{B}{}_{C}. \tag{4.9}$$

Putting this into (4.7) gives

$$d\tilde{\sigma}^{A\dot{A}} = (\Gamma_{C\dot{C}}{}^{A}{}_{B}\tilde{\sigma}^{B\dot{A}} + \overline{\Gamma}_{C\dot{C}}{}^{\dot{A}}{}_{\dot{B}}\tilde{\sigma}^{A\dot{B}}) \wedge \tilde{\sigma}^{C\dot{C}}. \tag{4.10}$$

The Γ's are again uniquely defined by (4.10), and by using (4.8) and (4.9) one can express them in terms of the expansion coefficients $c^{S\dot{S}}{}_{A\dot{A}B\dot{B}}$,

$$\begin{aligned}\Gamma_{A\dot{A}}{}^{S}{}_{B} = {} & (1/2N)c^{S\dot{S}}{}_{A\dot{A}S\dot{B}} - [1/2N(N-1)^2] \\ & \times\{(N-1)c^{T\dot{T}}{}_{A\dot{A}T\dot{T}}\delta^{S}{}_{B} + c^{T\dot{T}}{}_{\dot{A}BT\dot{T}}\delta^{S}{}_{A} \\ & - Nc^{S\dot{T}}{}_{A\dot{A}T\dot{B}} - c^{S\dot{T}}{}_{\dot{A}BT\dot{A}} - c^{T\dot{T}}{}_{TB\dot{A}T}\delta^{S}{}_{A}\}.\end{aligned} \tag{4.11}$$

Because the manifold we look upon is a group manifold it is possible to use the Mauer–Cartan equations[10] for invariant differential forms to compute the vector connection and vector curvature in the torsion-free case. By combining this method with the hyper Penrose–Cartan calculus we have a powerful tool to determine even the spin structure out of the knowledge of the structure constants of $\mathbf{U}_N$ alone.

Let $e_1,...,e_N$ be a basis of $d\,\mathbf{U}_N$, σ_j the corresponding left invariant vector fields on $\mathbf{U}_N$, and $\tilde{\sigma}^1,...,\tilde{\sigma}^N$ the left invariant one-forms determined by $\sigma^i(\tilde{\sigma}_j) = \delta^i{}_j$. Then

$$d\tilde{\sigma}^i = -\tfrac{1}{2}c^i{}_{jk}\tilde{\sigma}^j \wedge \tilde{\sigma}^k,$$

where $c^i{}_{jk}$ are the structural constants given by

$$[e_i,e_j] = c^k{}_{ij}e_k.$$

This means that the coefficients of the connection one-form are given by

$$w^i{}_{jk} = \tfrac{1}{2}c^i{}_{jk},$$

which is exactly the connection C obtained in (3.7). The curvature two-form $R^{a}{}_{b}$ is readily obtained using the Cartan calculus and applying the Jacobi identity,

$$R^{a}{}_{bjs} = \tfrac{1}{4}c^a{}_{bk}c^k{}_{js}.$$

The Ricci tensor is obtained by contracting over the a and j index,

$$R_{ab} = \tfrac{1}{4}c^i{}_{ak}c^k{}_{ib}.$$

These are exactly the results obtained in Sec. III [(3.8) and (3.10)]. The generators of $\mathbf{U}_N$ can be obtained in a standard way by employing Weyl's matrix units $E_i{}^{A}{}_{B} = \delta^{A}{}_{I}\delta^{I'}{}_{B}$. They satisfy the commutation relations

$$[E_i,E_j] = c^k{}_{ij}E_k, \quad c^k{}_{ij} = \delta^{I'}{}_{J}\delta^{K}{}_{I}\delta^{J'}{}_{K'} - \delta^{J'}{}_{I}\delta^{K}{}_{J}\delta^{I'}{}_{K'}. \tag{4.12}$$

The matrix units are also the generators of $\mathbf{GL}_N$. The N^2 generators of $\mathbf{U}_N$ can be arbitrarily labeled by

$$e^I{}_{I'} = iE^I{}_{I'}, \quad I = I',$$

$$e^i{}_{I'} = i(E^I{}_{I'} + E^{I'}{}_{I}), \quad I > I',$$

$$e^i{}_{I'} = E^I{}_{I'} - E^{I'}{}_{i}, \quad I < I',$$

where the indices on $e^I{}_{I'}$ no longer transform simply under $\mathbf{SL}_N$. To apply the exterior calculus of hyperspinors we have to replace each vector index by a pair of spinor indices on the structure constants, which we accomplish by using matrix unit coordinates and applying the spin metric μ. In the notation of (4.12) we define

$$c^{C\dot{C}}{}_{A\dot{A}B\dot{B}} := \tilde{\mu}^{C'\dot{C}}\mu_{\dot{A}A'}\mu_{\dot{B}B'}c^{C}{}_{C'}{}^{A'}{}_{A}{}^{B'}{}_{B}.$$

We observe that $c^j{}_{ij} = 0$ for $\mathbf{U}_N$. This results in a great simplification of (4.8), which becomes

$$w^{C\dot{C}}{}_{A\dot{A}B\dot{B}} = \tfrac{1}{2}c^{C\dot{C}}{}_{A\dot{A}B\dot{B}},$$

and (4.11), which turn into

$$\Gamma_t{}^{C}{}_{B} = \tilde{\sigma}^{A\dot{A}}{}_{t}\Gamma_{A\dot{A}}{}^{c}{}_{B} = (1/2N)c^{C\dot{S}}{}_{A\dot{A}S\dot{B}}\tilde{\sigma}^{A\dot{A}}{}_{t}.$$

This is exactly the same expression as (3.8), which was obtained in a more pedestrian way.

As in the case of Riemannian geometry we see that the exterior calculus, extended to our Finsler geometry, is not only formally elegant but can also be used efficiently for calculations.

V. CONCLUSIONS

We have shown that the unitary groups $\mathbf{U}_N$ possess a global hyperspin structure in addition to their Riemannian structure induced by the Killing metric, turning them into Bergmann manifolds of dimension N^2. The spin map on $\mathbf{U}_N$ is proportional to the left (or right) invariant differential form on the group. The resulting Bergmann manifolds are static, homogeneous, isotropic, and compact spaces of constant curvature. They also turn out to be solutions to the hypergravity field equations.

We have generalized the Cartan–Penrose exterior calculus from spinors to hyperspinors. This has enabled us to deduce the hyperspin structure from the Maurer–Cartan equations, which simplify the calculations greatly. The hyper Cartan–Penrose method with its conceptual simplicity is an important tool, not only for group manifolds, but for the study of Bergmann manifolds in general.

One of the main open problems of using $\mathbf{U}_N$ or any other Bergmann manifold in physics is to find a good dimensional reduction procedure to separate internal from external space. In particular, a $\mathbf{B}_N$ cannot be written as a product of two Bergmann manifolds, due to the different behavior of spin and vector dimensions. This problem is closely connected to the problem of introducing two different length scales on the manifold. We suggest deforming some of the generators of $\mathbf{U}_N$ which generate the internal space by multiplying them by a constant scale factor and holding the undeformed spin map fixed. This will result in a change of the chronometric. The method of group contractions also could be helpful in this context.

The dimensional constraints on Bergmann manifolds could be relaxed by allowing σ to be a singular map,[15] such that not all Hermitian matrices correspond to a time space vector. A similar idea of dimensional reduction was already considered by Einstein and Mayer[16] in their theory of semivectors.

Another possibility is that a strong suitable torsion in the internal dimensions could be responsible for dimensional reduction. The idea is that a torsion constrains motions onto a four-dimensional hypersurface. It would require very high energies to go off the shell and thus creating for us the illusion that we live in a four-dimensional world. The idea that torsion is responsible for the dimensional reduction has to our knowledge not yet been explored in the literature.

The $\mathbf{U}_N$ model, though not yet physical, might have some implications for earlier, more symmetric phases of the universe. A study of the energy spectrum of the simplest wave equation on $\mathbf{U}_N$, which is the neutrino equation, will be presented in a forthcoming paper.[17]

ACKNOWLEDGMENTS

I would like to thank D. Finkelstein for many stimulating discussions.

[1] D. Finkelstein, Phys. Rev. Lett. **56**, 1532 (1986).

[2] R. Penrose and W. Rindler, *Spinors and Space-Time* (Cambridge U. P., New York, 1984), Vol. I.

[3] P. G. Bergmann, Phys. Rev. **107**, 624 (1957).

[4] D. Finkelstein, S. R. Finkelstein, and C. Holm, Phys. Rev. Lett. **59**, 1265 (1987); S. R. Finkelstein, Ph.D. thesis, Georgia Institute of Technology, 1987 (unpublished).

[5] D. Finkelstein, S. R. Finkelstein, and C. Holm, Int. J. Theor. Phys. **25**, 441 (1986).

[6] H. Shimada, Tensor (N. S.) **33**, 365 (1979).

[7] G. S. Asanov, *Finsler Geometry, Relativity and Gauge Theories* (Reidel, Dordrecht, 1985).

[8] C. Holm, Int. J. Theor. Phys. **25**, 1209 (1986).

[9] K. Nomizu, *Lie Groups and Differential Geometry* (The Mathematical Society of Japan, Tokyo, 1956).

[10] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic, New York, 1978).

[11] E. Cartan and J. A. Schouten, Proc. Acad. Amsterdam **24**, 53 (1926).

[12] A. Einstein, Sitzungsber. Preuss. Akad. Wiss. Berlin, Phys.-Math. Kl. **1917**, 142.

[13] T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **66**, 213 (1980).

[14] J. M. Nester, in *An Introduction to Kaluza-Klein Theories*, edited by H. C. Lee (World Scientific, Singapore, 1984).

[15] D. Finkelstein (private communication).

[16] A. Einstein and W. Mayer, Berliner Berichte **1931**, 541.

[17] C. Holm, "Neutrino spectrum of Einstein universes," to be published.

Finite moment equations for a relativistic simple gas

A. Majorana
Dipartimento di Matematica, Università di Catania, Catania, Italy

(Received 25 August 1987; accepted for publication 18 November 1987)

This paper on the relativistic Boltzmann equation deals with the derivation of moment equations for a simple gas. It is shown that for a neutrino gas and for a hard-sphere model in the ultrarelativistic case, the moment equations at each order contain only a finite number of moments of the distribution function.

I. INTRODUCTION

It is clear that in relativistic fluid dynamics, equilibrium distributions have a very limited applicability. In general it is necessary to solve the Boltzmann equation. Fortunately, in most problems in astrophysics and cosmology we are only interested in small deviations from equilibrium. In this case a complete knowledge of the distribution function is unnecessary, but one thinks that it is sufficient to know the first few moments of $f(x,p)$ in order to have a clear description of phenomena. Of course, this requires the derivation of transport equations for these quantities and it is reasonable that they must be related in some sense to the Boltzmann equation. A well-known method developed in this direction is due to Grad,[1] in the nonrelativistic case. Relativistic generalizations have been given by Chernikov,[2] Marle,[3] Kranyš,[4] Anderson,[5] Israel and Stewart,[6] and many others. The method consists of trying to find a solution of the relativistic Boltzmann equation as an expansion in a series. Such an expansion leads to an infinite set of coupled equations for all moments.

It is the aim of this paper to show that, for some species of particles, the moment equations contain at each order a finite number of moments of the distribution function. We believe that this fact can be relevant for the kinetic foundations of relativistic irreversible thermodynamics, which has a close analogy with the moment equations.

The plan of the paper is as follows. In Sec. II we introduce the relativistic Boltzmann equation and the moment equations, and we describe briefly the relativistic version of the Grad method. In Sec. III we derive explicitly the moment equations for a neutrino gas and for the hard-sphere model in the ultrarelativistic limit. In Sec. IV we discuss the connection of our equations with the Grad method and the theory of extended thermodynamics.

II. THE RELATIVISTIC BOLTZMANN EQUATION

We consider a simple relativistic gas in Minkowski space described by the relativistic Boltzmann equation

$$p^\alpha \partial_\alpha f = \frac{1}{2}\int (f_* f_{*1} - f f_1) W(p,p_1|p_*,p_{*1})\,\omega_1 \omega_* \omega_{*1}, \tag{1}$$

where, as usual, we use the abbreviations f, f_1, f_*, f_{*1} for $f(x,p), f(x,p_1), f(x,p_*), f(x,p_{*1})$, $x = x^\alpha = (ct, x^1, x^2, x^3)$ being the space-time coordinates of a particle (c stands for the speed of light and t is the time coordinate) and $p = p^\alpha$ its four-momentum. We indicate the volume element with ω, i.e., $\omega = dp^1\,dp^2\,dp^3/p^0$ and W is the transition rate. Greek indices run from 0 to 3, and the signature of space-time is taken such that $g^{\alpha\beta} = \mathrm{diag}(-1,1,1,1)$. Here

$$W(p,p_1|p_*,p_{*1}) = \sigma s \delta^{(4)}(p + p_1 - p_* - p_{*1}), \tag{2}$$

with a differential cross section σ depending only on s, where

$$s := -P_\alpha P^\alpha, \tag{3}$$

$$P^\alpha := p^\alpha + p_1^\alpha, \tag{4}$$

and $\delta^{(4)}$ is the four-dimensional Dirac delta function, which describes the energy-momentum conservation law during the collisions.

From Eq. (1) one can obtain a moment equation relative to a given polynomial $g = g(p^\alpha)$ by integration, i.e.,

$$\partial_\alpha \int p^\alpha g f \omega = \frac{1}{2}\int f f_1 (g_* - g) W \omega^4, \tag{5}$$

where $\omega^4 = \omega\omega_1\omega_*\omega_{*1}$.

The Grad method is based on this equation. Let

$$f(x,p) = \Phi(x,p)[1 + \phi(x,p)], \tag{6}$$

where $\Phi(x,p)$ is a local Jüttner function. Since one requires that the five parameters of the function Φ correspond to the hydrodynamic four-velocity U^α, to the particle density n, and to the temperature T, then the deviation ϕ must verify the following conditions:

$$\int \Phi\phi p^\mu \omega = 0 \quad (\mu = 0,1,2,3), \tag{7}$$

$$\int \Phi\phi (p^\alpha U_\alpha)^2 \omega = 0. \tag{8}$$

The crucial point of the method is to assume that ϕ can be expanded as

$$\phi = X + X_\mu p^\mu + X_{\mu\nu} p^\mu p^\nu + \cdots, \tag{9}$$

where $X, X_\mu X_{\mu\nu}, \ldots$ are functions of x, $\tau = p^\alpha U_\alpha / K_B T$, K_B being the Boltzmann constant. Also, one assumes that ϕ is *small* with respect to Φ, such that it is possible to use the linear equation

$$\partial_\alpha \int p^\alpha g f \omega = \frac{1}{2}\int \Phi\Phi_1 (1 + \phi + \phi_1)(g_* - g) W \omega^4 \tag{10}$$

instead of Eq. (5).

In the 14-moment approximation one assumes that the particle four-flow N^α and the energy-momentum tensor $T^{\alpha\beta}$ are the only independent thermodynamical variables. Since

 0022-2488/88/040987-03$02.50

the conservation laws give five equations, it is necessary to derive the other nine equations. In the first approximation one assumes that ϕ is given by Eq. (9) (omitting the terms not explicitly written) and, moreover, that X is quadratic in τ, X_μ is linear in τ, and $X_{\mu\nu}$ depends only on x. By taking into account Eqs. (7) and (8) it is possible to obtain ϕ as a function of N^α, $T^{\alpha\beta}$, and p^α (for example, see de Groot *et al.*[7]). One finds that X is proportional to the viscous pressure, X_μ to the heat flow, and $X_{\mu\nu}$ to the traceless viscous pressure tensor. Therefore, by introducing the explicit expansion of $f=\Phi(1+\phi)$ into Eq. (10) for $g=p^\beta p^\gamma$ $(\beta,\gamma=0,1,2,3)$, one can obtain the nine additional equations. We remark that there exist procedures other than this one, but all are based on the expansion (9).

III. MOMENT EQUATIONS FOR NEUTRINO AND HARD-SPHERE GASES

Since σ depends only on s, it is possible to perform the integration easily over p^α_{*1} and partially over p^α_*. This gives

$$\partial_\alpha \int p^\alpha g f\omega = \int f f_1 s\sigma\delta(s+2P_\alpha p^\alpha_*)g_*\omega\omega_1\omega_* - \pi\int f f_1\sigma\sqrt{s^2-4sm^2c^2}\,g\omega_1\omega, \tag{11}$$

where m indicates the particle rest mass.

Successive integrations, which do not destroy the simple form of Eq. (11), can be obtained easily by specifying the function g. Linear and constant functions give conservation laws, which do not contain contributions of the collisional operator. Quadratic functions can be obtained by choosing

$g=p^\alpha p^\beta \quad (\alpha,\beta=0,1,2,3)$.

In this case it is easy to prove that Eq. (11) becomes

$$\begin{aligned}\partial_\alpha S^{\alpha\beta\gamma} = (\pi/3)c\int f f_1\sigma R\,\{&2(m^2c^2-2p_1^\mu p_\mu)p_1^\beta p^\gamma\\ &+2(p_1^\mu p_\mu-2m^2c^2)p^\beta p^\gamma\\ &+[(p_1^\mu p_\mu)^2-m^4c^4]g^{\beta\gamma}\}\omega\omega_1,\end{aligned} \tag{12}$$

where we have introduced the following definitions:

$$S^{\alpha\beta\gamma}:=c\int f p^\alpha p^\beta p^\gamma\omega, \tag{13}$$

$$R:=\sqrt{1-4m^2c^2/s}. \tag{14}$$

We remark that, until now, we have not put forward any hypothesis about the function f except the usual conditions on the integrability of the distribution function.

Now we consider two different species of particles. The first is a neutrino gas. We have

$$\sigma=\tfrac{1}{2}(G/\pi c\hbar^2)^2 s, \tag{15}$$

where G is the weak interaction constant and $\hbar$ is the Planck constant. Since in this case one can take $m=0$, we have $R=1$, and the integral on the right-hand side of Eq. (12) can be separated with respect to the integration because the integrand is ff_1 multiplied by a polynomial of p^α and p_1^α. Therefore it is easy to see that Eq. (12) becomes

$$\begin{aligned}\partial_\alpha S^{\alpha\beta\gamma} = (\pi/3)c(G/\pi c\hbar^2)^2(&-S^{\mu\lambda\nu}S_{\mu\lambda\nu}g^{\beta\gamma}\\ &+4S^{\mu\lambda\beta}S^{\gamma}_{\mu\lambda}-2T^{\mu\lambda}Q^{\beta\gamma}_{\mu\lambda}),\end{aligned} \tag{16}$$

where

$$Q^{\alpha\beta\gamma\delta}:=c\int f p^\alpha p^\beta p^\gamma p^\delta\omega. \tag{17}$$

The second example is the hard-sphere model. In this case σ is a constant and the following inequalities hold:

$$0\leqslant 1-R<4m^2c^2/s. \tag{18}$$

By approximating $R=1$ in Eq. (12) (which corresponds to the ultrarelativistic limit), we obtain the equation

$$\begin{aligned}\partial_\alpha S^{\alpha\beta\gamma} = (\pi/3)(\sigma/c)&\\ \times\{&2m^2c^2N^\beta N^\gamma-4T^{\mu\beta}T^\gamma_\mu+2N^\mu S^{\beta\gamma}_\mu\\ &-4m^2c^2AT^{\beta\gamma}+[T^{\mu\lambda}T_{\mu\lambda}-m^4c^4A^2]g^{\beta\gamma}\},\end{aligned} \tag{19}$$

where

$$A:=c\int f\omega. \tag{20}$$

Notice that (i) if f is a local Jüttner function, then the right-hand side of Eq. (19) vanishes identically; (ii) it is easy to obtain, by using inequalities (18), an exact upper bound for the error introduced by the approximation $R=1$; and (iii) Eq. (12) and Eq. (19) are equivalent when $m=0$.

Introducing the reciprocal effective temperature $z=mc^2/K_B T$, it is reasonable to think that in the ultrarelativistic case (i.e., when z is small enough), the approximate equation (19) is a good approximation. In any case, one can verify this in a given specific problem by using (ii).

Since Eq. (19) contains a few terms, such as $m^2c^2N^\beta N^\gamma$, of the same order of error, we delete these terms for consistency. After this reduction we obtain

$$\partial_\alpha S^{\alpha\beta\gamma} = (\pi/3)(\sigma/c)\times\{-4T^{\mu\beta}T^\gamma_\mu+2N^\mu S^{\beta\gamma}_\mu+T^{\mu\lambda}T_{\mu\lambda}g^{\beta\gamma}\}. \tag{21}$$

Now the right-hand side of the above equation vanishes identically if f is a *local ultrarelativistic Jüttner function*, that is, if we use the ultrarelativistic approximation for $T^{\alpha\beta}$ and $S^{\alpha\beta\gamma}$.

These results for a neutrino gas, as for the hard-sphere model, do not depend on the specific order of the equation, i.e., the third-order moment equation, but they hold in general. To prove this, we examine Eq. (11). Since $\sqrt{s^2-4sm^2c^2}=sR$, in the case $R\equiv 1$, the second integral on the right-hand side of Eq. (11) is separable. We show also that the first integral is separable in the case $R\equiv 1$. In fact, if

$$g=(p^\alpha k_\alpha)^n, \tag{22}$$

with $n\in N$, and k_α an arbitrary constant vector, then it is possible to prove that

$$\int \delta(s+2P_\alpha p^\alpha_\cdot)(p^\beta_\cdot k_\beta)^n\omega_\cdot$$

$$= \left(\frac{1}{2}\right)^n \pi R \sum_{h=0}^{[n/2]} \frac{1}{2h+1}\binom{n}{2h}$$

$$\times (P^\alpha k_\alpha)^{n-2h}[sk_\alpha k^\alpha + (P^\alpha k_\alpha)^2]^h R^{2h}, \qquad (23)$$

where $[n/2]$ is the largest integer not exceeding $n/2$. When $R=1$ the right side of Eq. (23) reduces to a polynomial, and therefore the first integral is also separable. We have chosen g of the kind (22) because it is a scalar, and this simplifies the calculation of integral (23). Moreover, we prove in the Appendix that any polynomial can be obtained by linear combination of polynomials of the kind (22).

IV. CONCLUSIONS

In this paper we have presented explicit simple forms of the moment equations. It is apparent that since they never give a closed system, it is still necessary to use some approximation procedure in order to obtain a consistent closed system. We limit our discussion to the case where N^α and $T^{\alpha\beta}$ are the only independent variables. If we use the Grad method described in Sec. II and linearize the right-hand side of Eq. (16), then we recover the known equations for a neutrino gas in the 14-moment approximation. Whereas if we use the approximate distribution function [Eqs. (6) and (9)], then, as one can easily imagine, Eq. (16) before linearization gives, together with the conservation laws, a different closed set of nonlinear equations. However, the applicability of this new set of equations is still limited by the hypothesis (9).

A different approach is offered by extended thermodynamics,[8] which provides an algorithm for determining the higher moments in terms of N^α and $T^{\alpha\beta}$. We note that now we have the advantage of knowing the explicit expression of the dissipative terms, which in general is not completely determined by extended thermodynamics.[8]

We believe that more general applications of the finite forms of moment equations are feasible when one assumes special forms, with adjustable parameters, of the distribution function in order to describe particular phenomena as, for example, in the case of the Mott-Smith method[9] for the study of the structure of shock waves.

Of course, we can repeat analogous considerations in the case of the hard-sphere model in the ultrarelativistic limit. These topics are currently under investigation.

ACKNOWLEDGMENTS

We have benefited from discussions with Professor Israel and Professor van Weert.

This research was supported by the Ministero della Pubblica Istruzione of Italy.

APPENDIX

We consider the set of all monomials in p real variables $x_1,x_2,\ldots,x_p$, with unit coefficients and degree n, which is the set

$$B^n_p = \{x_1^i x_2^j \cdots x_p^k:\ (i,j,\ldots,k)\in N^p \text{ and } i+j+\cdots+k=n\}. \qquad (\mathrm{A1})$$

It is well-known that B^n_p have d elements, with

$$d = [(n+p-1)(n+p-2)\cdots p]/n!. \qquad (\mathrm{A2})$$

We order B^n_p and we denote with b_h $(1<h<d)$ a generic element of B^n_p. Let V^n_p be the linear real spaces generated by B^n_p (including the zero element), with the usual operations between the polynomials. It is evident that B^n_p is a basis for V^n_p and d is its dimension. Letting $\langle c,x\rangle := \Sigma^p_{m=1} c_m x_m$, where $c=(c_1,c_2,\ldots,c_p)\in\mathbf{R}^p$, we prove the following.

Theorem 1: There exist $c^{(1)},c^{(2)},\ldots c^{(d)}\in\mathbf{R}^p$ such that

$$a_h := \langle c^{(h)},x\rangle^n \quad (h=1,2,\ldots,d) \qquad (\mathrm{A3})$$

are linearly independent; that is, the set (a_h) is a basis for V^n_p. By using Leibniz's formula, we have

$$\langle c,x\rangle^n = \sum \frac{n!}{i!j!\cdots k!}(c_1x_1)^i(c_2x_2)^j\cdots(c_px_p)^k$$

$$= \sum_{h=1}^d \frac{n!}{i!j!\cdots k!}(c_1)^i(c_2)^j\cdots(c_p)^k b_h$$

$$= \sum_{h=1}^d \frac{n!}{i!j!\cdots k!} b_h(c) b_h,$$

where we have indicated with $b_h(c)$ the function from $\mathbf{R}^p$ to $\mathbf{R}$ associated to b_h and evaluated at the point c.

To prove the theorem, it is sufficient to show that the determinant

$$\begin{vmatrix} b_1(c^{(1)}) & b_2(c^{(1)}) & \cdots & b_d(c^{(1)}) \\ b_1(c^{(2)}) & b_2(c^{(2)}) & \cdots & b_d(c^{(2)}) \\ \vdots & \vdots & & \vdots \\ b_1(c^{(d)}) & b_2(c^{(d)}) & \cdots & b_d(c^{(d)}) \end{vmatrix}$$

is nonsingular. One obtains this easily by *reductio ad absurdum* and using the Laplace theorem to evaluate the determinant.

[1] H. Grad, Commun. Pure Appl. Math. **2**, 331 (1949).
[2] N. A. Chernikov, Phys. Lett. **5**, 115 (1963).
[3] C. Marle, Ann. Inst. H. Poincaré A **10**, 127 (1969).
[4] M. Kranyš, Phys. Lett. A **33**, 77 (1970).
[5] J. L. Anderson, in *Proceedings of Midwest Conference on Relativity*, edited by M. Carmeli (Plenum, New York, 1970), pp. 109–131.
[6] W. Israel and J. M. Stewart, Phys. Lett. A **58**, 213 (1976); Ann. Phys. (NY) **118**, 341 (1979); Proc. R. Soc. London Ser. A **365**, 43 (1979).
[7] S. R. de Groot, W. A. van Leeuwen, and Cn. G. van Weert, *Relativistic Kinetic Theory* (North-Holland, Amsterdam, 1980).
[8] I-Shih Liu, I. Müller, and T. Ruggeri, Ann. Phys. (NY) **169**, 191 (1986).
[9] H. M. Mott-Smith, Phys. Rev. **82**, 885 (1951).

The ground-state energy of a system of identical bosons

Richard L. Hall
Department of Mathematics, Concordia University, 1455 de Maisonneuve Blvd. W, Montreal, Quebec H3G 1M8, Canada

(Received 29 June 1987; accepted for publication 2 December 1987)

A system of N identical bosons is studied, each having mass m, which interact in $\mathbf{R}^3$ via attractive central pair potentials and obey nonrelativistic quantum mechanics. A lower energy bound is found by the equivalent two-body method. An upper energy bound established previously on the basis of field theory is now derived by variational methods within conventional quantum mechanics. In the case of the linear potential $V_{ij} = \gamma|\mathbf{r}_i - \mathbf{r}_j|$ the bounds imply that the ground-state energy is given by $\mathscr{E} = C(N)(N-1)(\hbar^2/m)^{1/3}(\gamma N/2)^{2/3}$, where $2.3381 < C(N) < 2.343\,52$. The energy is therefore determined in this case with error $< 0.116\%$ for all $N \geqslant 2$. Similar results are given for other power-law potentials.

I. INTRODUCTION

We consider a system of N identical bosons which interact in $\mathbf{R}^3$ via attractive central pair potentials and obey nonrelativistic quantum mechanics. The Hamiltonian for the N-particle system (with the kinetic energy of the center of mass removed) is given explicitly by

$$H = \frac{1}{2m}\sum_{i=1}^{N}\mathbf{p}_i^2 - \frac{1}{2Nm}\left(\sum_{i=1}^{N}\mathbf{p}_i\right)^2 + \sum_{\substack{i,j=1\\ i<j}}^{N}\gamma f\left(\frac{r_{ij}}{a}\right), \qquad (1.1)$$

where m is the mass of a particle, $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is a pair distance, $f(r)$ is the potential *shape*, a is a length parameter, and γ is the coupling parameter.

The present paper is a continuation of the work on complementary energy bounds presented in Ref. 1. We make a minor change by setting up the problem in $\mathbf{R}^3$ rather than in $\mathbf{R}$, but we still use the equivalent two-body model both to formulate the problem and also to provide an energy *lower* bound. The main new result is a derivation (shown in Sec. III) of a general variational upper bound that is *identical* to the bound derived from field theory and used in Ref. 1. The mathematical uncertainties about this upper bound are therefore removed completely. As an illustration, we study the linear potential $V(\mathbf{r}) = \gamma|\mathbf{r}|$ in Sec. IV and provide a simple formula which determines the N-particle energy with an error of less than 0.116% for all $N \geqslant 2$. In Sec. V we consider more general power-law potentials, with shapes given by $f(r) = \mathrm{sgn}(\nu)r^\nu$.

II. THE EQUIVALENT TWO-PARTICLE MODEL

We present here a very brief summary of the formulation given in more detail in Refs. 1 and 2 and the literature quoted therein. It is important to be very careful about relative coordinates and also about the limit $N \to \infty$.

We suppose that new coordinates are defined by $\rho = B\mathbf{R}$, where $\rho = [\rho_i]$ and $\mathbf{R} = [r_i]$ are column vectors of the new and old coordinates, $\rho_1 = (1/\sqrt{N})\Sigma_{i=1}^N \mathbf{r}_i$ is the center-of-mass coordinate, $\rho_2 = (1/\sqrt{2})(\mathbf{r}_1 - \mathbf{r}_2)$ is a pair distance, and B is a real $N \times N$ *orthogonal* matrix. Such orthogonal coordinates, which include the generalized Jacobi coordinates given by

$$B = \begin{pmatrix} 1/\sqrt{N} & 1/\sqrt{N} & 1/\sqrt{N} & * & * & * & 1/\sqrt{N} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 & * & * & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 & * & * & 0 \\ * & * & * & * & * & * & * \\ * & * & * & * & * & * & * \\ 1/\sqrt{N(N-1)} & 1/\sqrt{N(N-1)} & * & * & * & * & 1-N/\sqrt{N(N-1)} \end{pmatrix}$$

as a special case, satisfy two important identities, namely,

$$\sum_{i=1}^{N}\mathbf{r}_i^2 = \sum_{i=1}^{N}\rho_i^2 \qquad (2.1)$$

and

$$\sum_{\substack{i,j=1\\ i<j}}^{N}|\mathbf{r}_i - \mathbf{r}_j|^2 = N\sum_{i=2}^{N}\rho_i^2. \qquad (2.2)$$

The column vectors Π and $\mathbf{P}$ of the new and old momenta are related by $\Pi = [B^{-1}]^T\mathbf{P} = B\mathbf{P}$. Meanwhile, the translation-invariant Hamiltonian (1.1) can be rewritten in the symmetrical form

$$H = \sum_{\substack{i,j=1\\ i<j}}^{N}\left\{\frac{1}{2Nm}(\mathbf{p}_i - \mathbf{p}_j)^2 + \gamma f\left(\frac{r_{ij}}{a}\right)\right\}. \qquad (2.3)$$

If we now compute expectations with respect to boson functions, we find from Eq. (2.3) that $\langle H\rangle = \langle\mathscr{H}\rangle$, where the reduced two-body Hamiltonian $\mathscr{H}$ is given by

 0022-2488/88/040990-05$02.50

$$\mathscr{H} = (N-1)\left(\frac{1}{2m}\Pi_2^2 + \frac{N}{2}\gamma f\left(\frac{\sqrt{2}\rho_2}{a}\right)\right). \qquad (2.4)$$

If Ψ represents a translation-invariant boson wave function then the lowest energy $\mathscr{E}$ of the N-particle system is given by the minimum Rayleigh quotient

$$\mathscr{E} = \inf_{\Psi}\,[(\Psi,\mathscr{H}\Psi)/(\Psi,\Psi)]\,. \qquad (2.5)$$

It is now convenient to define the dimensionless energy and coupling parameters E and v by

$$E = m\mathscr{E}a^2/[\hbar^2(N-1)], \quad v = m\gamma a^2 N/2\hbar^2. \qquad (2.6)$$

Equation (2.5) may then be further simplified and written in the form

$$E = F_N(v) = \inf_{\Psi}\,[(\Psi,\mathrm{H}\Psi)/(\Psi,\Psi)]\,, \qquad (2.7)$$

where the Hamiltonian H is defined in terms of the dimensionless variable $\mathbf{r} = (\mathbf{r}_1 - \mathbf{r}_2)/a = \sqrt{2}\rho_2/a$ and the Laplacian Δ with respect to $\mathbf{r}$ by

$$\mathrm{H} = -\Delta + vf(r), \qquad (2.8)$$

and Ψ is a translation-invariant N-boson wave function. We call the function $F_N(v)$ a *trajectory function* and the graph $(v,F_N(v))$, $v>0$, an *energy trajectory* for the N-boson problem. Since the permutation-symmetry constraint increases monotonically with N it is clear that, for each *fixed* v, the value of $F_N(v)$ increases monotonically with N, that is, $F_N(v) \geqslant F_M(v)$, $N>M$; consequently we have

$$F_2(v) \leqslant F_N(v) \leqslant F_\infty(v) \qquad (2.9)$$

provided the limit $N\to\infty$ exists. Keeping v constant means, of course, that the original coupling parameter γ must tend to zero as N increases. Boson systems with attractive pair potentials and a fixed coupling parameter collapse in the sense that $\mathscr{E}/N$ increases without bound as N increases.

We now look at variational estimates of the energy. If we could find a translation-invariant boson function with the single-product form

$$\Psi(\rho_2,\rho_3,\ldots,\rho_N) = \psi(\rho_2)g(\rho_3,\ldots,\rho_N), \qquad (2.10)$$

then, by substituting in the rhs of Eq. (2.7), we see that an upper bound to $F_N(v)$ is given by the Rayleigh quotient

$$F_\psi(v) = (\psi,\mathrm{H}\psi)/(\psi,\psi)\,. \qquad (2.11)$$

Expression (2.11) is exactly what we would use if we were to estimate variationally the bottom of the spectrum of H, a one-particle (or *reduced* two-particle) Hamiltonian. The catch in this is that (2.10) is a strong constraint for boson functions and it has in fact been proved[3] that the single-product form is achieved if and only if Ψ is a Gaussian function. However, in this case N disappears from the calculation and the result provides an upper trajectory bound valid for all N: After minimization with respect to scale, we call this $F_g(v)$. We can now summarize the results by writing

$$F_2(v) \leqslant F_N(v) \leqslant F_\infty(v) \leqslant F_g(v), \qquad (2.12)$$

where $F_g(v)$ is given by using $\psi(r) = e^{-\alpha r^2}$ in (2.11) and minimizing the resulting expression with respect to α. The corresponding energy bounds are recovered from the various trajectory functions by using the following general expression:

$$\mathscr{E} = (\hbar^2/ma^2)(N-1)F(m\gamma a^2 N/2\hbar^2). \qquad (2.13)$$

The inequalities (2.12) are important because they reduce the N-body energy problem, approximately, to a study of the single-particle operator H. It is now clear that the inequalities in (2.12) all collapse into equalities if and only if the potential has the harmonic oscillator shape $f(r) = r^2$; the common energy value in this case is $E = 3v^{1/2}$. In general, one has a family of nonintersecting trajectories labeled by N and bounded below by $F_2(v)$ and above by $F_\infty(v)$. An illustration of the family of trajectories for the exactly soluble delta potential (in one dimension) may be found in Ref. 1.

For a pure power-law potential shape $f(r)$ given by

$$f(r) = \mathrm{sgn}(\nu)r^\nu, \quad \nu \geqslant -1, \qquad (2.14)$$

simple scaling arguments show that the corresponding energy trajectories have the form

$$E = F_N^{(\nu)}(v) = F_N^{(\nu)}(1)v^{2/(\nu+2)}. \qquad (2.15)$$

Therefore, for each fixed power ν, the corresponding energy trajectories are all scaled images of any one of them, say, $N=2$.

III. A VARIATIONAL UPPER ENERGY BOUND

We now establish the following variational upper bound to $F_N(v)$:

$$F_N(v) \leqslant F_\phi(v) = \frac{1}{8}\int_{\mathbf{R}^3} \frac{(\nabla\phi(\mathbf{s}))^2}{\phi(\mathbf{s})}\,d^3\mathbf{s} + v\iint_{\mathbf{R}^6} \phi(\mathbf{s})f(|\mathbf{s}-\mathbf{s}'|)\phi(\mathbf{s}')d^3\mathbf{s}\,d^3\mathbf{s}', \qquad (3.1)$$

where the density function $\phi(\mathbf{s})$ satisfies the normalization condition

$$\int_{\mathbf{R}^3} \phi(\mathbf{s})\,d^3\mathbf{s} = 1.$$

Our main purpose in this section is to derive (3.1) from quantum mechanics and also to explain why it is that translation-invariant Gaussian wave functions used in the Rayleigh quotient $(\Psi, \mathrm{H}\Psi)/(\Psi,\Psi)$ and Gaussian densities used in (3.1) lead to precisely the same upper trajectory bound $F_g(v)$.

We find an upper estimate to the bottom of the spectrum of H by using a single-product boson wave function of the form

$$\Phi(\mathbf{r}_1, \mathbf{r}_2,\ldots,\mathbf{r}_N) = \varkappa(\mathbf{r}_1)\varkappa(\mathbf{r}_2)\cdots\varkappa(\mathbf{r}_N), \qquad (3.2)$$

where $\int_{\mathbf{R}^3}\varkappa^2(\mathbf{r})\,d^3\mathbf{r} = 1$. By using the symmetrical form (2.3) of the Hamiltonian H of relative motion and the permutation symmetry of Φ, we find

$$(\Phi,H\Phi) = (N-1)\iint_{\mathbf{R}^6} \varkappa(\mathbf{r})\varkappa(\mathbf{r}') \times\left[-\frac{\hbar^2}{4m}[\nabla_{\mathbf{r}} - \nabla_{\mathbf{r}'}]^2 + \frac{N}{2}\gamma f\left(\frac{|r-r'|}{a}\right)\right] \times\varkappa(\mathbf{r})\varkappa(\mathbf{r}')\,d^3\mathbf{r}\,d^3\mathbf{r}'. \qquad (3.3)$$

We now introduce the density function $\phi(\mathbf{s})$ defined by

$$a^{-3}\phi(\mathbf{r}/a) = \varkappa^2(\mathbf{r}). \tag{3.4}$$

Consequently,

$$\int_{\mathbf{R}^3} \varkappa^2(\mathbf{r}) d^3\mathbf{r} = \int_{\mathbf{R}^3} \phi(\mathbf{s}) d^3\mathbf{s} = 1, \quad \mathbf{s} = \mathbf{r}/a. \tag{3.5}$$

We let $\mathscr{E}$ represent the bottom of the spectrum of H and recall from (2.6) the definitions $F_N(v) = m\mathscr{E}a^2/[\hbar^2(N-1)]$ and $v = m\gamma a^2 N/2\hbar^2$ and note that the Laplacian ∇_r^2 has the scale transformation $\nabla_r^2 = a^{-2}\nabla_s^2$. We then drop the subscript s on the gradient ∇_s and note that $\nabla\phi = 2\varkappa\nabla\varkappa$ and also that, after integration, the cross terms in (3.3) involving $\nabla_r \cdot \nabla_{r'}$ all vanish. Having made these observations we may, with the aid of Stokes' theorem applied to the kinetic-energy term, transform (3.3) into the following form:

$$\begin{aligned} F_N(v) &= \frac{m\mathscr{E}a^2}{\hbar^2(N-1)} \\ &\leqslant \frac{m(\Phi, H\Phi)a^2}{\hbar^2(N-1)} \\ &= \frac{1}{8}\int_{\mathbf{R}^3} \frac{(\nabla\phi(\mathbf{s}))^2}{\phi(\mathbf{s})} d^3\mathbf{s} \\ &\quad + v \iint_{\mathbf{R}^6} \phi(\mathbf{s}) f(|\mathbf{s}-\mathbf{s}'|)\phi(\mathbf{s}') d^3\mathbf{s}\, d^3\mathbf{s}', \end{aligned} \tag{3.6}$$

which establishes the bound (3.1). Equation (3.6) is a curious result because, for a given fixed v, the bound does not depend on N. Hence the result is also an upper bound to the limiting trajectory $F_\infty(v)$ and can therefore only be close to the exact answer for *finite* N in cases where the trajectory functions F_N do not vary strongly with the particle number N. Similar upper bounds in which the center-of-mass energy has *not* been removed have been discussed by many authors: A recent exposition may be found in Ref. 4.

Now we turn to the special case of Gaussian functions. Let us suppose that $\varkappa(r) = C(\alpha)e^{-\alpha r^2}$ and $(\varkappa,\varkappa) = 1$, where α is a positive constant. Then, in terms of our orthogonal relative coordinates, we have by (2.1),

$$\begin{aligned} \Phi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) &= \Phi(\rho_1, \rho_2, \ldots, \rho_N) \\ &= \varkappa(\rho_1)\varkappa(\rho_2)\varkappa(\rho_3)\cdots\varkappa(\rho_N). \end{aligned} \tag{3.7}$$

Hence, apart from the additional factor $\varkappa(\rho_1)$, Φ has the special form (2.10) and consequently $ma^2(\Phi, H\Phi)/(N-1)\hbar^2 = (\varkappa, \mathrm{H}\varkappa)$, exactly as in (2.11). All the factors $\varkappa^2(\rho_i)$ for $i \neq 2$ integrate to unity and the result is the expectation of the one-body operator H with respect to a Gaussian function. The Gaussian is the only *translation-invariant* N-boson wave function with which it is comfortable to work and it leads to the same energy bound as (3.1) when a Gaussian density is used. The bound (3.1) is therefore superior because one can easily explore different densities ϕ and possibly improve on the special common Gaussian result. In the case of the harmonic oscillator with shape $f(r) = r^2$ and a Gaussian density, the bound (3.1), when minimized with respect to scale, yields the well-known exact energy (trajectory) $E = 3v^{1/2}$ of this particular N-boson system.

IV. THE LINEAR POTENTIAL

We now consider the linear potential shape given simply by

$$f(r) = r. \tag{4.1}$$

The general formula (2.15) for the energy trajectories of pure power-law potentials tells us that the energy trajectories for the linear potential (4.1) have the form

$$F_N(v) = F_N(1)v^{2/3}. \tag{4.2}$$

Meanwhile, from (2.12) and (3.1), we have the bounds

$$F_2(1) \leqslant F_N(1) \leqslant F_\phi(1). \tag{4.3}$$

The lower bound $F_2(1)$ is the bottom of the spectrum of the one-particle operator (2.8), which for the linear potential is given explicitly by

$$\mathrm{H} = -\Delta + r. \tag{4.4}$$

This well-known problem has ground-state energy given by the negative of the first zero of the Airy function. We truncate the decimal approximation of this zero so as to preserve the lower bound and therefore we find

$$F_2(1) > 2.338\,107. \tag{4.5}$$

The upper bound (3.1) requires a trial density ϕ. We first look at a one-parameter family of *central* densities on $\mathbf{R}^3$ given by

$$\phi(\mathbf{s}) = (4\pi b^3 I)^{-1} w(s/b), \quad s = |\mathbf{s}|, \tag{4.6}$$

where the *shape* $w(t)$ will be chosen later and the normalization integral I is given by

$$I = \int_0^\infty w(t)\, t^2\, dt. \tag{4.7}$$

If we now substitute (4.6) in (3.1), set $vf(r) = r$, and minimize the result with respect to the scale variable b we obtain the following formulas for the upper energy bound $F_\phi(1)$ and the optimal scale b:

$$E \leqslant F_\phi(1) = [3KV^2/8I^5]^{1/3}, \quad b = [3IK/8V]^{1/3}, \tag{4.8}$$

where the kinetic and potential energy integrals K and V are given by

$$K = \int_0^\infty \frac{(w'(t))^2}{w(t)} t^2\, dt \tag{4.9}$$

and

$$V = \int_0^\infty \int_t^\infty w(s)\, w(t)\, [st^4 + 3s^3t^2]\, ds\, dt. \tag{4.10}$$

For the potential energy, the integrals over the angles were first performed and then the remaining integrals were worked into the form (4.10) so as to avoid any integration over the absolute-value function.

We now choose the density shape $w(t)$ to be in a one-parameter family which includes the Gaussian $q = 2$ as a special case. Thus we set

$$w(t) = e^{-t^q}, \quad q > 0. \tag{4.11}$$

Consequently, the normalization and kinetic-energy integrals (4.7) and (4.9) become

$$I(q) = \Gamma(3/q)/q, \quad K(q) = (1+q)\Gamma(1+1/q). \tag{4.12}$$

For each q the potential energy integral (4.10) may now be performed numerically by Gaussian integration and the scale-optimized energy $E(q)$ can be calculated from (4.8). In Fig. 1 we exhibit a graph of $E(q)$ which has a minimum for q about 1.887. Thus we have

$$F_N(1) < E(1.887) < 2.343\,5154. \tag{4.13}$$

We have truncated the approximation in (4.13) so as to preserve the upper bound. In the case of the Gaussian density (or, equivalently, the Gaussian wave function) we can perform the upper-bound computation analytically to find the weaker result

$$E(2) = (81/2\pi)^{1/3} < 2.344\,78. \tag{4.14}$$

Graphs of the *radial* densities defined by

$$\phi(t) = [b^3 I]^{-1} t^2 e^{-(t/b)^q} \tag{4.15}$$

and normalized over $[0,\infty)$ are shown in Fig. 2 for $q = 1.887$ and 2.

With the notation $C(N)$ for $F_N(1)$, we can summarize our results for the linear potential by writing, from (2.13) and (4.2),

$$\mathscr{E} = C(N)(N-1)(\hbar^2/m)^{1/3}(\gamma N/2)^{2/3},$$

$$2.3381 < C(N) < 2.343\,52. \tag{4.16}$$

We have set $a = 1$ since the linear potential allows only one distinct parameter. The simple formula (4.16) determines the N-boson energy with an error of less than 0.116% for all $N \geqslant 2$.

V. POWER-LAW POTENTIALS

It is now straightforward to generalize the results obtained in Sec. IV to more general power-law potentials whose shapes are given by

$$f(r) = \operatorname{sgn}(\nu) r^\nu, \quad \nu \geqslant -1. \tag{5.1}$$

We shall consider this general problem as far as finding the basic upper and lower bounds by the equivalent two-body method and the use of a Gaussian trial function, and leave the final phase, the sharpening of the bounds by the method of Sec. IV, as a task to be carried out in answer to the needs of a specific application. As we shall see, even the bounds we obtain immediately are already very accurate.

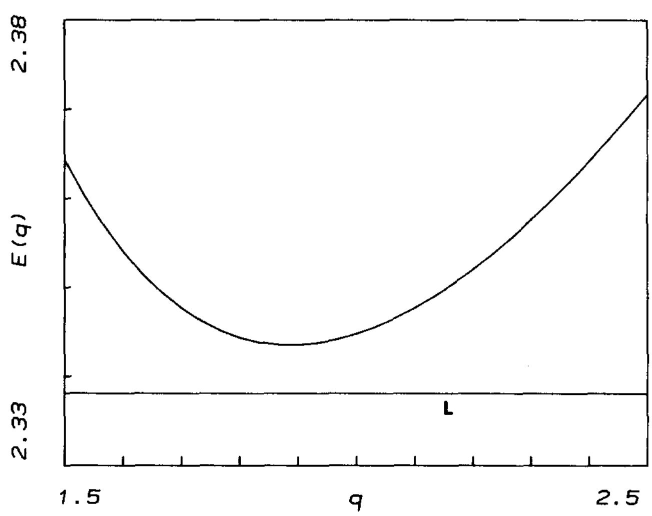


FIG. 1. The variational upper energy bound for the linear potential is first optimized with respect to the scale parameter b and the result $E(q)$ is then minimized with respect to the power parameter q. The graph shows $E(q)$ and also the constant lower energy bound $L = 2.3381$.

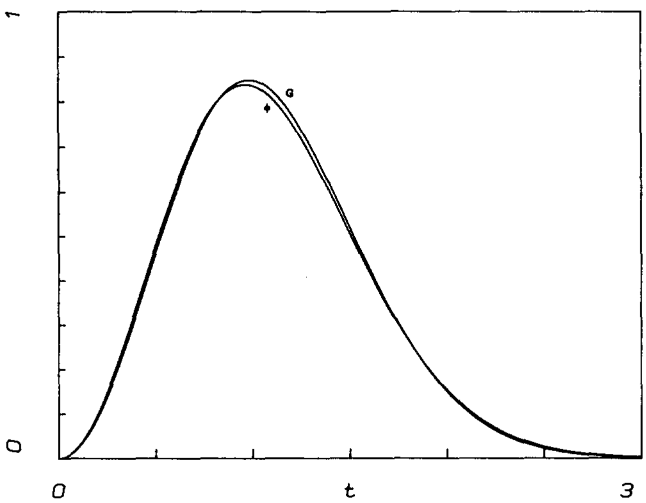


FIG. 2. The optimal *radial* density ϕ from the family (4.15), with $q = 1.887$ and the Gaussian density G, with $q = 2$.

From (2.12), (2.13), and (2.15) we see that the energy of the N-boson problem with the potential $V_{ij} = \gamma f(|\mathbf{r}_i - \mathbf{r}_j|)$ is given by

$$\mathscr{E} = C^{(\nu)}(N)(N-1)(\hbar^2/m)^{\nu/(\nu+2)}(\gamma N/2)^{2/(\nu+2)}, \tag{5.2}$$

where

$$F_2^{(\nu)}(1) \leqslant C^{(\nu)}(N) \leqslant F_g^{(\nu)}(1) \tag{5.3}$$

and $F_2^{(\nu)}(1)$ and $F_g^{(\nu)}(1)$ are, respectively, the bottom of the spectrum of the *one*-body operator

$$\mathbb{H} = -\Delta + \operatorname{sgn}(\nu) r^\nu \tag{5.4}$$

and the minimum of the expectation value $\langle \psi, \mathbb{H}\psi \rangle$ with respect to the scale of a normalized Gaussian trial function ψ. We have found the lower bounds by direct numerical integration. The Gaussian upper bounds can be shown to be given by the general formula

$$F_g^{(\nu)}(1) = \frac{3(\nu+2)}{4\nu} \times \left[\frac{|\nu| 2^{(\nu+4)/2} \Gamma((3+\nu)/2)}{3\pi^{1/2}} \right]^{2/(\nu+2)}. \tag{5.5}$$

We present our results for some powers ν in the range $0.1 \leqslant \nu \leqslant 1.0$ in Table I. We adopt the same style as we did for the linear potential, namely, we present lower and upper bounds to $C(N)$ and also an upper bound to the percentage error incurred if the mean of the bounds is used to estimate the energy of the N-boson problem; all decimals have been truncated in the appropriate directions to preserve the bounds. Thus, for these many-body problems, the energy is determined for all $N \geqslant 2$ by formula (5.2), with the error always less than 0.23%. The Gaussian trial function gives the worst results for this collection of potentials when ν is about $\frac{1}{2}$. By using the method of Sec. IV, the upper bounds can all be sharpened.

TABLE I. Lower and upper bounds to the coefficient $C(N)$ in the general formula (5.2) for the energy of the N-boson problem with pure power-law potentials.

ν	$F_2^{(\nu)}(1) < C(N) <$	$F_g^{(\nu)}(1)$	% error
0.1	1.2356	1.2382	0.1
0.2	1.4130	1.4176	0.17
0.3	1.5677	1.5741	0.2
0.4	1.7067	1.7143	0.22
0.5	1.8333	1.8416	0.23
0.6	1.9497	1.9582	0.22
0.7	2.0572	2.0657	0.21
0.8	2.1572	2.1653	0.19
0.9	2.2505	2.2581	0.17
1.0	2.3381	2.3448	0.15

VI. CONCLUSION

The energy upper bound (3.1), which can also be understood in terms of a field-theoretic analysis of the many-body problem, the so-called *collective field theory*,[1] was derived here by a simple variational argument based on single-product wave functions. This general approach to an upper bound is, of course, as old as wave mechanics itself. We are pleased to find that it is possible to work directly with the translation-invariant Hamiltonian: Contributions from the kinetic energy of the center of mass do *not* have to be accounted for specially or estimated in the large-N limit. It is interesting that the upper bound with a Gaussian density ϕ is the same as that given by the Rayleigh quotient $(\Psi, H\Psi)/(\Psi, \Psi)$, in which Ψ is a Jastrow translation-invariant Gaussian wave function. The upper bound (3.1) has the advantage that one can explore other shapes of density and possibly improve the energy estimate, as we have been able to do here for the linear potential.

ACKNOWLEDGMENTS

The author would like to thank the Natural Sciences and Engineering Research Council of Canada for partial financial support of this work under Grant No. A3438.

[1]R. L. Hall, Phys. Rev. A **36**, 4014 (1987); "Complementary pictures of the N-boson problem," to appear in Phys. Rev. A.
[2]R. L. Hall, J. Math. Phys. **24**, 324 (1983); **25**, 2708 (1984).
[3]R. L. Hall, Can. J. Phys. **50**, 305 (1972); Aequ. Math. **8**, 21 (1972).
[4]J. P. Blaizot and G. Ripka, *Quantum Theory of Finite Systems* (MIT, Cambridge, MA, 1986), p. 181.

Renewal theory for transport processes in binary statistical mixtures

C. D. Levermore and J. Wong
Lawrence Livermore National Laboratory, Livermore, California 94550
G. C. Pomraning
School of Engineering and Applied Science, University of California, Los Angeles, Los Angeles, California 90024

(Received 22 September 1987; accepted for publication 2 December 1987)

Renewal theory is used to analyze linear particle transport without scattering in a random mixture of two immiscible fluids, with the statistics described by arbitrary (non-Markovian) fluid chord length distributions. One conclusion (for unimodal distributions) that is drawn is that the mean and variance of the chord length distributions through each fluid is sufficient knowledge of the statistics to give a reasonably accurate description of the ensemble averaged intensity. Expressions for effective cross sections and an effective source to be used in the usual deterministic transport equation are also obtained. The use of these effective quantities allows statistical information to be introduced very simply into a standard transport equation. An analysis is given which shows how the transport description, including scattering, in a Markovian mixture can be modified to yield an approximate description of transport in a non-Markovian mixture. Numerical results are given to assess the accuracy of this model, as well as the simpler model involving the effective cross sections and source.

I. INTRODUCTION

A subject of recent interest has been the problem of describing particle transport in a statistical medium consisting of two randomly mixed immiscible fluids.[1–4] The first paper in this series of four by Levermore *et al.*[1] considered the very simple linear kinetic (transport) equation along a ray described by the distance x given by

$$\frac{d\psi(x)}{dx} + \sigma(x)\psi(x) = s(x). \tag{1}$$

Here $\psi(x)$ is the distribution function for the particle density, or intensity; $\sigma(x)$ is the collision (absorption) cross section; and $s(x)$ is the external source strength. The quantities σ and s in Eq. (1) were treated as random variables, each one taking one of two values at any point x. These two values, independent of position, are the values associated with each fluid component of the mixture. That is, as a particle traverses the mixture along the path x, it sees alternating packets of the two fluids, which we label by $i = 0,1$, and each fluid has a definite, spatially independent, value of σ and s, labeled σ_i and s_i. The transition from one fluid to the other was assumed to be a homogeneous Markov process. This implies that the chord length along x of fluid i follows a Poisson distribution; i.e., the chord length of each fluid is exponentially distributed. Using a projection operator technique, the method of smoothing as described by Keller[5–7] and Frisch,[8] an exact solution was obtained for $\Psi(x)$, the ensemble averaged value of the intensity, corresponding to Eq. (1).

The problem described above was revisited by Vanderhaegen,[2] in which many of the results given by Levermore *et al.*[1] were reproduced. The new contributions of this paper were two. First, it was shown that this problem could be addressed by the use of a Chapman–Kolmogorov, or master, equation for the joint probability density.[8–10] Second, by considering the isotropic diffusion limit of the transport equation given by Eq. (1), it was shown that the concept of an effective cross section σ_{eff} arises naturally. That is, in this diffusion limit, the stochastic problem can be replaced with a standard deterministic problem, involving a single effective cross section made up of σ_0, σ_1, and the two mean chord lengths of the fluid components, λ_0 and λ_1. It is worth noting, however, that this deterministic diffusion description was obtained under the restrictive assumption that the ratio s/σ is not stochastic, i.e., $s_0/\sigma_0 = s_1/\sigma_1$.

It is clear that these two papers described above treated a very simple situation as described by Eq. (1). In particular, Eq. (1) is time independent and, more significantly, neglects the scattering of particles. Furthermore, both of these papers assumed homogeneous Markov statistics and additionally assumed that σ_0 and σ_1 are spatially independent. Levermore[3] recognized that the treatment of Vanderhaegen could be extended by a master equation approach to analyze the general linear kinetic equation, including time dependence, scattering, and spatially dependent cross sections and sources. Levermore was also able to analyze inhomogeneous (spatially dependent) statistics, but it must be emphasized that this work was still restricted to Markovian statistics. The transport equation treated by Levermore[3] is, with the assumption of isotropic scattering,

$$\frac{1}{c}\frac{\partial\psi}{\partial t} + \mathbf{\Omega}\cdot\nabla\psi + \sigma\psi = \frac{c\sigma_s}{4\pi}e + s, \tag{2}$$

where

$$e = \frac{1}{c}\int_{4\pi} d\mathbf{\Omega}\,\psi. \tag{3}$$

Here t denotes time; c is the particle speed; $(1/c)\partial/\partial t + \mathbf{\Omega}\cdot\nabla = d/dx$ is the advection operator with $\mathbf{\Omega}$ denoting the particle flight direction and x parametrizing the space-time ray; and σ_s is the scattering cross section.

The quantities σ, σ_s, and s in Eq. (2), and hence ψ and e, are random variables associated with the two-fluid random mixture. Under the assumption of inhomogeneous Markov statistics, Levermore[3] used the master equation formalism to obtain a description for Ψ, the ensemble average of ψ, as the solution of the two coupled equations

$$\left(\frac{1}{c}\frac{\partial}{\partial t}+\mathbf{\Omega}\cdot\nabla\right)\begin{bmatrix}\Psi\\ \chi\end{bmatrix}+\begin{bmatrix}\Sigma & \nu\\ \nu & \hat{\Sigma}\end{bmatrix}\begin{bmatrix}\Psi\\ \chi\end{bmatrix}=\frac{c}{4\pi}\begin{bmatrix}\Sigma_s & \nu_s\\ \nu_s & \hat{\Sigma}_s\end{bmatrix}\begin{bmatrix}E\\ \eta\end{bmatrix}+\begin{bmatrix}S\\ T\end{bmatrix}. \tag{4}$$

Equation (4) is known to be rigorously correct in the case of time independent transport in general geometry with no scattering. In this instance, Eq. (4) reduces to an infinite uncoupled set of ordinary differential equations, for which the master equation approach is well understood.[10] In the more general time-dependent multidimensional case including scattering, this transport description must at this time be considered a phenomenological model. This model, while not rigorously derived, is a reasonable model since it reduces to the proper result in all known limits, and is robust away from these limits. A much more detailed discussion of Eq. (4) in this regard will be given shortly in a forthcoming article.[3] The new results given in this paper use Eq. (4) only in the time-independent, purely absorbing (no scattering) context. As just noted, in this limit Eq. (4) is well founded. Our only reason for including Eq. (4) here in full generality (multidimensional time-dependent transport with scattering) is to make this introduction a complete summary of work done to date. The quantity Ψ in Eq. (4) is the ensemble-averaged intensity corresponding to Eq. (2), and χ is an auxiliary function describing a cross correlation between the random variables ψ and σ. Corresponding to Ψ and χ are the angular integral quantities

$$E=\frac{1}{c}\int_{4\pi}d\mathbf{\Omega}\,\Psi;\quad \eta=\frac{1}{c}\int_{4\pi}d\mathbf{\Omega}\,\chi. \tag{5}$$

The other quantities in Eq. (4) are given in terms of the fluid properties (σ_i, σ_{si}, s_i) and the Markov statistical properties (p_i, λ_c) as

$$S=p_0s_0+p_1s_1;\quad T=(p_0p_1)^{1/2}(s_0-s_1), \tag{6}$$

$$\Sigma=p_0\sigma_0+p_1\sigma_1;\quad \hat{\Sigma}=p_1\sigma_0+p_0\sigma_1+\lambda_c^{-1}, \tag{7}$$

$$\Sigma_s=p_0\sigma_{s0}+p_1\sigma_{s1};\quad \hat{\Sigma}_s=p_1\sigma_{s0}+p_0\sigma_{s1}, \tag{8}$$

$$\nu=(p_0p_1)^{1/2}(\sigma_0-\sigma_1);\quad \nu_s=(p_0p_1)^{1/2}(\sigma_{s0}-\sigma_{s1}). \tag{9}$$

The statistical quantity $p_i(\mathbf{r},t)$ is the probability of finding material i at position $\mathbf{r}$ and time t. The second statistical quantity $\lambda_c(\mathbf{r},t,\mathbf{\Omega})$ is a correlation length for the inhomogeneous Markov statistics. Its definition follows by considering $P_{ij}(x,y)$, the conditional probability of finding material i at position x along a ray, given that material j is at position y along the same ray. For inhomogeneous Markov statistics, this conditional probability can be written[3]

$$P_{ij}(x,y)=p_i(x)+(-1)^{i+j}[q(x,y)/p_j(y)]e^{-\Gamma(x,y)}, \tag{10}$$

where

$$q(x,y)=[p_0(x)p_1(x)p_0(y)p_1(y)]^{1/2}, \tag{11}$$

and

$$\Gamma(x,y)=\left|\int_y^x d\xi\,\lambda_c^{-1}(\xi)\right|. \tag{12}$$

The function $\lambda_c(x)$ in Eq. (12) is an arbitrary (subject to minimal constraints) function, whose functional dependence in any given instance depends upon the details of the statistics. In view of the way λ_c appears in Eq. (10), it is logically identified as a correlation length. For homogeneous Markov statistics, one has[1,3]

$$p_i=\lambda_i/(\lambda_0+\lambda_1);\quad 1/\lambda_c=1/\lambda_0+1/\lambda_1, \tag{13}$$

where, as defined earlier, λ_i is the mean chord length in material i associated with the exponential distribution of chord lengths.

For Markov statistics, Eq. (4) gives a complete description for linear transport through a two-fluid mixture, with the underlying kinetic equation given by Eq. (2). The remaining generalization required is the removal of the restriction of Markov statistics. This generalization was suggested by Vanderhaegen,[4] under the simplifying assumptions of no scattering, homogeneous statistics, a spatially constant cross section σ_i, and the ratio s/σ not stochastic, i.e., $s_0/\sigma_0=s_1/\sigma_1$. Vanderhaegen assumed that the chord length for each fluid in the two fluid mixture, along any ray x, is described by an arbitrary chord length distribution, with this distribution independent of position along the ray. Homogeneous Markov statistics is a special case of this corresponding to an exponential chord length distribution for each fluid. It was shown that the transport problem using arbitrary chord length distributions could be formulated exactly using renewal theory.[11] More precisely, the situation treated by Vanderhaegen is generally called an alternating renewal process,[11] and we shall refer to it simply as renewal statistics. The resulting integral equations are of the convolution type, and easily solved by Laplace transformation. By again considering the diffusion limit, Vanderhaegen showed that one can obtain a simple deterministic description involving an effective cross section σ_{eff} based upon σ_0, σ_1, and a simple integral quantity involving the chord length distribution. This generalized his earlier use of diffusion theory to obtain σ_{eff} for a Markov (exponential distribution) process.[2]

In the present paper, we will show that the renewal analysis of Vanderhaegen can be generalized to include inhomogeneous statistics, spatially varying cross sections, and general sources, not restricted by $s_0/\sigma_0=s_1/\sigma_1$. We use this generalized analysis to investigate several specific areas. We first asked the question as to how important are the details of the statistics, i.e., the chord length distributions, on the transport description. By considering various distributions that are unimodal with the same mean, we found that the mean alone is not sufficient statistical information to characterize the chord length distribution; the solution for Ψ depends significantly upon which distribution is used. We then asked if knowing the mean and the variance was a sufficient characterization of the chord length distribution. Here we find that for the distributions investigated, the solution for Ψ

is relatively insensitive to which distribution is used, provided that the mean and the variance of the distributions are the same.

By solving the general renewal equations in the special case of spatially homogeneous statistics and fluid properties σ_i and s_i independent of position, we are able to obtain expressions for an effective cross section σ_{eff}, and an effective source S_{eff}, which can be used in a deterministic transport equation of the usual form. Simple expressions for these effective quantities are given for general renewal statistics, i.e., for an arbitrary chord length distribution. The expression for σ_{eff} agrees with that given by Vanderhaegen[4] from his diffusion limit argument. In the special case of Markov statistics, our results for σ_{eff} and S_{eff} agree with those obtained from a certain asymptotic limit of Eq. (4), as given by Levermore.[3] We give numerical results to assess the accuracy of using a standard transport equation involving σ_{eff} as an approximate treatment of transport through a random medium.

We also show that the Markov equations given by Eq. (4) can be used to describe non-Markov renewal statistics in a well defined approximate way by replacing λ_c in Eq. (7) with an effective correlation length, which we denote by λ_{eff}. Specifically, these equations treat non-Markov statistics in the sense that they yield the proper expressions for σ_{eff} and S_{eff} when analyzed asymptotically following Levermore,[3] or are solved in the spatially homogeneous case as alluded to in the preceding paragraph. Thus the Markov equations given by Eq. (4), with λ_c taken as λ_{eff}, can be considered as a reasonable approximate description of transport in a non-Markovian (renewal) two-fluid mixture, including scattering. The accuracy of this approximation is assessed by comparing the solution of Eq. (4), with λ_c replaced by λ_{eff}, to the exact solution based upon the renewal equations for various non-Markov statistics. We find that Eq. (4) does a reasonable job in reproducing the exact results.

The details of the investigations we have just outlined are given in the next two sections of this paper. Section II is devoted to the renewal analysis, and Sec. III gives various numerical results. Section IV gives a few concluding remarks, including what we feel are additional needed areas of investigation in this continuing study of linear transport through a statistical mixture of immiscible fluids.

II. RENEWAL ANALYSIS

The equation we analyze is the simple one given by Eq. (1) on the interval $0 \leqslant x < \infty$. We imagine the entire line $-\infty < x < \infty$ populated statistically with alternating segments of two materials labeled 0 and 1. The point $x = 0$ is chosen at random on this infinite line. We assign a stochastic boundary condition to Eq. (1) of the form

$$\psi(0) = \begin{cases} \psi_0 & \text{if } x = 0 \text{ is in material } 0, \\ \psi_1 & \text{if } x = 0 \text{ is in material } 1. \end{cases} \tag{14}$$

We define $p_i(x)$ as the probability of finding material i at position x. We also define two conditional probabilities, namely,

$$Q_i(y,x) = \text{Prob}\{[y,x] \text{ is in } i | x \text{ is in } i \text{ and } x + dx \text{ is not}\}, \tag{15}$$

and

$$R_i(y,x) = \text{Prob}\{[y,x] \text{ is in } i | x \text{ is in } i\}. \tag{16}$$

From these definitions, we see that $Q_i(y,x)$ is the probability that the entire interval $[y,x]$ is in material i given that x is the right boundary of material i. By contrast, $R_i(y,x)$ is the probability that the entire interval $[y,x]$ is in material i, given that x is anywhere in material i. These probabilities describe in general inhomogeneous (spatially dependent) statistics. If the statistics are homogeneous, then p_i is independent of x, and the two conditional probabilities defined by Eqs. (15) and (16) depend only upon the single displacement argument, $x - y$.

Associated with each material, we assign a cross section $\sigma_i(x)$ and an external source $s_i(x)$ which, as indicated, are in general position dependent. We let $\phi_i[y,x;b]$ denote the solution of Eq. (1) at x given that the solution is b at y, and given that the entire interval $[y,x]$ consists of material i. This is a nonstochastic solution which is easily found from Eq. (1) to be

$$\phi_i[y,x;b] = be^{-\tau_i(y,x)} + \int_y^x dx'\, s_i(x')e^{-\tau_i(x',x)}, \tag{17}$$

where

$$\tau_i(y,x) = \int_y^x d\xi\, \sigma_i(\xi). \tag{18}$$

Given this definition of $\phi_i[y,x;b]$, we can write the general renewal equation

$$\Phi_0(x) = \phi_0[0,x;\psi_0]Q_0(0,x) + \int_0^x d_y Q_0(y,x)\phi_0[y,x;\Phi_1(y)], \tag{19}$$

where $\Phi_i(x)$ is the expected value of the random variable $\psi(x)$, given that x is the right-hand end point of material i. [The subscript y on the differential $d_y Q_0$ means that this differential only acts on the y variable of $Q_0(y,x)$.] The first term on the right-hand side of Eq. (19) is the contribution to $\Phi_0(x)$ if the interval $[0,x]$ is entirely made up of material 0, and the integral term accounts for the possibility that this is not so. A second equation complementary to Eq. (19) follows by interchanging the indices 0 and 1.

Equation (19) and that obtained by interchanging the indices represent two equations for the two unknowns $\Phi_0(x)$ and $\Phi_1(x)$. These can, in principle, be solved once the statistics, as reflected in the functions $Q_i(y,x)$, have been specified. However, knowing the $\Phi_i(x)$ is not sufficient since this function is only defined at the right-hand end point of material i. What is required is an analogous function defined for any point x. This function, which we denote by $\Psi_i(x)$, follows from the auxiliary renewal equation

$$\Psi_0(x) = \phi_0[0,x;\psi_0]R_0(0,x) + \int_0^x d_y R_0(y,x)\phi_0[y,x;\Phi_1(y)]. \tag{20}$$

Here $\Psi_i(x)$ is the expected value of the random variable $\psi(x)$, given that x is (anywhere) in material i. The corresponding equation for $\Psi_1(x)$ is found from Eq. (20) by interchanging the two indices. As with Eq. (19), the first term on the right-hand side of Eq. (20) accounts for the possibility that the interval $[0,x]$ is composed entirely of material 0, and the integral term accounts for the possibility of one or more material boundaries in $[0,x]$. Once the $\Phi_i(x)$ are obtained from Eq. (19) and its complement, these solutions can be inserted into Eq. (20) and its complement, and these can then in principle be simply evaluated for the $\Psi_i(x)$. The expected value of the random variable $\psi(x)$, which we denote by $\Psi(x)$, is then found by weighting the $\Psi_i(x)$ by the probability that point x is in material i, namely,

$$\Psi(x) = p_0\Psi_0(x) + p_1\Psi_1(x). \tag{21}$$

Equations (19) and (20) and their complementary equations found by interchanging the indices 0 and 1 then represent our generalization of the renewal equations given by Vanderhaegen[4] to include inhomogeneous statistics and spatially dependent cross sections $\sigma_i(x)$ and external sources $s_i(x)$.

The special case of homogeneous statistics is described entirely by the chord length distribution function $f_i(z)$, such that $f_i(z)dz$ is the probability that material i has a chord length between z and $z+dz$. This distribution function is independent of position x on the line $-\infty < x < \infty$. That is, the length of any segment of material i on this infinite line is chosen from the same distribution $f_i(z)$. If we define

$$Q_i(z) = \int_z^\infty dz' f_i(z'), \tag{22}$$

then $Q_i(z)$ is interpreted as the probability of a given segment of material i exceeding a length z. The average chord length in material i, which we denote by λ_i, is given by

$$\lambda_i = \int_0^\infty dz\, z f_i(z) = \int_0^\infty dz\, Q_i(z). \tag{23}$$

If we define

$$R_i(z) = \frac{1}{\lambda_i}\int_z^\infty dz'\, Q_i(z'), \tag{24}$$

it is easily argued that $R_i(z)$ is the probability that the right-hand boundary of a segment of material i is a distance greater than z from an arbitrary (random) point in the segment. For homogeneous statistics, the conditional probabilities defined for general inhomogeneous statistics by Eqs. (15) and (16) are given in terms of the single variable quantities defined by Eqs. (22)–(24) as

$$Q_i(y,x) = Q_i(x-y), \tag{25}$$

$$R_i(y,x) = R_i(x-y). \tag{26}$$

Additionally, for homogeneous statistics the probability p_i of finding material i at any point on the line $-\infty < x < \infty$ is given by

$$p_i = \lambda_i/(\lambda_0 + \lambda_1). \tag{27}$$

Homogeneous Markov statistics as treated earlier[1] correspond to a simple exponential distribution, i.e.,

$$\lambda_i f_i(x) = Q_i(x) = R_i(x) = e^{-x/\lambda_i}. \tag{28}$$

A periodic medium, in which case each segment of material i has the same length λ_i corresponds to $f_i(x) = \delta(x-\lambda_i)$, and hence

$$Q_i(x) = 1, \quad R_i(x) = (\lambda_i - x)/\lambda_i, \quad x < \lambda_i, \tag{29}$$

$$Q_i(x) = R_i(x) = 0, \quad x > \lambda_i. \tag{30}$$

A family of chord length distributions we will consider in our numerical calculations reported in the next section is the two-parameter gamma distribution given by

$$f_i(x) = \left[m_i(m_i x/\lambda_i)^{(m_i-1)}/\lambda_i\Gamma(m_i)\right] \times \exp(-m_i x/\lambda_i), \tag{31}$$

where $\Gamma(z)$ is the gamma function. Here λ_i is the mean of the distribution as given by Eq. (23), and the parameter m_i is related to variance according to $V_i = \lambda_i^2/m_i$ The special cases $m_i = 1$ and $m_i = \infty$ correspond to Markov (exponential) and periodic statistics, as given by Eqs. (28) and (29), (30), respectively. Three other simple chord length distributions we will use in the next section are the "block," defined by

$$f_i(x) = \begin{cases} (2\lambda_i)^{-1}, & x < 2\lambda_i, \\ 0, & x > 2\lambda_i, \end{cases} \tag{32}$$

the "tent," defined by

$$f_i(x) = \begin{cases} x/\lambda_i^2, & x < \lambda_i \\ (2\lambda_i - x)/\lambda_i^2, & \lambda_i < x < 2\lambda_i, \\ 0, & x > 2\lambda_i, \end{cases} \tag{33}$$

and the "ramp," defined by

$$f_i(x) = \begin{cases} 8x/(9\lambda_i^2), & x < (3/2)\lambda_i, \\ 0, & x > (3/2)\lambda_i. \end{cases} \tag{34}$$

If we specialize the general renewal equations given by Eqs. (19) and (20) to the case of homogeneous statistics, we find, after interchanging orders of integration in the double integral that arises from explicit use of Eq. (17) and subsequently performing the inner integration,

$$\begin{aligned}\Phi_0(x) = {} & \psi_0 e^{-\sigma_0 x} Q_0(x) \\ & + \int_0^x d_y Q_0(x-y) e^{-\sigma_0(x-y)}\Phi_1(y) \\ & + \int_0^x dy\, e^{-\sigma_0(x-y)} s_0(y) Q_0(x-y),\end{aligned} \tag{35}$$

$$\begin{aligned}\Psi_0(x) = {} & \psi_0 e^{-\sigma_0 x} R_0(x) \\ & + \int_0^x d_y R_0(x-y) e^{-\sigma_0(x-y)}\Phi_1(y) \\ & + \int_0^x dy\, e^{-\sigma_0(x-y)} s_0(y) R_0(x-y),\end{aligned} \tag{36}$$

with two complementary equations obtained by interchanging the indices in Eqs. (35) and (36). In writing Eqs. (35) and (36) we have allowed the external source $s_0(x)$ to be spatially dependent, but have taken the cross section σ_0 to be constant, independent of x. In this case Eqs. (35) and (36) and their complements are easily solved by Laplace transformation since the integral terms are of the convolution form. In the more general case of a spatially dependent cross section, the exponential terms in these two equations would be replaced by the corresponding exponentials involving the optical depths, i.e.,

$$e^{-\sigma_0(x-y)} \to e^{-\tau_0(y,x)}, \tag{37}$$

with $\tau_i(y,x)$ given by Eq. (18).

As noted above, the general solution of Eqs. (35) and (36) is easily found. However, this general solution is algebraically complex, and we give here only the solution we will need to define an effective cross section σ_{eff}. This solution corresponds to the source-free ($s_i = 0$) case with a nonstochastic boundary condition $\psi_0 = \psi_1$. If we define the Laplace transform of any function $h(x)$ by $\tilde{h}(r)$, i.e.,

$$\tilde{h}(r) = \int_0^x dx\, h(x) e^{-rx}, \tag{38}$$

we then find for this special case

$$\tilde{\Psi}(r) = p_0 \tilde{\Psi}_0(r) + p_1 \tilde{\Psi}_1(r) = \frac{(r + p_0\sigma_1 + p_1\sigma_0)}{(r+\sigma_0)(r+\sigma_1)} - \frac{p_0 p_1 (\sigma_0 - \sigma_1)^2}{(r+\sigma_0)(r+\sigma_1)\lambda_c} \cdot \left[\frac{\tilde{Q}_0 \tilde{Q}_1}{(r+\sigma_0)\tilde{Q}_0 + (r+\sigma_1)\tilde{Q}_1 - (r+\sigma_0)(r+\sigma_1)\tilde{Q}_0\tilde{Q}_1} \right], \tag{39}$$

where the argument of Q_i in Eq. (39) is $(r + \sigma_i)$, and we have defined

$$1/\lambda_c = 1/\lambda_0 + 1/\lambda_1. \tag{40}$$

In writing Eq. (39) we have taken a unit incoming intensity, i.e., $\psi_0 = \psi_1 = 1$. Knowing the statistics, one could in principle compute the $Q_i(r + \sigma_i)$, and a Laplace inversion of Eq. (39) would then give $\Psi(x)$, the solution to the problem.

We choose here to use Eq. (39) in another way. Specifically, if we compute l, the average distance to collision for a particle in this half-space, we have

$$l = \int_0^\infty dx\, x \left| \frac{d\Psi(x)}{dx} \right| = \int_0^\infty dx\, \Psi(x) = \tilde{\Psi}(0). \tag{41}$$

Setting $r = 0$ in Eq. (39) we find

$$l = \tilde{\Psi}(0) = \frac{1 + (p_0\sigma_1 + p_1\sigma_0) q \lambda_c}{p_0\sigma_0 + p_1\sigma_1 + \sigma_0\sigma_1 q \lambda_c}, \tag{42}$$

where we have defined

$$q = \frac{1}{\sigma_0} \left[\frac{1}{\tilde{Q}_0(\sigma_0)} - \frac{1}{\lambda_0} \right] + \frac{1}{\sigma_1} \left[\frac{1}{\tilde{Q}_1(\sigma_1)} - \frac{1}{\lambda_1} \right] - 1. \tag{43}$$

We now approximate the exact solution for $\Psi(x)$ by a single decaying exponential $\Psi_a(x)$ according to

$$\Psi_a(x) = e^{-\sigma_{\text{eff}} x}, \tag{44}$$

and define σ_{eff} such that $\Psi_a(x)$ gives the correct mean distance to collision. Use of Eq. (44) for $\Psi(x)$ in Eq. (41) gives $l = 1/\sigma_{\text{eff}}$, and equating this to Eq. (42) we find

$$\sigma_{\text{eff}} = \frac{p_0\sigma_0 + p_1\sigma_1 + \sigma_0\sigma_1 q \lambda_c}{1 + (p_0\sigma_1 + p_1\sigma_0) q \lambda_c}. \tag{45}$$

From the physics of the problem, l is non-negative for any statistics, and hence σ_{eff} given by Eq. (45) is clearly non-negative. We will also shortly prove this directly by showing that $q \geqslant 0$ for all chord length distribution functions. Equation (45) is the same result obtained by Vanderhaegen[4] from diffusion limit considerations. In the case of Markov statistics given by Eq. (28), we find $q = 1$ and Eq. (45) reduces to the result found by Levermore[3] from an asymptotic analysis of Eq. (4).

To obtain a corresponding expression for an effective source, we take s_i in Eqs. (35) and (36) to be constant, independent of x, and seek the deep-in ($x \gg 1$) solution. This solution will be a constant, independent of x. Omitting the algebraic details, we find

$$\Psi(x = \infty) = \frac{p_0 s_0 + p_1 s_1 + (p_0\sigma_1 s_0 + p_1\sigma_0 s_1) q \lambda_c}{p_0\sigma_0 + p_1\sigma_1 + \sigma_0\sigma_1 q \lambda_c}, \tag{46}$$

with q given once again by Eq. (43). We now define an effective source S_{eff} by the equation

$$\Psi(x = \infty) = S_{\text{eff}} / \sigma_{\text{eff}}. \tag{47}$$

Since σ_{eff} is non-negative, and from the physics so is $\Psi(x = \infty)$, we conclude that S_{eff} is non-negative for any renewal statistics. The motivation for this definition of S_{eff} is that a usual transport equation employing σ_{eff} and S_{eff} (to be written shortly) will give the correct deep-in solution in the case of homogeneous non-Markov statistics when σ_i and s_i are spatially independent. From Eqs. (45)–(47), we then deduce

$$S_{\text{eff}} = \frac{p_0 s_0 + p_1 s_1 + (p_0\sigma_1 s_0 + p_1\sigma_0 s_1) q \lambda_c}{1 + (p_0\sigma_1 + p_1\sigma_0) q \lambda_c}. \tag{48}$$

We note the similarity of Eqs. (45) and (48). In particular, if the source and cross section are related according to $s_0/\sigma_0 = s_1/\sigma_1 = B$, we then find

$$S_{\text{eff}} = B \sigma_{\text{eff}}. \tag{49}$$

In the Markov case given by Eq. (28), we have $q = 1$ and Eq. (48) reduces to the result obtained earlier by Levermore[3] from an asymptotic analysis of Eq. (4).

In terms of the effective quantities σ_{eff} and S_{eff} given by Eqs. (45) and (48), we suggest

$$\frac{d\Psi(x)}{dx} + \sigma_{\text{eff}}(x) \Psi(x) = S_{\text{eff}}(x), \tag{50}$$

as the simplest transport equation which incorporates non-Markov statistical effects into a standard transport equation. The x dependences of σ_{eff} and S_{eff} in Eq. (50) arise from using local values of the quantities on the right-hand sides of Eqs. (45) and (48). Equation (50) is a robust equation in that both σ_{eff} and S_{eff} are non-negative, and it has the property of predicting the correct mean distance to collision as well as the deep-in solution for homogeneous non-Markov statistics when σ_i and s_i are spatially independent. We note that as $\sigma_i \lambda_c$ tends to zero for both $i = 0$ and $i = 1$, Eqs. (45) and (48) predict the so-called atomic mix limit

$$\sigma_{\text{eff}} = p_0\sigma_0 + p_1\sigma_1 = \Sigma, \tag{51}$$

$$S_{\rm eff} = p_0 s_0 + p_1 s_1 = S, \tag{52}$$

which is the expected result. In the next section of this paper, we give a few numerical results to assess the accuracy of Eq. (50) as an approximation to the true situation.

One very major drawback of Eq. (50) is that it does not contain scattering. This motivates us to consider a second independent procedure for deriving a standard transport equation involving effective quantities to take statistical effects into account. This procedure includes scattering and, as we shall see, reduces to Eq. (50) in the absence of scattering. Further, it yields, in addition to a standard scalar transport equation with effective quantities, a higher level approximation in the form of a two-component vector description, as is Eq. (4) for Markov statistics. In fact, this scheme is based upon a modification of the exact Markov description given by Eq. (4). Specifically, we assume that Eq. (4) holds (approximately) in the non-Markov case by replacing λ_c in Eq. (7) by $\lambda_{\rm eff}$, an effective correlation length. We define $\lambda_{\rm eff}$ by

$$\lambda_{\rm eff} = q\lambda_c, \tag{53}$$

where q is given by Eq. (43) and λ_c in Eq. (53) is given by Eq. (40). In both Eqs. (40) and (43), λ_i is the mean chord length for material i for whatever non-Markov (renewal) statistics is obeyed by the mixture.

The motivation behind the definition of $\lambda_{\rm eff}$ as given by Eq. (53) is found in an asymptotic analysis of Eq. (4), with λ_c in Eq. (7) replaced with $\lambda_{\rm eff}$. The scaling for this analysis corresponds physically to a small amount of large cross-section material admixed with a large amount of small cross-section material. The result of this scaling, in low order, leads to the renormalized transport equation[3]

$$\frac{1}{c}\frac{\partial\Psi}{\partial t} + \mathbf{\Omega}\cdot\nabla\Psi + \sigma_{\rm eff}\Psi = \frac{c}{4\pi}\sigma_{s,\rm eff}E + S_{\rm eff}, \tag{54}$$

where

$$\sigma_{\rm eff} = \frac{p_0\sigma_0 + p_1\sigma_1 + \sigma_0\sigma_1\lambda_{\rm eff}}{1 + (p_0\sigma_1 + p_1\sigma_0)\lambda_{\rm eff}}, \tag{55}$$

$$\sigma_{s,\rm eff} = \sigma_{\rm eff} - \sigma_{a,\rm eff}, \tag{56}$$

$$\sigma_{a,\rm eff} = \frac{p_0\sigma_{a0} + p_1\sigma_{a1} + \sigma_{a0}\sigma_{a1}\lambda_{\rm eff}}{1 + (p_0\sigma_{a1} + p_1\sigma_{a0})\lambda_{\rm eff}}, \tag{57}$$

$$S_{\rm eff} = \frac{p_0 s_0 + p_1 s_1 + (p_0\sigma_{a1}s_0 + p_1\sigma_{a0}s_1)\lambda_{\rm eff}}{1 + (p_0\sigma_{a1} + p_1\sigma_{a0})\lambda_{\rm eff}}, \tag{58}$$

with $\sigma_{ai} = \sigma_i - \sigma_{si}$. In the absence of scattering ($\sigma_{si} = 0$, so that $\sigma_{ai} = \sigma_i$) and time dependence, we see that Eqs. (50) and (54) are identical, and Eqs. (55) and (58) reproduce Eqs. (45) and (48) if one chooses $\lambda_{\rm eff}$ equal to $q\lambda_c$.

Thus by a proper choice of the single parameter $\lambda_{\rm eff}$ in the expression [see Eq. (7)]

$$\hat{\Sigma} = p_1\sigma_0 + p_0\sigma_1 + \lambda_{\rm eff}^{-1}, \tag{59}$$

the Markov description has been modified to give both the correct $\sigma_{\rm eff}$ and $S_{\rm eff}$ for general non-Markov renewal statistics. Accordingly, we suggest Eq. (54) as a reasonable low order approximate description of transport in a non-Markov mixture, including the effects of scattering. A better (we show this by example in the next section), but still approximate, treatment would be to use Eq. (4) with $\hat{\Sigma}$ given by Eq. (59). One advantage of using this modified Eq. (4) over Eq. (54) is that it reduces to an exact treatment in the special case of Markov statistics. More generally, in the case of non-Markov statistics, the modified Eq. (4) would be expected to be more accurate than Eq. (54). In particular, in the time-independent source-free ($s_i = 0$) case with constant cross sections, Eq. (4) predicts a solution which is the sum of two exponentials, whereas Eq. (54) obviously gives a single-exponential solution according to Eq. (44). A two-exponential form clearly has a much better chance of accurately predicting the true solution than does a single exponential. In the next section we will see by example that the modified Eq. (4) is, in fact, more accurate than Eq. (54). The disadvantage of the modified Eq. (4) over Eq. (54) is that it is obviously more complex, being a two-component vector description rather than a standard scalar model.

For this treatment to be robust, we would like $\lambda_{\rm eff}$ to be non-negative. This is easily shown to be the case. From the definition of $\lambda_{\rm eff}$ given by Eq. (53), it obviously suffices to show that $q \geqslant 0$, where q is defined by Eq. (43). We first bound $\tilde{Q}_i(\sigma_i)$, given by

$$\tilde{Q}_i(\sigma_i) = \int_0^\infty dx\, e^{-\sigma_i x} Q_i(x). \tag{60}$$

We wish to maximize $\tilde{Q}_i(\sigma_i)$ over all admissible functions $Q_i(x)$. From the definition of $Q_i(x)$ given by Eq. (22), this family of functions satisfies

$$Q_i(0) = 1, \quad Q_i(\infty) = 0, \tag{61}$$

$$\frac{dQ_i(x)}{dx} \leqslant 0, \quad \int_0^\infty dx\, Q_i(x) = \lambda_i. \tag{62}$$

Since the integrand in Eq. (60) contains $\exp(-\sigma_i x)$, a decreasing function of x, it is clear that $\tilde{Q}_i(\sigma_i)$ will be maximized if the function $Q_i(x)$ is as concentrated as possible near $x = 0$, subject to the constraints given by Eqs. (61) and (62). In fact, it is easily shown that the function $Q_i(x)$ which maximizes $\tilde{Q}_i(\sigma_i)$ is that corresponding to a periodic lattice as given by Eqs. (29) and (30). Thus we conclude

$$\tilde{Q}_i(\sigma_i) \leqslant (1 - e^{-\sigma_i\lambda_i})/\sigma_i, \tag{63}$$

and use of this in Eq. (43) gives

$$q \geqslant \left[\frac{1}{1 - e^{-\sigma_0\lambda_0}} - \frac{1}{\sigma_0\lambda_0} - \frac{1}{2}\right] + \left[\frac{1}{1 - e^{-\sigma_1\lambda_1}} - \frac{1}{\sigma_1\lambda_1} - \frac{1}{2}\right]. \tag{64}$$

Since

$$1/(1 - e^{-z}) - 1/z - \tfrac{1}{2} \geqslant 0, \tag{65}$$

we then have $q \geqslant 0$, thus completing the proof that $\lambda_{\rm eff} \geqslant 0$.

In the next section, we use the renewal analysis considered here to give a few numerical examples. Specifically, we show by example: (1) the transport results for $\Psi(x)$ are generally sensitive to the details of the statistics in the sense that different chord length distributions with the same mean can give noticeably different results; (2) the transport results for $\Psi(x)$ are generally insensitive to different chord length distributions which have the same mean and variance; and (3) the use of the two-component vector Markov equations with λ_c replaced by $\lambda_{\rm eff}$ is significantly more accurate than the use of the scalar transport equation involving σ_{eff} in predicting transport in non-Markov mixtures.

Before leaving this section, however, we indicate a few characteristics of the use of the gamma distribution as given by Eq. (31) in this renewal theory context. This distribution is particularly interesting since it includes Markov statistics ($m_i = 1$) and a periodic lattice ($m_i = \infty$) as special cases. The two parameters in this distribution can be chosen to match a given mean and variance. As given by Eq. (31), λ_i is the mean, and the variance is given in terms of λ_i and m_i by $V_i = \lambda_i^2/m_i$. Lastly, as we now indicate, in the special case that the parameter m_i is the same for both materials, i.e., $m_0 = m_1$, the Laplace transform given by Eq. (39) is easily inverted to give $\Psi(x)$ as a finite sum of exponentials.

If we take $f_i(x)$ to be the gamma distribution as given by Eq. (31), we find by direct computation

$$\tilde{Q}_i(r+\sigma_i) = (A_i^m - 1)/(r+\sigma_i)A_i^m, \tag{66}$$

where

$$A_i = 1 + \lambda_i(r+\sigma_i)/m. \tag{67}$$

In writing Eqs. (66) and (67) we have set $m_0 = m_1 = m$, a common value. The singularities of $\tilde{\Psi}(r)$ given by Eq. (39) in the complex r plane are easily shown to consist entirely of simple poles at values of r determined by

$$(A_0A_1)^m = 1. \tag{68}$$

(The apparent poles at $r = -\sigma_i$ are, in fact, not singular points.) Taking the mth root of Eq. (68) gives

$$[1+\lambda_0(r+\sigma_0)/m][1+\lambda_1(r+\sigma_1)/m] = e^{2\pi n i/m}, \quad 1 \leqslant n \leqslant m, \tag{69}$$

where $i = (-1)^{1/2}$ on the right-hand side of Eq. (69). Thus we have $2m$ simple poles at values of r given by Eq. (69). This in turn implies that the inversion of $\tilde{\Psi}(r)$ given by Eq. (39) will give a result for $\Psi(x)$ which is the sum of $2m$ exponentials. All that is required to obtain this solution explicitly is to solve the m quadratic equations given by Eq. (69), and find the residues corresponding to these simple poles. In a similar fashion, the explicit solution for the case $m_1 = km_0$ can be reduced to the problem of solving m_0 polynomial equations, each of degree $k+1$.

III. NUMERICAL RESULTS

In this section we give a few representative numerical results that address two questions. In the first place, we investigate the sensitivity of the solution for the ensemble averaged intensity to the details of the statistics. Second, we compare the accuracy of the scalar Eq. (54) involving effective quantities to that of the two-component vector Eq. (4) with $\hat{\Sigma}$ given by Eq. (59). All of the results given in this section are for the source-free ($s_i = 0$) time-independent problem with no scattering ($\sigma_{si} = 0$). Additionally, we assume homogeneous statistics and spatially independent cross sections σ_i. Thus we will be comparing solutions of the simplest transport equation

$$\frac{d\psi(x)}{dx} + \sigma(x)\psi(x) = 0. \tag{70}$$

We take Eq. (70) to hold on the semi-infinite line $0 \leqslant x < \infty$, and assign the nonstochastic boundary condition $\psi(0) = 1$.

The ensemble-averaged solution $\Psi(x)$ for this problem is simply the ensemble average of a decaying exponential according to

$$\Psi(x) = \langle \exp(-\tau) \rangle, \tag{71}$$

where τ in Eq. (71) is given by

$$\tau = \int_0^x dx' \, \sigma(x'). \tag{72}$$

Equation (71) can be evaluated exactly from

$$\Psi(x) = p_0\Psi_0(x) + p_1\Psi_1(x), \tag{73}$$

where $\Psi_0(x)$ and $\Psi_1(x)$ are the solutions of the renewal equations given by Eqs. (35) and (36) and their complements in the source-free ($s_i = 0$) case. The approximate solution corresponding to the scalar Eq. (54) with $S_{\text{eff}} = \sigma_{s,\text{eff}} = 0$ is given by the single exponential according to Eq. (44). The approximate solution corresponding to the vector Eq. (4), with $\hat{\Sigma}$ given by Eq. (59), is the sum of two decaying exponentials.

We give numerical results corresponding to three different sets of physical parameters p_i (the probability) and $\tau_i = \sigma_i\lambda_i$ (the mean optical chord length). The quantities λ_i (the mean geometric chord length) and σ_i (the cross section) follow from

$$\lambda_i = (\tau_0 + \tau_1)p_i, \quad \sigma_i = \tau_i/\lambda_i. \tag{74}$$

In all three cases, we choose the unit of length so that Σ, the ensemble-averaged cross section as given by Eq. (7), is unity. These three sets of parameters we consider are summarized in Table I. Case I corresponds physically to a small amount ($p_1 = 0.1$) of large cross section ($\sigma_1 = 9.09$) material admixed with a large amount ($p_0 = 0.9$) of small cross section ($\sigma_0 = 0.10$) material. The optical depths in this case are small ($\tau_0 = 0.1$) and moderate ($\tau_1 = 1.0$). Case II involves the same p_i and σ_i as case I, but the optical depths for case II are ten times larger than for case I, namely $\tau_0 = 1.0$ and $\tau_1 = 10$. Case III describes an equal amount ($p_0 = p_1 = 0.5$) of the two fluids, one with a very small cross section ($\sigma_0 = 0.02$) and one with a moderate cross section ($\sigma_1 = 1.98$). In this case the optical depths are small ($\tau_0 = 0.1$) and large ($\tau_1 = 10.0$). The chord length distributions we investigated numerically are the two-parameter gamma distribution given by Eq. (31), and the block, tent, and ramp functions given by Eqs. (32)–(34), respectively.

Figure 1 compares exact results for $\Psi(x)$ corresponding to case I parameters for five different chord length distributions, namely Markov (gamma distribution with $m_0 = m_1 = 1$), block, tent, ramp, and a periodic medium (gamma

TABLE I. The parameters used in the numerical results.

Case	p_0	p_1	τ_0	τ_1	λ_0	λ_1	σ_0	σ_1
I	0.9	0.1	0.1	1.0	0.99	0.11	0.10	9.09
II	0.9	0.1	1.0	10.0	9.90	1.10	0.10	9.09
III	0.5	0.5	0.1	10.0	5.05	5.05	0.02	1.98

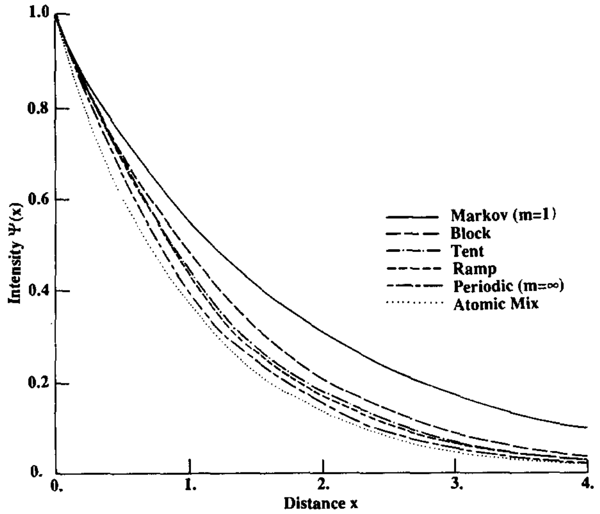


FIG. 1. Comparison of exact results for different statistics for case I parameters.

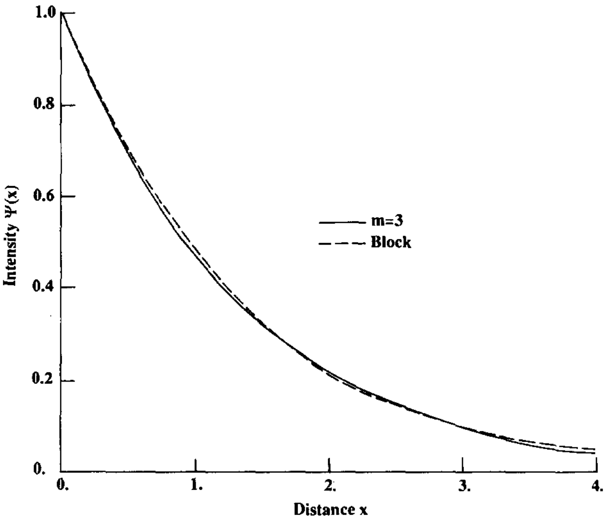


FIG. 3. Comparison of block and gamma ($m = 3$) exact results for case I parameters.

distribution with $m_0 = m_1 = \infty$). Also included in this figure is the atomic mix result, which is simply

$$\Psi_{\rm AM}(x) = e^{-x}, \tag{75}$$

since Σ, the ensemble-averaged cross section, is unity. We see from this figure a fair amount of spread among the curves, from which we conclude that a knowledge of the mean of the chord length distribution is in general not sufficient information concerning the statistics to accurately predict $\Psi(x)$. In particular, we see that the Markov curve (which corresponds to exponentially distributed chord lengths) is further away from the other curves than is the atomic mix curve. Accordingly, in this case if the true statistics correspond to one of the four middle curves, it is more accurate (except for small x) to ignore statistical effects entirely (i.e., use atomic mix) rather than use the Markov statistical description.

Figure 2 gives the same comparison as does Fig. 1, except now the six curves correspond to case II parameters. Again we see the same trend, but in this case the atomic mix result is very far from the other five curves. We note that in both Figs. 1 and 2 the atomic mix curve is the lowest (expected on physical grounds) and the Markov curve is the highest. In all of our numerical calculations, we observed that the atomic mix and Markov results always bound the result for non-Markov statistics from below and above, respectively.

Having concluded from these and other numerical comparisons that the mean alone is not sufficient information concerning the chord length distribution to accurately predict $\Psi(x)$, we asked if the mean and variance together constitute sufficient information. The results given in Figs. 3–5, and other similar results, answer this question in the affirmative; $\Psi(x)$ is relatively insensitive to which chord length distribution is used, given that the mean and variance is preserved. Figure 3 compares the block distribution with the $m_0 = m_1 = 3$ gamma distribution ($m = 3$ matches the block variance) for case I parameters. Likewise, Fig. 4 compares the tent distribution with the $m_0 = m_1 = 6$ gamma distribu-

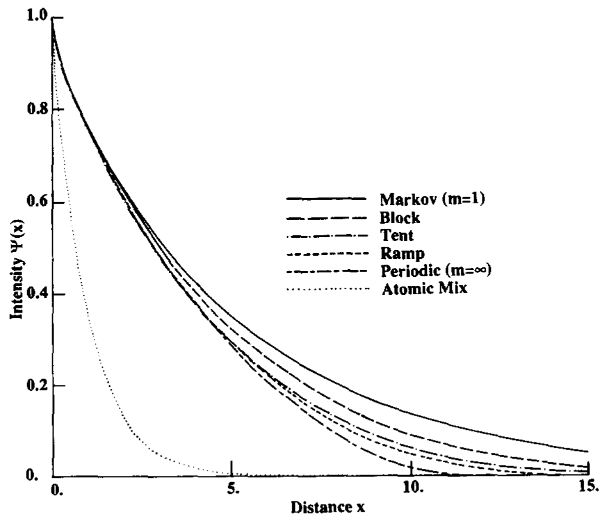


FIG. 2. Comparison of exact results for different statistics for case II parameters.

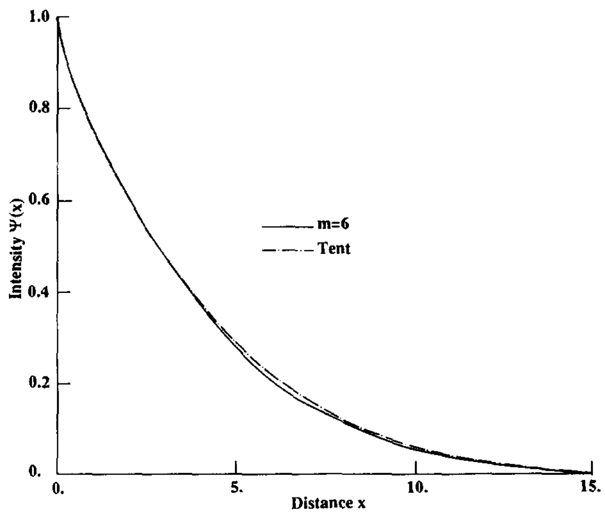


FIG. 4. Comparison of tent and gamma ($m = 6$) exact results for case II parameters.

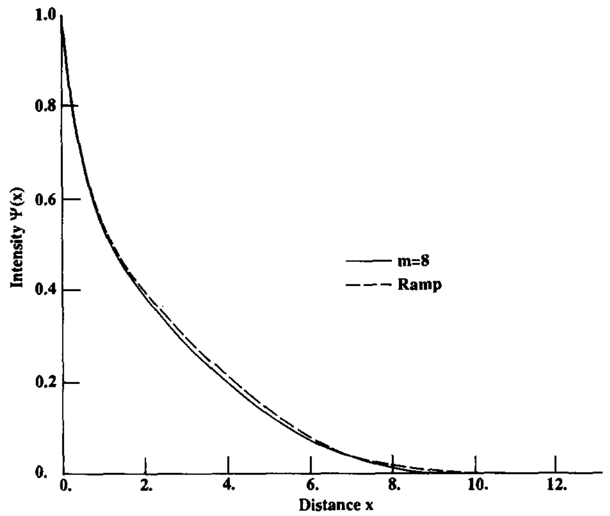


FIG. 5. Comparison of ramp and gamma ($m = 8$) exact results for case III parameters.

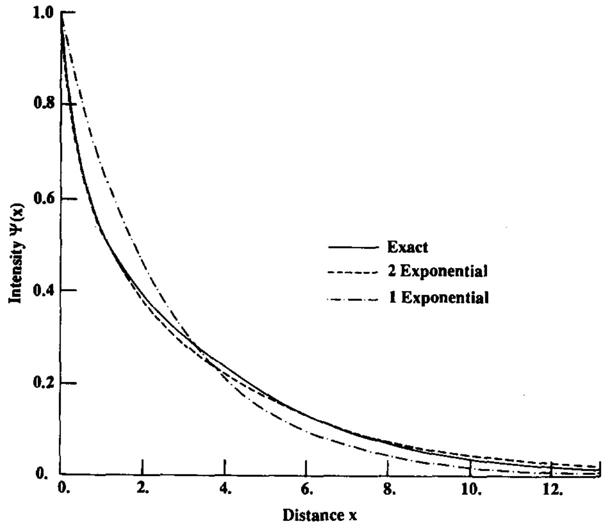


FIG. 7. Comparison of one- and two-exponential approximations for gamma ($m = 2$) statistics and case III parameters.

tion ($m = 6$ matches the tent variance) for case II parameters, and Fig. 5 compares the ramp distribution with the $m_0 = m_1 = 8$ gamma distribution ($m = 8$ matches the ramp variance) for case III parameters. In all three figures we see that the two curves are very close. Thus we conclude, at least for the unimodal distributions we have investigated, that the solution for $\Psi(x)$ is relatively insensitive to the higher (greater than second) moments of the chord length distribution. It follows then that any distribution with two parameters which can be chosen to match an arbitrary mean and variance (such as the gamma distribution) can be used to represent an arbitrary distribution insofar as predicting $\Psi(x)$ with a small error is concerned.

The second major question we addressed with our numerical calculations concerns the accuracy of the scalar [Eq. (54)] (involving effective quantities) and the vector [Eq. (4)] [in conjunction with Eq. (59)] as approximations to the true solution for $\Psi(x)$. Figure 6 considers this comparison for case I parameters, and a gamma distribution with $m_0 = m_1 = 2$. We see that both the one-exponential [corresponding to Eq. (54)] and the two-exponential corresponding to Eq. (4)] approximations are very close to the exact result. This is not unexpected since case I parameters correspond exactly to the asymptotic scaling which leads to a renormalized transport equation involving effective quantities. Figure 7 makes the same comparison as in Fig. 6, again for a gamma distribution with $m_0 = m_1 = 2$, but for case III parameters. Here we see that the one-exponential approximation is noticeably different from the exact solution, whereas the two-exponential result is a quite good fit to the exact curve. Based upon these and other similar results, we conclude that Eq. (4) in conjunction with Eq. (59) is a reasonably accurate model for non-Markov statistics, whereas Eq. (54), while robust, is in general noticeably less accurate.

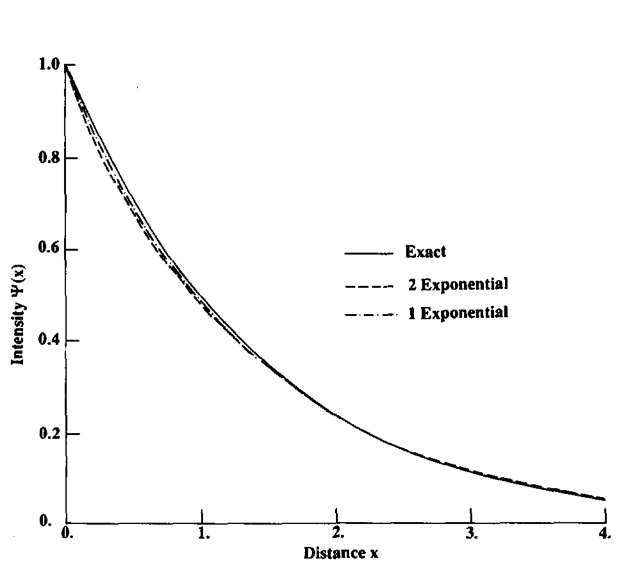


FIG. 6. Comparison of one- and two-exponential approximations for gamma ($m = 2$) statistics and case I parameters.

It would be useful to extend these numerical comparisons to more complex (multimodal) chord length distributions, and involving both external sources ($s_i \neq 0$) and scattering ($\sigma_{si} \neq 0$). These comparisons, in particular including scattering, would be much more difficult to make, but would be very useful to test the robustness of our two tentative conclusions. To summarize, these conclusions are the following: (1) the mean and the variance of the chord length distribution together constitute sufficient information of the statistics to obtain a good approximation for $\Psi(x)$, whereas the mean alone is not sufficient information; and (2) the modified Markov description given by Eqs. (4) and (59) is a quite accurate approximation to the true non-Markov situation, whereas Eq. (54) is in general less accurate. In this regard, it should be emphasized that at the present time no simple exact formulation of the non-Markov situation with scattering [analogous to Eq. (4) for Markov statistics] is known. Hence to obtain an exact solution for non-Markov statistics including scattering, one would have to populate physical space statistically with the two fluids, perform a transport calculation for each realization of this statistical population, and then average a large number of these calculations to obtain an estimate of the ensemble averaged solution. This procedure would be prohibitively time consuming (expensive) even on the most powerful computer.

IV. CONCLUDING REMARKS

As pointed out in the Introduction, with Eq. (2) as the underlying transport equation, Eq. (4) constitutes a complete transport description for the ensemble-averaged intensity in the case of inhomogeneous Markov statistics, including the scattering process. In the absence of scattering, Eqs. (19) and (20), together with the complementary equations found by interchanging the indices, give the ensemble-averaged transport description for arbitrary inhomogeneous renewal statistics. What is missing at this point is an exact transport description for arbitrary renewal statistics with the scattering interaction included in the underlying transport equation. How to treat this case is a major unanswered question in this general area of linear transport in randomly mixed immiscible fluids. One approach is to search for and justify a master equation for general renewal statistics in a multidimensional setting including scattering. At this time, such a master equation is unknown.

Other generalizations of the results obtained to date suggest themselves. One could extend all of the results in this and the earlier papers[1-4] to the case of more than two fluids. Such an extension should be entirely straightforward. As mentioned at the end of the last section, more numerical studies (say on multimodal chord length distributions) could be undertaken to test the limits of the conclusions we have tentatively drawn in this paper. The question of the physical realizability of the renewal statistics we have used could be studied. For example, can one envision an ensemble of partitioning of all space into two (or more) materials, such that the same prescribed homogeneous chord length distribution is extant along an arbitrary ray? It is known[12] that this is indeed possible in the special case of an exponential distribution (Markov statistics). One can also use the master equation approach to derive equations for the higher moments of the random intensity field, such as the variance. All of our work to date has focused on obtaining a transport description for the ensemble average.

We close by suggesting one concrete line of inquiry that we hope to pursue in the near future. Let us assume that all chord length distributions with the same mean and variance give essentially the same result for the ensemble averaged intensity, within some acceptable accuracy limits. For the unimodal distributions considered in the last section, this seems to be the case. Then, insofar as a treatment of non-Markov statistics including scattering is concerned, it suffices to develop a theory for a single two-parameter chord length distribution whose parameters can be chosen to match the mean and variance of any distribution of interest. That is, it is not necessary to develop a theory for general renewal statistics corresponding to an arbitrary chord length distribution. One only has to treat the case corresponding to one particular chord length distribution to obtain acceptable accuracy for any distribution. The hope is that such a particular distribution can be found which lends itself to an exact treatment, including scattering, for the ensemble averaged intensity. If this cannot be done, one still has Eqs. (4) and (59) as an alternative treatment of non-Markov statistics. The numerical results given in the last section, and others, suggest this as a reasonably accurate approximation.

In this regard, we note that the current ability to predict the statistical characteristics of a random mixture is in a relatively rudimentary state. Hence, as a practical matter, an approximate but reasonably accurate transport description involving simple integral characteristics of the statistics [such as $\widetilde{Q}_i(\sigma_i)$ given by Eq. (60)] may be just as useful or perhaps even more so, than an exact description requiring all of the details of the statistics. It may well be the case that Eqs. (4) and (59), from a practical point of view, constitute a theory which is sufficiently accurate at this point in time. However, from an understanding and completeness point of view, more needs to be done. While it seems that a quite good start has been made, much remains to be done before this area of inquiry can be considered to be unfruitful for further research.

ACKNOWLEDGMENTS

The authors acknowledge helpful discussions with Larry George of the Lawrence Livermore National Laboratory. The authors are also indebted to a reviewer who made helpful comments.

The work of the first (CDL) and last (JW) authors was supported by the U. S. Department of Energy, and that of the second (GCP) author was partially supported by the National Science Foundation.

[1] C. D. Levermore, G. C. Pomraning, D. L. Sanzo, and J. Wong, J. Math. Phys. **27**, 2526 (1986).
[2] D. Vanderhaegen, J. Quant. Spectros. Radiat. Transfer **36**, 557 (1986).
[3] C. D. Levermore, "Transport in randomly mixed media with inhomogeneous anisotropic statistics," in preparation.
[4] D. Vanderhaegen, "Impact of a mixing structure on radiative transfer in random media," to be published in J. Quant. Spectros. Radiat. Transfer.
[5] J. B. Keller, "Wave propagation in random media," in *Proceedings of the Symposium on Applied Mathematics* (Am. Math. Soc., Providence, RI, 1962), Vol. 13, p. 227.
[6] J. B. Keller, "Stochastic equations and wave propagation in random media," in *Proceedings of the Symposium on Applied Mathematics* (Am. Math. Soc., Providence, RI, 1964), Vol. 16, p. 145.
[7] J. B. Keller, "Effective behavior of heterogeneous media," in *Statistical Mechanics and Statistical Methods in Theory and Application*, edited by U. Landman (Plenum, New York, 1977), p. 632.
[8] U. Frisch, "Wave propagation in random media," in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968), p. 75.
[9] J. A. Morrison, J. Math. Phys. **13**, 299 (1972).
[10] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, (North-Holland, Amsterdam, 1981).
[11] D. R. Cox, *Renewal Theory* (Chapman and Hall, London, 1982).
[12] P. Switzer, Am. Math. Statist. **36**, 1859 (1965).

Bäcklund transformations for the anti-self-dual Yang–Mills equations

L. Mason[a)]
New College, Oxford University, England
S. Chakravarty and E. T. Newman
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 1 December 1987; accepted for publication 16 December 1987)

Beginning from any given (local) solution of the GL(n,C) anti-self-dual Yang–Mills (ASDYM) equations on Minkowski space, a simple technique for the generation of large classes of solutions (perhaps in some sense all) is given. The origin of this technique is described in terms of two versions of the Ward construction. The resulting description of Bäcklund transformations is sufficiently simple that it is then possible to identify the group generated by the collection of all such Bäcklund transformations and the space on which it acts in terms of concrete functions.

I. INTRODUCTION

In the past several years a great deal of progress has been made in understanding the structure of the anti-self-dual Yang–Mills (ASDYM) equations and the various solution generating methods. In one approach,[1,2] the ASDYM equations are thought of as the integrability condition on a pair of first-order linear differential equations, the so-called "Lax pair." The Lax pair leads, by several ingenious techniques, to the subject of Bäcklund transformations,[3,4] where the basic idea is to generate new solutions of the ASDYM equations from a seed solution. Another approach by Ward, and with a slight variation by Sparling (unpublished), establishes the correspondence between solutions of the ASDYM equations and holomorphic vector bundles on twistor space. Solutions are then generated by a Riemann–Hilbert splitting of the patching functions defining the bundle.

The purpose of this work is twofold.

(i) We will describe an extremely simple technique for the generation of large classes of solutions (in some sense all) of the GL(n,C) ASDYM equations on Minkowski space, beginning from any given (local) seed solution. From a given seed solution, i.e., a Yang–Mills (YM) potential (or connection) chosen in a suitable gauge, we show that one can take specific combinations of the components of the given potential to produce a completely new potential for the ASDYM field. The important fact is that this technique is completely algebraic and only involves matrix multiplication. (An interesting feature arising from these transformations is that from the inhomogeneous part of the transformation it is possible to obtain a nontrivial YM field both from the pure gauge and even from the zero seed solution.)

(ii) We wish to describe the geometric origin, via the Ward and Sparling approach, of these BT's and show its relationship to the work of others. In particular we wish to show that the BT's become considerably simpler and less mysterious when represented on the twistor data for the ASDYM fields. Indeed, the simplification is sufficient to allow one to identify the infinite-dimensional transformation groups generated by the BT's and the appropriate extension of the solution space on which they act.

In Sec. II we present our notation and the new solution generating method. In Sec. III we describe the Sparling and Ward approaches to the ASDYM and how the BT's arise. In particular we show the geometric origin (from the gauge theory point of view) of the Lax pair and the Yang–Pohlmeyer equation. In Sec. IV we discuss the action of BT on the twistor data and how it relates to the problem of reconstructing the space-time field from the twistor data. In Sec. V we discuss the transformation group to which the BT's give rise. We also briefly relate these ideas to the analogous ones for the stationary axisymmetric vacuum equations.

(In this paper we are concerned only with the local problem and do not consider boundary conditions for the YM equations. This is because there are many different boundary conditions for, and reductions of, the ASDYM equations of interest and it would take us too far afield to discuss them in any detail. Furthermore, they are not necessary for the understanding of the BT's and the infinite-dimensional group to which they give rise.)

II. THE BÄCKLUND TRANSFORMATIONS

A. Notation

We begin with our notation and the discussion of several useful geometric ideas.

The ASDYM equations obtained below are three matrix-valued differential equations on the matrix-valued vector potential γ_a. Define the YM field by

$$F_{ab} = 2\nabla_{[a}\gamma_{b]} - [\gamma_a,\gamma_b].$$

Then the ASD equations are

$$F^*{}_{ab} = -iF_{ab}\,, \tag{2.1}$$

where the asterisk is the Hodge duality operator. A useful alternative set of equations equivalent to (2.1) are[5]

$$F_{ab}L^{ab} = F_{ab}\overline{M}^{ab} = F_{ab}N^{ab} = 0, \tag{2.2}$$

where L, $\overline{M}$, and N are any three independent self-dual antisymmetric tensors. Equation (2.2) follows from the orthogonality of self-dual and anti-self-dual forms. A succinct version of (2.2) is

[a)] Present address: Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

 0022-2488/88/041005-09$02.50

$$F_{ab}\overline{m}^{ab}=0\,, \tag{2.3}$$

with $\overline{m}^{ab}$ a self-dual skew tensor written as

$$\overline{m}^{ab}=L^{ab}+\overline{M}^{ab}\zeta+N^{ab}\zeta^2, \tag{2.4}$$

where ζ is an arbitrary point on the (complex) Riemann plane $C+(\infty)$.

A particularly useful form for the L, $\overline{M}$, and N is obtained as follows. Define on Minkowski space the normalized null tetrad and the coordinates u, v, w, $\overline{w}$ by

$$\begin{aligned} D^0&\equiv l^{0a}\nabla_a=\partial_t+\partial_z=\partial_v,\\ \Delta^0&\equiv n^{0a}\nabla_a=\partial_t-\partial_z=\partial_u\,, \end{aligned} \tag{2.5a}$$

$$\begin{aligned} \delta^0&\equiv m^{0a}\nabla_a=-\partial_x+i\,\partial_y=-\partial_w,\\ \overline{\delta}^0&\equiv \overline{m}^{0a}\nabla_a=-\partial_x-i\,\partial_y=-\partial_{\overline{w}}\,, \end{aligned} \tag{2.5b}$$

with $l^0\cdot n^0=-m^0\cdot\overline{m}^0=1$ and all other products vanishing. Next, from the null tetrad, we define the following vectors:

$$L^a(\zeta)=l^{0a}+\zeta m^{0a}, \tag{2.6}$$

$$\overline{M}^a(\zeta)=\overline{m}^{0a}+\zeta n^{0a}. \tag{2.7}$$

Then using (2.6) and (2.7) we make the following choice for (2.4):

$$\overline{m}^{ab}\equiv L^{[a}\overline{M}^{b]}=l^{0[a}\overline{m}^{0b]}+\zeta\{l^{0[a}n^{0b]}+m^{0[a}\overline{m}^{0b]}\}+\zeta^2m^{0[a}n^{0b]}. \tag{2.8}$$

The skew tensor $L^{[a}\overline{M}^{b]}$ at any point defines a self-dual two-surface through that point. As ζ ranges over the complex Riemann sphere, $L^{[a}\overline{M}^{b]}$ ranges over all self-dual totally null two-planes through that point. The vectors $L^a(\zeta)$ and $\overline{M}^a(\zeta)$ are two independent vectors in these planes. The set of all such two-surfaces in Minkowski space is (projective) twistor space.

B. The Bäcklund transformations

We now give the simple prescription for obtaining the BT from any given seed ASDYM connection γ_a.

Given a seed solution γ_a of the ASDYM equations, we first construct the expressions $\gamma_a(x^b)L^a(\zeta)$ and $\gamma_a(x^b)\overline{M}^a(\zeta)$ and then seek some matrix $\mathscr{B}(x^a,\zeta)$ such that the right-hand side (rhs) of

$$\mathscr{B}\gamma_a(x^b)L^a(\zeta)\mathscr{B}^{-1}+L^a(\zeta)\nabla_a\mathscr{B}\cdot\mathscr{B}^{-1}=\gamma'_a(x^b)L^a(\zeta), \tag{2.9a}$$

$$\mathscr{B}\gamma_a(x^b)\overline{M}^a(\zeta)\mathscr{B}^{-1}+\overline{M}^a(\zeta)\nabla_a\mathscr{B}\cdot\mathscr{B}^{-1}=\gamma'_a(x^b)\overline{M}^a(\zeta) \tag{2.9b}$$

is to have the same ζ behavior as the $\gamma\cdot L$ and $\gamma\cdot\overline{M}$, i.e., (2.9a) and (2.9b) are to be linear in ζ. Assuming that such a $\mathscr{B}$ exists (the proof by construction will be given later), we claim that the γ'_a defined by (2.9) is a new connection satisfying the ASDYM equations and further is (in general) gauge *inequivalent* to the original γ_a.

The proof of this consists of first noting that the field corresponding to the connection γ'_a is given by

$$F'_{ab}=2\nabla_{[a}\gamma'_{b]}-[\gamma'_a,\gamma'_b] \tag{2.10}$$

and that, using Eqs. (2.9) one obtains

$$F'_{ab}L^a\overline{M}^b=\mathscr{B}F_{ab}L^a\overline{M}^b\mathscr{B}^{-1}. \tag{2.11}$$

By Eqs. (2.3) and (2.8), the rhs vanishes and hence F'_{ab} is an ASDYM field and γ' is an ASDYM connection. That γ and γ' are, in general, gauge inequivalent follows from the fact that Eqs. (2.9) mix the connection components, while a standard gauge transformation does not. An additional feature of (2.9) is that if two seed solutions γ and γ^* are gauge *equivalent*, then the two new solutions γ' and γ'^* obtained from (2.9) are, in general, gauge *inequivalent*.

Before discussing the construction of $\mathscr{B}$ for the $\mathrm{GL}(n,\mathbb{C})$ case, we illustrate the idea for the specific case of $\mathrm{GL}(2,\mathbb{C})$.

To begin, any seed solution γ of the $\mathrm{GL}(2,\mathbb{C})$ equations must first be written in a special gauge (always obtainable) which is defined by the vanishing of certain matrix elements of γ_a. We require that the matrices

$$\gamma_aL^a(\zeta_0)\ \text{and}\ \gamma_a\overline{M}^a(\zeta_0) \tag{2.12}$$

are upper triangular and the matrices

$$\tilde{\zeta}_0\gamma_aL^a(-\tilde{\zeta}_0^{-1})\ \text{and}\ \tilde{\zeta}_0\gamma_a\overline{M}^a(-\tilde{\zeta}_0^{-1}) \tag{2.13}$$

are lower triangular, where $\zeta_0\equiv\zeta_0(x^a)$ and $\tilde{\zeta}_0\equiv\tilde{\zeta}_0(x^a)$ are functions of x^a obtained, respectively, as solutions of the algebraic equations

$$f(x^aL_a(\zeta_0),x^a\overline{M}_a(\zeta_0),\zeta_0)=0\,, \tag{2.14a}$$

$$\tilde{f}(x^aL_a(-\tilde{\zeta}_0^{-1}),x^a\overline{M}_a(-\tilde{\zeta}_0^{-1}),-\tilde{\zeta}_0^{-1})=0\,, \tag{2.14b}$$

with f and $\tilde{f}$ arbitrary analytic functions of their three arguments.

[A particularly simple choice would be $\zeta_0=\tilde{\zeta}_0=0$, which leads to

$$L^a(\zeta_0)=l^{0a},\quad \overline{M}^a(\zeta_0)=\overline{m}^{0a},$$
$$\tilde{\zeta}_0L^a(-\tilde{\zeta}_0^{-1})=-m^{0a},$$
$$\tilde{\zeta}_0\overline{M}^a(-\tilde{\zeta}_0^{-1})=-n^{0a}.]$$

When γ is in the special gauge defined by (2.12) and (2.13), then it is checked easily that $\mathscr{B}$ of the form, either

$$\mathscr{B}=\begin{pmatrix}\Lambda & 0\\ 0 & 1\end{pmatrix}\ \text{or}\ \mathscr{B}=\begin{pmatrix}0 & 1\\ \Lambda & 0\end{pmatrix}, \tag{2.15}$$

with

$$\Lambda=(\zeta-\zeta_0)(1+\zeta\tilde{\zeta}_0)^{-1}, \tag{2.16}$$

satisfies Eqs. (2.9) automatically.

With the special choice $\zeta_0=\tilde{\zeta}_0=0$ we have

$$\mathscr{B}=\begin{pmatrix}0 & 1\\ \zeta & 0\end{pmatrix}. \tag{2.17}$$

Equations (2.9), in this case, become a set of algebraic equations between different components of γ_a and γ'_a (both being in the "special gauge"); one can solve them for γ'_a as follows:

$$\begin{aligned} l^{0a}\gamma'_a&=-l^{0a}\gamma_a+E_{11}l^{0a}\gamma_aE_{22}-E_{12}m^{0a}\gamma_aE_{12}\,,\\ m^{0a}\gamma'_a&=-m^{0a}\gamma_a+E_{22}m^{0a}\gamma_aE_{11}-E_{21}l^{0a}\gamma_aE_{21}\,,\\ \overline{m}^{0a}\gamma'_a&=-\overline{m}^{0a}\gamma_a+E_{11}\overline{m}^{0a}\gamma_aE_{22}+E_{12}n^{0a}\gamma_aE_{12}\,,\\ n^{0a}\gamma'_a&=-n^{0a}\gamma_a+E_{22}n^{0a}\gamma_aE_{11}+E_{21}\overline{m}^{0a}\gamma_aE_{21}\,, \end{aligned} \tag{2.18}$$

where the E_{ij}'s are 2×2 matrices with 1 as their ijth element and all other elements being zero.

Remarks: (i) Equations (2.18) can be made somewhat simpler using spinors; in the spin frame (o^A,ι^A) given by $l^{0a}=o^A\bar{o}^{A'}$, $n^{0a}=\iota^A\iota^{A'}$, $m^{0a}=o^A\iota^{A'}$, etc., the gauge conditions above imply that the connection γ_a can be written in the form

$$\gamma_a=\begin{pmatrix} a_{AA'} & b_A o_{A'} \\ c_A\iota_{A'} & d_{AA'} \end{pmatrix}.$$

The new connection γ'_a is then given by interchanging b_A with c_A and $a_{AA'}$ with $d_{AA'}$:

$$\gamma'_a=\begin{pmatrix} d_{AA'} & c_A o_{A'} \\ b_A\iota_{A'} & a_{AA'} \end{pmatrix}.$$

(ii) It should be pointed out that even within the special gauge there is still considerable gauge freedom. In fact, in the $\zeta_0=\tilde{\zeta}_0=0$ case, the gauge transformation $\gamma_a\to\gamma^*_a=F\gamma_aF^{-1}+\nabla_aF\cdot F^{-1}$, with

$$F(x)=\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}, \tag{2.19}$$

satisfying

$$f_{12}/f_{11}=AJ(x^al^0{}_a,x^a\bar{m}^0{}_a)/\{1-BJ(x^al^0{}_a,x^a\bar{m}^0{}_a)\},$$

$$f_{21}/f_{22}=CH(x^am^0{}_a,x^an^0{}_a)/\{1-DH(x^am^0{}_a,x^an^0{}_a)\}$$

preserves the special gauge conditions. Here A, B, C, and D are particular combinations of integrals of matrix elements of γ_a, while J and H are arbitrary functions of their arguments. The BT's obtained from two gauge related γ's are gauge *inequivalent* when the J and H are different from zero, i.e., when F is not diagonal.

(iii) An important point to emphasize is that we can compose two different $\mathscr{B}$'s, say $\mathscr{B}_1$ and $\mathscr{B}_2$, to obtain another allowable $\mathscr{B}=\mathscr{B}_1\mathscr{B}_2$. Since the inverse of a $\mathscr{B}$ is also allowable, the set of $\mathscr{B}$'s form, roughly speaking, a transformation group on the solution space of the ASDYM equations. Successive applications of this procedure result in the analog of the Kinnersley–Chitre type of transformation for ASDYM equations[6,7]. This technique (in a specialized form) was also used by Corrigan *et al.*[8] and Prasad *et al.*[9] to generate the infinite hierarchy of Atiyah and Ward.[10] We conjecture that they act transitively on this space.

(iv) As a simple illustration of the above solution generation technique in the GL(2,C) case, let us begin with the trivial seed solution, i.e., $\gamma_a=0$. Then Eqs. (2.9) reduce to

$$L^a\gamma'_a=L^a\nabla_a\mathscr{B}\cdot\mathscr{B}^{-1},\quad \bar{M}^a\gamma'_a=\bar{M}_a\nabla_a\mathscr{B}\cdot\mathscr{B}^{-1},$$

which with Eq. (2.15) become

$$L^a\gamma'_a=\mathrm{diag}(0,a),\quad \bar{M}^a\gamma'_a=\mathrm{diag}(0,b),$$

where

$$a=-(\delta^0\zeta_0+\zeta\delta^0\ln\tilde{\zeta}_0),\quad b=-(\Delta^0\zeta_0+\zeta\Delta^0\ln\tilde{\zeta}_0)$$

is linear in ζ, as expected. The components of $\gamma'$are then given by

$$l^{0a}\gamma'_a=\mathrm{diag}(0,-\delta^0\zeta_0),$$
$$\bar{m}^{0a}\gamma'_a=\mathrm{diag}(0,-\Delta^0\zeta_0),$$
$$n^{0a}\gamma'_a=\mathrm{diag}(0,-\Delta^0\ln\tilde{\zeta}_0),$$
$$m^{0a}\gamma'_a=\mathrm{diag}(0,-\delta^0\ln\tilde{\zeta}_0).$$

We are thus able to produce nontrivial Maxwell fields by applying the BT on the trivial GL(2,C) ASDYM solution.

C. Generalizations

The ideas used in the GL(2,C) case are generalized easily to GL(n,C). We consider again the two functions $\zeta_0=\zeta_0(x)$ and $\tilde{\zeta}_0=\tilde{\zeta}_0(x)$ defined as the solutions of Eqs. (2.14) and Λ defined by (2.16). We next consider any $n\times n$ projection operator E, which is simply a diagonal matrix with only ones and zeros on the diagonal ($E^2=E$). Then $\mathscr{B}(x^a,\zeta)$ is defined by

$$\mathscr{B}(x^a,\zeta)=(I-E)+\Lambda E. \tag{2.20}$$

Before the BT [Eqs. (2.9], can be implemented, one must again impose a special gauge condition on the seed solution. These conditions are as follows.

In the new gauge we construct the four components of γ,

$$\gamma\cdot L(\zeta_0),\quad \gamma\cdot\bar{M}(\zeta_0),$$
$$\tilde{\zeta}_0\gamma\cdot L(-\tilde{\zeta}_0{}^{-1}),\quad \tilde{\zeta}_0\gamma\cdot\bar{M}(-\tilde{\zeta}_0{}^{-1}),$$

and require the vanishing of certain specific matrix elements of the above components, namely,

$$\begin{aligned} &E\gamma\cdot L(\zeta_0)(I-E)=0,\quad E\gamma\cdot\bar{M}(\zeta_0)(I-E)=0,\\ &(I-E)\gamma\cdot L(-\tilde{\zeta}_0{}^{-1})E=0,\\ &(I-E)\gamma\cdot\bar{M}(-\tilde{\zeta}_0{}^{-1})E=0. \end{aligned} \tag{2.21}$$

These conditions can always be satisfied. As in the GL(2,C) case there is still considerable freedom within these conditions, producing an equivalence class of gauge related connections all satisfying (2.21). Two seed solutions in this class, in general, yield after the BT, two solutions which are gauge inequivalent. Also, products of two transformations and inverses of transformations are transformations and hence again we have a transformation group.

Remark: An alternative, but equivalent point of view to our procedure is the following. Consider an ASDYM connection γ on which we perform an ordinary gauge transformation that depends on the parameter ζ, i.e,

$$\gamma^*{}_a(x,\zeta)=\mathscr{B}\gamma_a\mathscr{B}^{-1}+\nabla_a\mathscr{B}\cdot\mathscr{B}^{-1},\quad \mathscr{B}=\mathscr{B}(x,\zeta).$$

We now seek $\mathscr{B}$ such that $\gamma^*{}_a$ has the form

$$\gamma^*{}_a(x,\zeta)=\gamma'_a(x)+\alpha L_a(\zeta)+\beta\bar{M}_a(\zeta),$$

with α and β any two matrices. If such a $\mathscr{B}$ can be found then $\gamma'_a(x)$ automatically is an ASDYM connection of the type just discussed. The proof of this consists in simply contracting $\gamma^*{}_a$ with L^a and $\bar{M}^a$ and comparing with (2.9). There remains an interesting and perhaps an important question—what is the real geometric meaning of this procedure? It appears to be related to ordinary gauge transformations—not on a vector bundle over the Minkowski space, but on a vector bundle over the spin bundle of the Minkowski space. This question is being studied. We will see later that the BT corresponds to a singlular gauge transformation on a bundle on twistor space.

III. THE GEOMETRIC BACKGROUND

In this section we will describe both the Sparling and Ward approach to the ASDYM equations and the insight they give to the BT's of Sec. II.

A. The Sparling equation

We begin by considering Minkowski space M, with x^a an arbitrary interior point in M and C_x its future null cone. Each null generator of C_x is labeled by a pair $(\zeta,\bar{\zeta})$ (stereographic co-ordinates on S^2 the sphere of null directions) and denoted by $l_x(\zeta,\bar{\zeta})$. (Almost always we will consider M to be the real Minkowski space although on occasion we will allow a "small" complex thickening.) In addition, we have on M a $\mathrm{GL}(n,\mathbb{C})$ ASDYM field with connection γ. In the associated $\mathrm{GL}(n,\mathbb{C})$ bundle we define our basic variable, a linear map $G(x^a,\zeta,\bar{\zeta})$, which (via the connection γ) parallelly propagates, along $l_x(\zeta,\bar{\zeta})$, an arbitrary vector in the fiber over x^a to the fiber at the end point of $l_x(\zeta,\bar{\zeta})$, i.e., to future null infinity ($\mathscr{I}^+$).

[Recall that future null infinity $\mathscr{I}^+$ is a null cone, topologically $S^2\times R$, with coordinates u along the R factor and $(\zeta,\bar{\zeta})$ holmorphic (affine) stereographic coordinates on the S^2 factor. It can be thought of as the space of end points of all future pointing null geodesics, with the null geodesic in the direction $l^a(\zeta,\bar{\zeta})$ through the point x^a and ending at $(u,\zeta,\bar{\zeta}) = (x_a l^a(\zeta,\bar{\zeta}),\zeta,\bar{\zeta})$.]

Now $G(x,\zeta,\bar{\zeta})$, which is a function of the starting point x^a and the generator $(\zeta,\bar{\zeta})$ and plays a fundamental role in all that follows, is defined formally as

$$G = O\exp\int_{l_x(\zeta,\bar{\zeta})}\gamma_a\, dy^a = O\exp\int_{l_x(\zeta,\bar{\zeta})}\gamma_a l^a\, ds, \quad (3.1)$$

where $l^a(\zeta,\bar{\zeta})\nabla_a = \partial_s$ is the normalized null tangent vector along $l_x(\zeta,\bar{\zeta})$ and O indicates path ordered integral. We will show that G satisfies the following first-order linear differential equation (the Sparling equation),[11] which is equivalent to the original ASDYM equations:

$$(1+\zeta\bar{\zeta})\partial_{\bar{\zeta}} G = -G\bar{A}(u,\zeta,\bar{\zeta})|_{u=x^a l_a(\zeta,\bar{\zeta})}, \quad (3.2)$$

where $\bar{A}(u,\zeta,\bar{\zeta})|_{u=x^a l_a(\zeta,\bar{\zeta})}$ is the restriction of the free YM data $\bar{A}(u,\zeta,\bar{\zeta})$ given on $\mathscr{I}^+$ to the intersection of C_x with $\mathscr{I}^+$. This intersection is described by $u\equiv x^a l_a(\zeta,\bar{\zeta})$.

We now outline a proof of the Sparling equation.

Proof: We first introduce the closed path parallel propagator (also called the holonomy operator) associated with an infinitesimal closed loop (in fact, a triangle). This loop, which lies *completely in a self-dual surface*, is defined on the null cone C_x as an infinitesimal triangle $\bar{\Delta}_x$ formed by two neighboring geodesics $l_x(\zeta,\bar{\zeta})$ and $l_x(\zeta,\bar{\zeta}+d\bar{\zeta})$ and is closed at $\mathscr{I}^+$ by a connecting vector $\bar{M}^a\, d\zeta$. Since $\bar{\Delta}_x$ lies in a self-dual blade and our field and connection is anti-self-dual, the associated holonomy operator becomes the identity operator, i.e., the connection is integrable on self-dual surfaces. Using the definition, Eq. (3.1), of the parallel propagator G, its inverse, and the definition of the holonomy operator, we have

$$I = G^{-1}(x^a,\zeta,\bar{\zeta})G(x^a,\zeta,\bar{\zeta}+d\bar{\zeta})\{I+P^{-1}\bar{A}\, d\bar{\zeta}\}, \quad (3.3)$$

where I is the identity operator and $\bar{A}$ is the component of the connection along the connecting vector $\bar{m}^a$ on $\mathscr{I}^+$, i.e.,

$$\bar{A} = P\gamma_a\bar{m}^a|_{\mathscr{I}^+}, \quad P\equiv 1+\zeta\bar{\zeta}, \quad (3.4)$$

where P is used for the normalization of $\bar{m}^a$.

Expanding $G(x^a,\zeta,\bar{\zeta}+d\bar{\zeta})$ in Eq. (3.3) and retaining terms linear in $d\bar{\zeta}$ we obtain the Sparling equation, namely,

$$P\partial_{\bar{\zeta}}G = -G\bar{A}.$$

We have just seen from the definition of G and from the given anti-self-dual connection γ that G satisfies (3.2). We now sketch the proof of the converse, namely, that a solution of (3.2), $G(x,\zeta,\bar{\zeta})$, which is a regular function of the $(\zeta,\bar{\zeta})$ sphere, defines an anti-self-dual connection γ. This result, which we refer to as the reconstruction theorem, is composed of two parts: (i) We must show that G does define a connection, and (ii) the connection is anti-self-dual.

(i) The appropriate relationship between G and γ is suggested by differentiating Eq. (3.1) in the l^a direction, yielding

$$l^a(\zeta,\bar{\zeta})\nabla_a GG^{-1} = \gamma_a l^a(\zeta,\bar{\zeta}). \quad (3.5)$$

What we must do is show that if G does satisfy the Sparling equation then indeed (3.5) defines a $\gamma_a(x)$ independent of $(\zeta,\bar{\zeta})$. This is done most easily by taking an explicit representation of $l^a(\zeta,\bar{\zeta})$, namely,

$$\begin{aligned} l^a(\zeta,\bar{\zeta}) &= P^{-1}\{(l^{0a}+\zeta m^{0a})+\bar{\zeta}(\bar{m}^{0a}+\zeta n^{0a})\} \\ &= 2\sqrt{2}(1+\zeta\bar{\zeta},\zeta+\bar{\zeta},i(\zeta-\bar{\zeta}),\zeta\bar{\zeta}-1). \quad (3.6)\end{aligned}$$

We also make use of

$$\begin{aligned}\bar{m}^a(\zeta,\bar{\zeta}) &\equiv P\left(\frac{\partial}{\partial\bar{\zeta}}\right) l^a \\ &= \{(\bar{m}^{0a}+\zeta n^{0a})-\zeta(l^{0a}+\zeta m^{0a})\} \quad (3.7)\end{aligned}$$

and the identity

$$\left(\frac{\partial}{\partial\bar{\zeta}}\right)\{P\bar{m}^a\} = 0.$$

Applying the operator $P(\partial/\partial\bar{\zeta})$ on (3.5), after some manipulation with the Sparling equation (using $P\partial_{\bar{\zeta}}\nabla_a G = -\nabla_a G\bar{A} - G\{(\partial/\partial u)\bar{A}\}l_a$), yields

$$\bar{m}^a(\zeta,\bar{\zeta})\nabla_a GG^{-1} = \gamma_a\bar{m}^a(\zeta,\bar{\zeta}) + \left\{P\left(\frac{\partial}{\partial\bar{\zeta}}\right)\gamma_a\right\} l^a(\zeta,\bar{\zeta}). \quad (3.8a)$$

Applying $(\partial/\partial\bar{\zeta})P$ to (3.8a) and using the identity and again the Sparling equation shows that $(1+\zeta\bar{\zeta})(\partial/\partial\bar{\zeta})\gamma_a = 0$. This, with the assumed regularity of G, proves that $\gamma_a = \gamma_a(x)$ and hence

$$\bar{m}^a(\zeta,\bar{\zeta})\nabla_a GG^{-1} = \gamma_a\bar{m}^a(\zeta,\bar{\zeta}). \quad (3.8b)$$

[A more formal way to see this is by applying the edth-bar operator twice on the lhs of (3.5). Using the Sparling equation one can show that it vanishes, which leads to the above result.]

(ii) To show that the above defined γ is anti-self-dual, we construct the following basic equations which are linear combinations of (3.5) and (3.8):

$$L^a(\zeta)\nabla_a GG^{-1} = \gamma_a L^a(\zeta), \quad (3.9a)$$

$$\bar{M}^a(\zeta)\nabla_a GG^{-1} = \gamma_a\bar{M}^a(\zeta), \quad (3.9b)$$

with [via (2.6) and (2.7)]

$$L^a(\zeta) = l^{0a}+\zeta m^{0a}\equiv l^a-\bar{\zeta}\bar{m}^a, \quad (3.10)$$

$$\bar{M}^a(\zeta) = \bar{m}^{0a}+\zeta n^{0a}\equiv\zeta l^a+\bar{m}^a. \quad (3.11)$$

Equations (3.9), derived from the reconstruction theorem, are the Lax pair of the ASDYM equations and are a generalization of the Lax pair of Pohlmeyer, Forgács, Zakharov, and others. Their integrability conditions, namely, $\bar{M}^a(\zeta)\nabla_a$ applied to (3.9a) minus $L^a(\zeta)\nabla_a$ applied to (3.9b), yields $L^{[a}\bar{M}^{b]}F_{ab}=0$, i.e., (2.3), the condition for γ to be anti-self-dual.

We have thus shown the correspondence between solutions of the Sparling equation and solutions of the ASDYM equations. □

B. Left and right gauge transformations

We now return to the Sparling equation and some of the properties of its regular solutions.

An important point is that the Sparling equation remains invariant under $G\to f(x^a)G$, where f is a nonsingular matrix-valued function of x^a only. This transformation of G induces the standard gauge transformation on γ:

$$\gamma_a\to\gamma^*{}_a=f\gamma_a f^{-1}+\nabla_a ff^{-1}.$$

We refer to this transformation of G as the left gauge transformation (LGT). Also, G can be transformed by multiplying it on the rhs by a matrix-valued function $Q\equiv Q(l,\bar{m},\zeta,\bar{\zeta})$, where

$$l\equiv l(x^a,\zeta,\bar{\zeta})=x^a l_a(\zeta,\bar{\zeta}),\quad \bar{m}=x^a\bar{m}_a(\zeta,\bar{\zeta}).$$

We refer to this as the right gauge transformation (RGT). Under a RGT the YM free data transforms as

$$A^*=QAQ^{-1}+P\partial_{\bar{\zeta}}QQ^{-1},$$

but the ASDYM connection γ_a remains invariant,[12] i.e.,

$$\gamma^*{}_a(x^b)=\gamma_a(x^b).$$

Both the LGT and the RGT play an important role. The RGT is simply a standard gauge transformation on the data (which is a component of the connection on the data surface, $\mathscr{I}^+$), while the LGT, as mentioned, is a standard gauge transformation at the field point x^a. Both of these transformations follow from the definition of G as the parallel propagator.

To reconstruct explicitly the γ from Eqs. (3.9), one simply chooses any two values of $(\zeta,\bar{\zeta})$, i.e., $(\zeta_1,\bar{\zeta}_1)$ and $(\zeta_2,\bar{\zeta}_2)$, and substitutes them into Eqs. (3.9), giving four equations to determine $\gamma_a(x^b)$ in terms of the two values of G and their derivatives. If we choose $(\zeta_1,\bar{\zeta}_1)=(0,0)$ and $(\zeta_2,\bar{\zeta}_2)=(\infty,\infty)$, calling $G(x^a,0,0)=G_S(x^a)$ and $G(x^a,\infty,\infty)=G_N(x^a)$, we have

$$\begin{aligned}l^{0a}\gamma_a&=D^0G_S\cdot G_S^{-1},\quad \bar{m}^{0a}\gamma_a=\bar{\delta}^0G_S\cdot G_S^{-1},\\ n^{0a}\gamma_a&=\Delta^0G_N\cdot G_N^{-1},\quad m^{0a}\gamma_a=\delta^0G_N\cdot G_N^{-1}.\end{aligned}\tag{3.12}$$

Two of the three self-dual equations are satisfied identically by Eq. (3.12), while the third becomes, with

$$g=G_N^{-1}G_S,\tag{3.13}$$

the Yang–Pohlmeyer equation

$$\Delta^0(D^0g\cdot g^{-1})-\delta^0(\bar{\delta}^0g\cdot g^{-1})=0.\tag{3.14}$$

[Alternatively, the Yang–Pohlmeyer equation is the integrability conditions on the Lax pair after using (3.12).] There are several things things to note about Eqs. (3.13) and (3.14): They are applicable to any gauge group and furthermore they are gauge independent, as can be seen by applying the LGT to G. If, however, we apply the RGT to G, the solutions of Eq. (3.14) are mapped to formally new solutions:

$$g\to g^*=U^{-1}gV,$$

with

$$U=Q|_{\zeta=\bar{\zeta}=\infty},\quad V=Q|_{\zeta=\bar{\zeta}=0}$$

— the known invariance group of Eq. (3.14).[13]

Remark: There are two simple points of comparison of our work with that of others.

(i) If we make the choice of gauge that $G_N=I$ [accomplished easily from an arbitrary G by the choice of the LGT $f(x^a)=G_N^{-1}(x^a)$] we have the gauge condition most frequently used, namely, $n^{0a}\gamma_a=0$ and $m^{0a}\gamma_a=0$.

(ii) If we now consider $G(x^a,\zeta,\bar{\zeta})$ to be analytic in ζ and $\bar{\zeta}$ separately (now referring to $\bar{\zeta}$ rather than $\bar{\zeta}$) then $G_0(x^a,\zeta)=G(x^a,\zeta,0)$ and $G_\infty(x^a,\zeta)=G(x^a,\zeta,\infty)$ are the two "splitting" functions of a Riemann–Hilbert problem, i.e,

$$G_\infty{}^{-1}G_0=\mathscr{P}(L_ax^a,\bar{M}_ax^a,\zeta),$$

with $\mathscr{P}$ analytic in an annulus on the ζ Riemann sphere [and L and $\bar{M}$ as defined in Eqs. (2.6) and (2.7)]. The variable ζ is then what most workers refer to as the spectral parameter of the problem. We will return to this point in Sec. III C.

C. Relationship with BT's

Returning to the BT, we investigate how the BT, via Eq. (2.9) on the γ and γ', can be used to relate the two Sparling functions $G(x^a,\zeta,\bar{\zeta})$ and $G'(x^a,\zeta,\bar{\zeta})$.

For the pair (G',γ'_a) we can write the equations equivalent to the Lax pair, Eqs. (3.9). By substituting Eqs. (2.9) into (3.9) and using Eqs. (3.9) again, we obtain the following pair of equations to determine the G' from the G:

$$L^a\nabla_aG'G'^{-1}=\mathscr{B}L^a\nabla_aGG^{-1}\mathscr{B}^{-1}+L^a\nabla_a\mathscr{B}\cdot\mathscr{B}^{-1},\tag{3.15a}$$

$$\bar{M}^a\nabla_aG'G'^{-1}=\mathscr{B}\bar{M}^a\nabla_aGG^{-1}\mathscr{B}^{-1}+\bar{M}^a\nabla_a\mathscr{B}\cdot\mathscr{B}^{-1}.\tag{3.15b}$$

From Eqs. (3.15), one sees that the BT for G' is given as

$$G'=\mathscr{B}GQ,\tag{3.16}$$

where Q, a *singular* RGT, is needed in order to make the G' regular, since $\mathscr{B}$ (being a function independent of $\bar{\zeta}$) is singular at $\zeta=\zeta_0$ and $\zeta=-\bar{\zeta}_0{}^{-1}$. (Here $\mathscr{B}$ is redefined by postmultiplying the previous $\mathscr{B}$ [as in Eq. (2.15)] with $H(x^a)$, a nonsingular matrix-valued function of x^a only. See also Eq. (4.1).) The choice of Q depends upon the choice of $\mathscr{B}$. Once a $\mathscr{B}$ is chosen, the corresponding Q is given by

$$Q=Q_0\{(I-E)+\mathscr{F}E\},\tag{3.17}$$

where $\mathscr{F}(L_ax^a,\bar{M}_ax^a,\zeta,\bar{\zeta})$ is the ratio of two twistor functions $F(L_ax^a,\bar{M}_ax^a,\zeta,\bar{\zeta})$ and $\tilde{F}(L_ax^a,\bar{M}_ax^a,\zeta,\bar{\zeta})$, each of which is regular on the $(\zeta,\bar{\zeta})$ sphere and has a single zero at $\zeta=\zeta_0(x)$ and $\zeta=-\bar{\zeta}_0{}^{-1}(x)$, respectively. [Such an F (and $\tilde{F}$) can be constructed, among several ways, by first choosing a twistor function $f(L_ax^a,\bar{M}_ax^a,\zeta)$, which vanishes at $\zeta=\zeta_0(x)$ and is holomorphic in a neighborhood of $\zeta=\zeta_0(x)$; the F is then obtained by replacing the three arguments $(L_a x^a, \bar{M}_a x^a, \zeta)$ in f by $(WL_a x^a, W\bar{M}_a x^a, W\zeta)$, where W is another twistor function of $(L_a x^a, \bar{M}_a x^a, \zeta, \tilde{\zeta})$ chosen so that when $\tilde{\zeta} = \bar{\zeta}$ for all ζ, the values of $(WL_a x^a, W\bar{M}_a x^a, W\zeta)$ remain in a neighborhood of $(L(\zeta_0), \bar{M}(\zeta_0), \zeta_0)$ and $W(\zeta_0, \tilde{\zeta}) = 1 = W(\zeta, 0)$. We obtain $\tilde{F}$ similarly from an $\tilde{f}(L_a x^a, \bar{M}_a x^a, \zeta)$, which vanishes at $\zeta = -\tilde{\zeta}_0^{-1}(x)$. Thus $\mathscr{F}$ has the further property that

$$\mathscr{F}(x^a L_a x^a \bar{M}_a, \zeta, 0) = f(L_a x^a, \bar{M}_a x^a, \zeta)$$

and

$$\mathscr{F}(L_a x^a, \bar{M}_a x^a, \zeta, \infty) = \tilde{f}^{-1}(x^a L_a(\zeta), x^a \bar{M}_a(\zeta), \zeta).$$

We will not show here how such a W can be constructed.]

To make G' regular, the nonsingular $H(x) \in \mathrm{LGT}$ and a regular $Q_0 \in \mathrm{RGT}$ are needed to make HGQ_0 lower triangular at $\zeta = \zeta_0$ and upper triangular at $\zeta = -\tilde{\zeta}_0^{-1}$ for all $\tilde{\zeta}$. One can easily construct such an H and Q_0. We, however, will not discuss the details here since a simpler and more powerful method will be discussed later.

Remark: As an alternative to discussing the action of the BT on the G, the action of the BT on the data of the Sparling equation, $\bar{A}$, can be found as follows: $\bar{A}$ is just a component of the restriction of the YM connection γ_a to $\mathscr{I}^+$. However, since the ASDYM field and connection are conformally invariant, as far as the ASDYM field is concerned, $\mathscr{I}^+$ is on an equal footing with any other null cone in Minkowski space; an inversion can be used to interchange any light cone with $\mathscr{I}^+$. So let us assume that the region of interest is a neighborhood of the vertex of $\mathscr{I}^+$; then the application of the BT, as described in Sec. II, is just as well defined on this neighborhood as it is in the interior of M. We can therefore apply the above technique on this neighborhood and then restrict the connection to $\mathscr{I}^+$ and obtain the new $\bar{A}$. (Despite the fact that we are working at infinity, as far as the YM field is concerned, we are still only working locally.)

Finally, we consider the question of obtaining the new solution $g'(x^a)$ of the Yang–Pohlmeyer equation via the BT. There are two ways of constructing the $g'(x^a)$.

(i) From a regular G' obtained in Eq. (3.16), we at first evaluate G'_N and G'_S and then construct the new $g'(x^a)$ by using Eq. (3.13).

(ii) We use Eq. (3.12), now both for γ and γ', and substitute them in Eq. (2.9), which then generates a set of linear, first-order partial differential equations between the components of (G_N, G_S) and (G'_N, G'_S). The solutions of these equations, along with Eq. (3.13), then yield $g'(x^a)$. This approach is similar to the methods given by several authors[3,4] to obtain the $g'(x^a)$ from a seed $g(x^a)$.

D. The Ward construction

Ward[14] has shown that a solution of the ASDYM equations on a region in M determines and is determined by a holomorphic vector bundle on a corresponding region in twistor space.

This can be seen from the ideas presented above as follows. The space $M \times S^2$ with coordinates (x^a, ζ) for $\zeta \neq \infty$ [or $(x^a, \eta = \zeta^{-1})$ for $\zeta \neq 0$] is the bundle of projective primed (self-dual) spinors. If we allow the coordinates x^a to take on complex values, the space $M \times S^2$ is foliated by the integral two-planes of the vector fields, L^a and $\bar{M}^a$, i.e., the set of self-dual planes. The space of these two-planes, as was mentioned earlier, is a region in (projective) twistor space $PT \cong CP^3$.

The coordinates

$$z_i = (z_0, z_1, z_2) = \{L_a x^a, \bar{M}_a x^a, \zeta\} \tag{3.18}$$

are constant along L^a and $\bar{M}^a$ and are the (holomorphic) coordinates on $\zeta \neq \infty$ in PT; note that PT is naturally a complex manifold on which z_i are holomorphic coordinates. [A neighborhood of $\zeta = \infty$ can be coordinatized by $z'_i = (z_0/\zeta, z_1/\zeta, 1/\zeta)$.]

Given an ASDYM field on a region U in M, we can construct a holomorphic vector bundle on the corresponding region U' in PT. (U' consists of all the self-dual two-planes in $U \times S^2$, if U lies entirely within the Euclidean real slice of M, then U' is just $U \times S^2$.) A holomorphic frame of the vector bundle on U' is determined by a frame of the original vector bundle pulled back to $U \times S^2$, which is holomorphic in ζ and covariantly constant along L^a and $\bar{M}^a$. Such frames exist because the connection is integrable on the self-dual surfaces. This definition is "natural" (i.e., not dependent on a choice of gauge).

Such a holomorphic frame can be represented in terms of some frame on space-time by means of a solution $G(x^a, \zeta)$ of the Lax pair, Eqs. (3.9), which is holomorphic in ζ:

$$L^a(\zeta)\nabla_a G G^{-1} = \gamma_a L^a(\zeta), \tag{3.9a}$$

$$\bar{M}^a(\zeta)\nabla_a G G^{-1} = \gamma_a \bar{M}^a(\zeta). \tag{3.9b}$$

[Since the operators in (3.9) do not depend on $\bar{\zeta}$, we consider in the above context solutions $G(x^a, \zeta)$ which are independent of $\bar{\zeta}$.]

Note that a solution $G(x, \zeta)$ of (3.9) is unique up to postmultiplication $G \to GF$ by some matrix function $F(z_i)$ of the twistor coordinates z_i alone.

E. The patching function

The matrix function $G(x, \zeta)$ contains more information than just that of the holomorphic bundle on PT. To specify the bundle, we need only the "patching function" $\mathscr{P}$. This is a matrix function on PT that translates from a frame defined on $|\zeta| > \frac{1}{2}$ to one defined on $|\zeta| < 2$. In particular we consider two solutions of (3.9): G_0, which is to be regular over the region $\zeta \neq \infty$, say $|\zeta| \leqslant 2$; and G_∞, which is to be regular over the region $\zeta \neq 0$, say $|\zeta| > \frac{1}{2}$. It is seen easily from the fact that the only freedom in the solution of (3.9) is right multiplication by a matrix constant along L^a and $\bar{M}^a$ that G_0 and G_∞ are related by

$$G_0 = G_\infty \mathscr{P}, \tag{3.19}$$

with $\mathscr{P} = \mathscr{P}(z_i)$. (See the remark of Sec. III E.) The matrix $\mathscr{P}$ is the patching or transition matrix which characterizes the bundle most economically; it provides the transition between the components of a section in the frame determined by G_0 in its region of definition and the components in the frame G_∞ in its region of definition.

Remarks: (i) An attractive feature of the twistor description of ASDYM fields is that the patching function is effectively freely prescribable. We shall see that it plays a similar role in the twistor description, as does the function $\bar{A}(u,\zeta,\bar{\zeta})$ in the Sparling approach.

(ii) The patching function $\mathscr{P}$ can only be reduced to the identity if a solution $G(x,\zeta)$ can be found which is regular for all ζ including ∞. This would, by Liouville's theorem, imply that the G were independent of ζ. This, however, is only possible when the ASDYM field is flat.

F. Reconstruction of the field from the patching function

In the previous paragraph we discussed the construction of the patching function $\mathscr{P}$ from knowledge of G_0 and G_∞. The converse problem of beginning with a given patching function $\mathscr{P}(z_i)$ and constructing from it the two functions $G_0(x^a,\zeta)$ and $G_\infty(x^a,\zeta)$ in their respective domains, satisfying (3.19), is the heart of Ward's method of solving the ASDYM equations. In order to do this the x^a in (3.19) is held constant, the $\mathscr{P}$ becomes a function of ζ, and the x^a play the role of parameters. So, given the data $\mathscr{P}$, we must solve

$$G_0(x,\zeta) = G_\infty(x,\zeta)\mathscr{P}(L_a x^a,\bar{M}_a x^a,\zeta)$$

for G_0 and G_∞, with G_0 defined for $|\zeta|<2$ and G_∞ defined for $|\zeta|>\frac{1}{2}$. The problem of splitting the $\mathscr{P}$ and thus finding the two G's is the Riemann–Hilbert problem. Generically solutions exist.

The YM connection γ can now be reconstructed from G by reproducing the Lax pair equations (3.9). This is analogous to the reconstruction in the Sparling case. The proof that the Lax pair exists follows from (3.19) and uses a generalization of Liouville's theorem; since $L^a\nabla_a\mathscr{P}=0$, we can write

$$\gamma_a L^a(\zeta) = L^a(\zeta)\nabla_a G_0 G_0^{-1} = L^a(\zeta)\nabla_a G_\infty G_\infty^{-1}.$$

We see that the first equation implies that $\gamma\cdot L$ is regular for $|\zeta|<2$ and the second implies that $\gamma\cdot L$ has a simple pole at $\zeta=\infty$; thus $\gamma\cdot L$ must be linear in ζ and is therefore

$$L\cdot\gamma = l^0\cdot\gamma + \zeta m^0\cdot\gamma,$$

with $l^0\cdot\gamma$ and $m^0\cdot\gamma$ independent of ζ; this yields the l^0 and m^0 components of the connection. An identical argument applied to the $\bar{M}$ equation yields $n^{0a}\gamma_a$ and $\bar{m}^{0a}\gamma_a$. These connection components are identical to those obtained by the Sparling approach [Eq. (3.12)]. This once again shows the correspondence between the Ward and Sparling versions of the twistor construction.

Remarks: (i) In both the Ward and Sparling versions of the twistor procedures, the data are effectively freely prescribable. Although it is on occasion possible to analyze the ASDYM field directly from its data on twistor space it is often desirable to evaluate the field explicitly on space-time. Unfortunately there is no general explicit method of solving directly either Eqs. (3.2) or (3.19). One falls back on either special *Ansätze* (cf. Ref. 10) or solution generating techniques such as the BT's of, for example, this paper.

(ii) The Sparling equation approach to the ASDYM equations can be thought of as a Dolbeault version of the Ward construction.

In a Dolbeault approach one represents a holomorphic vector bundle using a frame on PT that depends on z_i and $\bar{z}_i$, but which has the advantage of being global on the $(\zeta,\bar{\zeta})$ Riemann sphere. The matrix functions $H(z_i,\bar{z}_i)$, which rotate from the given frame to a holomorphic one, are characterized as the solutions of a first-order linear partial differential equation, namely, the vanishing of the covariant $\bar{\partial}$ operator acting on H;

$$\bar{\partial}_{\bar{A}}H \equiv \frac{\partial H}{\partial\bar{z}_i}d\bar{z}_i + H\bar{A}^i d\bar{z}_i = 0$$

for some $\bar{A}^i$ chosen so that $\bar{\partial}_{\bar{A}}^2=0$ (this last condition is the integrability condition for the existence of solutions to $\bar{\partial}_{\bar{A}}H=0$).

In the asymptotic formalism developed earlier, such a nonholomorphic global frame is obtained relative to $\mathscr{I}^+$ as follows: Each twistor plane (i.e., integral surface of L^a and $\bar{M}^a$) given by $z_i=$ const on (x^a,ζ) space intersects $\mathscr{I}^+$ at the point

$$(u,\zeta,\bar{\zeta}) = ((2P\sqrt{2})^{-1}(z_0+\bar{\zeta}z_1),\zeta,\bar{\zeta}).$$

(Note that in general u is complex.) An ordinary frame for the YM field at each point of $\mathscr{I}^+$ determines one on PT; the frame at a point z_i in PT is given by that at the point at which the corresponding twistor plane intersects $\mathscr{I}^+$. This frame is global in ζ, but depends on $\bar{\zeta}$ as well as z_i. A rotation from the given global frame to a local holomorphic frame is given by $H(u,\zeta,\bar{\zeta})$ if it is covariantly constant along $\partial_{\bar{\zeta}}$, that is,

$$\partial_{\bar{\zeta}}H = -H\bar{A}(u,\zeta,\bar{\zeta}),$$

where $[u=(2P\sqrt{2})^{-1}(z_0+\bar{\zeta}z_1),\zeta,\bar{\zeta}]$ is the point where the twistor z_i intersects $\mathscr{I}^+$ and $\bar{A}$ is the coefficient of $d\bar{\zeta}$ in $\gamma_a\, dx^a|_{\mathscr{I}^+}$. As can be seen, this is a version of the above $\bar{\partial}$ operator in a special gauge.

In order to evaluate the field on space-time, it was necessary to find a global holomorphic frame over each Riemann sphere in PT given by holding $x^a=$ const in (3.18). For a fixed x^a the matrix that rotates from the given nonholmorphic frame to a holomorphic one on the Riemann sphere reduces to finding solutions $G(x,\zeta,\bar{\zeta})$ of the Sparling equation that are global in $(\zeta,\bar{\zeta})$.

(iii) Reality structures. The version of the ASDYM equations we have been using has been written with the Lorentzian signature. To obtain the perhaps more familar Euclidean version, we send $t\to it$. If we then put

$$2v = t+iz \quad\text{and}\quad 2w = x+iy,$$

we have $D^0=-i\partial/\partial\bar{v}$, $\Delta^0=-i\,\partial/\partial v$, $\delta^0=-\partial/\partial w$, $\bar{\delta}^0=-\partial/\partial\bar{w}$, so that the Yang–Pohlmeyer equation (3.14) becomes

$$(g_{,\bar{v}}g^{-1})_{,v} + (g_{,\bar{w}}g^{-1})_{,w} = 0.$$

Usually one is interested in producing real solutions to the YM equations. The appropriate reality conditions when the field is anti-self-dual are those for Euclidean or (2,2) signature (i.e., g should be Hermitian on the appropriate real slice).

The complex conjugation on complex Minkowski space that preserves the Euclidean slice induces a conjugation on PT, given in the above coordinates by

$$(z_0,z_1,z_2)\to(\hat{z}_0,\hat{z}_1,\hat{z}_2) = (-\bar{z}_1,\bar{z}_0,-\bar{z}_2^{-1}),$$

which we can write as $z_i\to\hat{z}_i$, $i=0,1,2$. The conjugation is antiholomorphic and has no fixed points.

A bundle on twistor space is real if its pullback under the conjugation is its Hermitian conjugate. In particular, one can choose holomorphic frames on $|z|<2$ and $|z|>\frac{1}{2}$ so that the twistor data P satisfy

$$\overline{\mathscr{P}}(z)=\mathscr{P}^{t}(\hat{z}),$$

where t denotes transpose. Alternatively one can choose a RGT for the Sparling equation such that

$$A(z_i,\bar{\zeta})=(\bar{\zeta}/\zeta)\overline{A}^{t}(\hat{z}_i,-1/\zeta).$$

[The signature (2,2) case is the same as above, except that

$$\hat{z}_0=+\bar{z}_1 \text{ and } \hat{z}_2=+\bar{z}_2^{-1}.]$$

IV. BT's AND THE TWISTOR DATA

In this section we will translate the BT's of Sec. II into the twistor formalism. We start with a patching function for which it is possible to solve the Riemann–Hilbert problem. The BT will be seen to correspond to inserting manually new singularities which are sufficiently simple into the patching function $\mathscr{P}$ (in fact simple poles and zeros) such that it is again possible to solve the new Riemann–Hilbert problem. As mentioned previously, this process only requires algebraic manipulations.

Let $\mathscr{P}(z_i)$ be some patching matrix for which we know how to solve Eq. (3.19), i.e., we have $\mathscr{P}=G_\infty{}^{-1}G_0$, with $G_0(x^a,\zeta)$ nonsingular for $|\zeta|<2$ and $G_\infty(x^a,\zeta)$ nonsingular for $|\zeta|>\frac{1}{2}$. We now consider a new patching matrix $\mathscr{P}'=B_\infty{}^{-1}\mathscr{P}B_0$. For $\mathscr{P}'$ to also be a patching matrix it is required that B_∞ and B_0 should be holomorphic in the twistor coordinates and nonsingular in the range $\frac{1}{2}<|\zeta|<2$. (By a nonsingular matrix we mean a matrix with all its entries regular and which is invertible.) We also require of B_0 and B_∞ that (a) $\mathscr{P}'$ is not gauge equivalent to $\mathscr{P}$, and (b) from knowledge of the splitting of Eq. (3.19) for $\mathscr{P}$, we can find the solution or splitting of Eq. (3.19) for $\mathscr{P}'$.

The matrix $\mathscr{P}'$ is gauge equivalent to $\mathscr{P}$ if B_∞ is nonsingular for $|\zeta|>\frac{1}{2}$ and B_0 is nonsingular for $|\zeta|<2$. So for $\mathscr{P}'$ to be an inequivalent patching matrix, B_∞ must be singular for some ζ with $|\zeta|>2$ and B_0 must be singular for some ζ with $|\zeta|<\frac{1}{2}$. We shall make such an *Ansatz* for B_∞ and B_0, from which condition (b) follows also. However, in order to motivate the *Ansätze* we first explain the procedure.

We have

$$\mathscr{P}=G_\infty{}^{-1}G_0$$

and we wish to find G'_0 and G'_∞ with the appropriate regularity properties such that

$$\mathscr{P}'=G_\infty{}'^{-1}G_0{}'$$

and hence

$$\mathscr{P}'=B_\infty{}^{-1}\mathscr{P}B_0=B_\infty{}^{-1}G_\infty{}^{-1}G_0B_0=G_\infty{}'^{-1}G_0{}'.$$

As a result of the singularities in B_∞ on $|\zeta|>2$, $G_\infty B_\infty$ is not regular on $|\zeta|>2$ and so cannot be a good candidate for $G_\infty{}'$. However, if the singularity structure of B_∞ is simple enough, we can compensate by finding a (singular) matrix function $\mathscr{B}(x^a,\zeta)$ and setting

$$G_\infty{}'=\mathscr{B}G_\infty B_\infty \text{ and } G_0{}'=\mathscr{B}G_0B_0.$$

If $\mathscr{B}$ can be chosen so that $G_0{}'$ and $G_\infty{}'$ are regular on the relevant regions, then we are done.

First choose a constant projector matrix $E=\mathrm{diag}(I,0)$, where I is the $r\times r$ identity matrix and 0 is the $s\times s$ zero matrix $(r+s=n)$. We shall decompose all our matrices into 2×2 block form with respect to this E. We take

$$B_\infty=\begin{pmatrix}1 & 0\\ 0 & \tilde{f}\end{pmatrix},\quad B_0=\begin{pmatrix}1 & 0\\ 0 & f^{-1}\end{pmatrix},$$

where f is a twistor function holomorphic on $|\zeta|<2$, with only simple zeros on $|\zeta|<\frac{1}{2}$ and $\tilde{f}$ is a twistor function holomorphic on $|\zeta|>\frac{1}{2}$, with simple zeros only in $|\zeta|>2$.

We now choose $\mathscr{B}$ so as to eliminate the zeros in $G_\infty B_\infty$ on $|\zeta|>2$ and the poles of B_0 in $|\zeta|<\frac{1}{2}$. The zero sets of f and $\tilde{f}$ can be written as $\zeta-\zeta_0(x^a)=0$ and $1+\tilde{\zeta}_0(x)\zeta=0$, respectively. We first choose (for gauge fixing) a matrix function of x^a, $H(x^a)$ such that

$$EH(x)G_0(x^a,\zeta_0(x))(1-E)=0$$

and

$$(1-E)H(x)G_\infty(x^a,-\tilde{\zeta}_0{}^{-1}(x))E=0.$$

In the block decomposition above this implies that HG_0 is lower triangular at $\zeta=\zeta_0$ and HG_∞ is upper triangular at $\zeta=-\tilde{\zeta}_0^{-1}$. [If we can write $G_\infty(x,\zeta_0)$ and G_0 in block form as

$$G_\infty(x,-\tilde{\zeta}_0{}^{-1})=\begin{pmatrix}\tilde{A}_0 & \tilde{B}_0\\ \tilde{C}_0 & \tilde{D}_0\end{pmatrix},\quad G_0(x,\zeta_0)=\begin{pmatrix}A_0 & B_0\\ C_0 & D_0\end{pmatrix},$$

then it is easily checked that such an H exists when $\tilde{A}_0$ and D_0 are invertible. These conditions only require the nonvanishing of two determinants and so will be satisfied generically.] This implies that we can write

$$HG_0=\begin{pmatrix}A & (\zeta-\zeta_0)B\\ C & D\end{pmatrix}$$

for some matrices A, B, C, and D regular on $|\zeta|<2$. (HG_∞ can be similarly represented.)

We now put

$$\mathscr{B}=\{E+\Lambda(1-E)\}H=\begin{pmatrix}1 & 0\\ 0 & \Lambda\end{pmatrix}H, \tag{4.1}$$

where $\Lambda=(\zeta-\zeta_0(x))(1+\tilde{\zeta}_0(x)\zeta)^{-1}$.

A short calculation shows that $\mathscr{B}G_0B_0$ is now regular on $|\zeta|<2$ and $\mathscr{B}G_\infty B_\infty$ is regular on $|\zeta|>\frac{1}{2}$. For $\mathscr{B}G_0B_0$ we obtain

$$\begin{aligned}\mathscr{B}G_0B_0&=\{E+\Lambda(1-E)\}HG_0\{E+f^{-1}(1-E)\}^{-1}\\ &=\begin{pmatrix}1 & 0\\ 0 & \Lambda\end{pmatrix}\begin{pmatrix}A & (\zeta-\zeta_0)B\\ C & D\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & f^{-1}\end{pmatrix}\\ &=\begin{pmatrix}A & (\zeta-\zeta_0)f^{-1}B\\ \Lambda C & \Lambda f^{-1}D\end{pmatrix}.\end{aligned}$$

We see that the vanishing of the determinant of $\mathscr{B}$ cancels the singularity of that of B_0 and the choice of H eliminates the pole one would otherwise have in the top right entry of the above matrix. Similar remarks hold for $\mathscr{B}G_\infty B_\infty$, with zeros interchanged with poles. Now $\mathscr{B}(x,\zeta)$ is the same matrix as was used in Sec. II to first define the BT.

If we wish to preserve reality conditions we must have that the twistor function $\tilde{f}(z_i)$ is determined by the following relation:

$$\tilde{f}(z_i)=\bar{f}(\hat{z}_i).$$

This, in turn, implies $\tilde{\zeta}_0(x) = \bar{\zeta}_0(x)$.

Remarks: (i) This procedure does not respect the gauge invariance on twistor space, i.e., if $\mathscr{P}$ and $\mathscr{Q}$ are gauge equivalent patching functions, $B_0\mathscr{P}B_\infty{}^{-1}$ and $B_0\mathscr{Q}B_\infty{}^{-1}$ are not, in general, gauge equivalent. This allows us to generate more solutions, since if we can split $\mathscr{P} = G_\infty{}^{-1}G_0$, then we can split $\mathscr{Q}$; thus we can perform the above procedure on $\mathscr{Q}$ to obtain a new, distinct solution.

This fact does, however, mean that the BT's only generate an infinite-dimensional group on the space of patching functions and not on the space of solutions of the SDYM equations.

(ii) As it stands, the above transformations are a special case of those in Forgács *et al.*[6] To obtain the general ones, one must repeat the above transformations n times, alternating with gauge transformations on the patching matrix.

V. DISCUSSION AND CONCLUSIONS

The results we have presented here are an extension and generalization of the work of many others.[1–5,7,9,15–18] This work frequently involves the application of BT's to solutions $g(x^a)$ of the SL(2,C) or SL(3,C) Yang–Pohlmeyer equation, obtaining new solutions $g'(x^a)$ or the application of BT's to the direct problem of the solution of the Lax pair, Eqs. (3.9). An important feature of our work is that these methods are reduced to algebraic manipulations and apply to arbitrary groups and arbitrary seed solutions. It also relates, in a geometric manner, many of the solution generating techniques and ideas.

One of the more important features of our work is that it is now clear what the appropriate generalization of the infinite-dimensional transformation groups on the solution space of various integrable systems is in the case of ASDYM fields. (Recall that in many integrable systems, the BT's generate a group that acts on either the solution space or some extension thereof. These groups are typically loop groups, with the infinite-dimensional groups consisting of maps from the circle to some semisimple finite-dimensional group. Their Lie algebras are Kac–Moody Lie algebras.) In the complex case, the appropriate extension of the solution space to consider is the space Π of patching functions $\mathscr{P}$, nonsingular matrix functions of z_i on the domain $\frac{1}{2} < |\zeta| < 2$. It is straightforward to convince oneself that BT's, as presented in Sec. IV, together with allowable gauge transformations, are dense in the space Γ of pairs of matrix functions $(B_0(z_i), B_\infty(z_i))$ defined on $\frac{1}{2} < |\zeta| < 2$. These act on $\mathscr{P}$ by

$$\mathscr{P} \to B_\infty{}^{-1}\mathscr{P}B_0.$$

With Euclidean reality conditions, we must also have $B_\infty(z_i) = \bar{B}_0(\hat{z}_i)$ and $P(z_i) = \bar{P}(\hat{z}_i)$. This group is clearly transitive in the complex case.

An example of how the above ideas relate to a perhaps more familar integrable system is as follows: In the case of stationary axisymmetric SL(2,R) solutions (satisfying a certain extra symmetry condition) the ASDYM equations are equivalent to the stationary axisymmetric Einstein vacuum equations. This is a well-known integrable system whose BT's generate a group known as the Geroch group. This group is generated by the basic BT's with $\zeta_0 = \bar{\zeta}_0 = 0$ (compensated by a gauge transformation to preserve explicit axisymmetry) together with constant gauge transformations.[7] Kinnersly and Chitre[16] showed that the Lie algebra of this group is a Kac–Moody type algebra associated to a type of loop group. In Woodhouse and Mason[18] this group is shown to be related closely to the subgroup of Γ that depends only on the holomorphic stationary axisymmetric coordinate $\alpha = z_0 + z_1/\zeta$ on PT.

This same technique [for the SL(2,C) ASDYM case] also generates the instanton solutions via the Atiyah and Ward[10] *Ansätze*.

ACKNOWLEDGMENTS

The authors thank the relativity groups at the University of Pittsburgh and Mathematics Institute, Oxford University for valuable comments. One of us (LM) would like to thank in particular NMJ Woodhouse for scientific comments and the Andrew Mellon Foundation and Esmée Fairbairn for financial support.

This work was supported by the National Science Foundation under Grant No. PHY80023.

[1] P. Forgács, *Non-Linear Equations in Classical and Quantum Field Theory*, edited by N. Sanchez (Springer, Berlin, 1985).

[2] A. A. Belavin and V. E. Zakharov, Phys. Lett. B **73**, 53 (1978).

[3] K. Pohlmeyer, Commun. Math. Phys. **72**, 37 (1980).

[4] M. K. Prasad, A. Sinha, and L. L. Wang, Phys. Lett. B **87**, 237 (1979).

[5] E. T. Newman, Phys. Rev. D **22**, 3023 (1980).

[6] P. Forgács, Z. Horváth, and L. Palla, Phys. Rev. D **23**, 1876 (1981).

[7] N. M. J. Woodhouse, Class. Quant. Grav. **4**, 799 (1987).

[8] E. F. Corrigan, D. B. Fairlie, R. G. Yates, and P. Goddard, Commun. Math. Phys. **58**, 223 (1978).

[9] M. K. Prasad, A. Sinha, and L. C. Wang, Phys. Rev. D **23**, 2321 (1981).

[10] M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977).

[11] E. T. Newman, J. Math. Phys. **27**, 2797 (1986).

[12] S. Chakravarty and E. T. Newman, J. Math. Phys. **28**, 334 (1987).

[13] R. Ward, Gen. Relativ. Gravit. **15**, 105 (1983).

[14] R. Ward, Commun. Math. Phys. **80**, 563 (1981).

[15] R. Geroch, J. Math. Phys. **13**, 394 (1972).

[16] W. Kinnersly and D. M. Chitre, J. Math. Phys. **18**, 1538 (1977); **19**, 1926, 2037 (1978).

[17] D. Kramer and G. Neugebauer, Ann. Phys. (Leipzig) **24**, 62 (1969).

[18] N. M. J. Woodhouse and L. J. Mason, "The Geroch group and non-Hausdorff twistor spaces," to appear in Nonlinearity.

A remark on BRST quantization

P. Thomi
Institute for Theoretical Physics, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland

(Received 30 September 1987; accepted for publication 18 November 1987)

Using rigged Hilbert space techniques, the scalar product on the BRST cohomology for certain bosonic systems is rigorously defined.

I. INTRODUCTION

In the last few years BRST methods have become an important tool in the attempts to covariantly quantize theories with gauge symmetries, e.g., strings.[1,2] In the path integral formalism the methods introduced by Fradkin and his collaborators[3] considerably enlarge the scope of the formalism as compared to the Fadeev–Popov description. For instance, systems of rank $\geqslant 1$ not covered by the Fadeev–Popov procedure (such as gravity) can be treated. In Yang–Mills (YM) theories, gauges with extra ghosts can be chosen. Also the operator formalism is very useful, e.g., it provides one of the simplest ways to arrive at the critical parameters of string theory.[2]

We will be concerned here with the operator formalism, namely, with the BRST cohomology and its inner product structure.

In BRST formalism one works in an extended phase space containing the usual canonical variables and the ghosts. This extended phase space carries a Poisson bracket structure. In quantum theory one has to represent the operator algebra corresponding to this structure on a space of states. As a result of the gauge symmetry of the considered systems this space is too large. To arrive at the true physical space of states one has to impose certain constraints and make some identifications. The BRST formalism gives a prescription showing how to do this. There exists an operator $\hat{\Omega}$ acting on the "big" space that is Hermitian w.r.t. the Hermitian form on the big space and nilpotent, i.e., $\hat{\Omega}^2 = 0$. Note that in order for this to be possible the Hermitian form has to be nondefinite. A state is physical if it is annihilated by the BRST charge $\hat{\Omega}$. Moreover, one has to identify states that differ by $\hat{\Omega}|\chi\rangle$. Since $\hat{\Omega}$ is Hermitian and nilpotent we have $\mathrm{Im}\,\hat{\Omega} \subset \ker \hat{\Omega}$ and the true physical space is really the cohomology of the BRST operator, H^*_{BRST}. The states in $\mathrm{Im}\,\hat{\Omega}$ all have zero norm and are orthogonal to every physical state, i.e., they completely decouple. We see that in order for all this to work in the indicated way, it is essential to have on the big space of states a well-defined Hermitian form w.r.t. which $\hat{\Omega}$ is Hermitian.

One would like the inner product on the physical space to be induced from the Hermitian form on the big space. Further, one wants the BRST formalism to give the correct quantum theory for systems where it is known, such as YM theories and the relativistic particle. Of course, one wants the whole procedure to be unique (at least up to equivalence) and well defined.

There are problems with this, however, even in such simple systems as the free relativistic particle. Usually one takes as the Hermitian form on the big space [compare (6)]:

$$(\psi,\phi) = \int dp\, d\eta\, \psi^* \phi. \tag{$*$}$$

It is stated in the literature that for simple systems (i.e., when no topological complications occur) H^*_{BRST} is isomorphic to the zero ghost states in $\ker \hat{\Omega}$. The problem is that the above Hermitian form is usually not well defined for these. Indeed, for the particle the zero ghost states in $\ker \hat{\Omega}$ have to satisfy the Klein–Gordon equation and hence do not belong to L^2. Rather they are proportional to $\delta(p^2 + m^2)$. The Hermitian form is thus proportional to $\delta(0)$ (arising from the p integration) times zero (arising from the Berezin integration over η). One way out is that the ghosts naturally regulate it, but "so far, this hope has only been substantiated by heuristic arguments."[4] Another is to define the scalar product only for equivalence classes of physical states. However, having no Hermitian form on the big space one loses the argument that shows that the states in $\mathrm{Im}\,\hat{\Omega}$ completely decouple, and thus their factoring out is not very well motivated any more. For further discussions on this, see Ref. 4 and Sec. 8 of Ref. 5. A third way is given in this paper.

Another problem is that H^*_{BRST} depends on the boundary conditions one chooses the states to satisfy. This was first noted in Ref. 6.

We found a way to rigorously define the Hermitian form on the big space for bosonic systems. In order to be able to explicitly calculate H^*, we had to make the simplifying assumption that the constraints only depend on the momenta. For such systems the constraints automatically commute strongly. This form (constraints only depending on the momenta) can be achieved locally for any bosonic system by a canonical transformation in the extended phase space.[5] However, topological obstructions may prevent one from bringing an arbitrary system to this form globally, and thus with the above assumption we restrict ourselves to a subclass of bosonic systems. This subclass includes the relativistic particle and therefore also the zero modes of the bosonic string equivalent to the particle.[6]

We start from ($*$), expand the wave functions in the ghosts and do the Berezin integral. We arrive at an expression where sums of products of the form $\overset{(j)}{\psi}\,\overset{(2m-j)}{\phi}$ are integrated over. Here $\overset{(j)}{\psi}$ is a coefficient function of j ghosts in the expansion of ψ and $2m$ is the number of ghosts [see (15)]. This is well defined if whenever $\overset{(j)}{\psi}$ is a distribution then $\overset{(2m-j)}{\phi}$ is a test function (or vice versa) or when both are in L^2. So the Hermitian form on the big space is well defined if the coefficient functions $\overset{(j)}{\psi}$ are chosen from suitable parts of a rigged Hilbert space (Proposition 1). There are several ways of doing this, and in general they lead to different $H^*_{\rm BRST}$. This is similar to the dependence of $H^*_{\rm BRST}$ on boundary conditions mentioned above. We define a character function $\chi(j)$ (Definition 2) that indicates in which function spaces the coefficient functions of j ghosts lie. We show that for all $\chi(j)$ that lead to reasonable H^* [these are the monotonically decreasing ones with $\chi(0)=1$] the BRST cohomology is $H^* = S'(\mathscr{C}) \oplus S(\mathscr{C})$, where $\mathscr{C}$ is the constraint surface (Theorems 1 and 1'). Thus the arbitrariness of χ does not show up in the final result, all (consistent) choices of χ are equivalent. We see that H^* has (more than) twice (because the first summand is S', not S) as many degrees of freedom as the physical space in more traditional quantization schemes, which is $S(\mathscr{C})$. This "doubling" is an intrinsic feature of our formalism, it occurs even for trivial topology. If topology is nontrivial, additional doublings may occur, but this is not considered here. The induced Hermitian form on H^* is given in (28). It is not positive definite.

In order to have a positive definite inner product we have to choose a linear subspace Π of H^*. This Π must be left invariant by the algebra of BRST observables, especially by the Hamiltonian. There is a whole infinity of such Π's but they all lead to mutually equivalent quantum theories. These quantum theories are also equivalent to the usual quantum theory in systems in which we know what the correct quantum theory is. Thus by selecting the subspace we also remove the doubling. So to speak, the two problems, doubling and nondefiniteness of the induced form, cancel each other. Finally we have to complete Π w.r.t. the induced inner product to arrive at the physical Hilbert space. Although there is some arbitrariness along the way, the final Hilbert space is *unique*.

One should, of course, try to extend these methods to more general systems. For instance, one should look at theories of arbitrary rank. For these, factor-ordering problems in the BRST operator can lead to complications. It would also be interesting to examine the consequences of nontrivial topology and second-class constraints. An important issue is to generalize our procedure to fermionic systems. We have some partial results that indicate that similar methods work at least in the case of the spinning particle of Galvao and Teitelboim.[7] But this is left for a future publication.

The paper is organized as follows: In Sec. II we give a short review of BRST quantization of bosonic systems. This is mainly intended to fix notation. In Sec. III we define the Hermitian form and determine $H^*_{\rm BRST}$. In Sec. IV we discuss the selection of the subspace on which the induced form is positive definite. We show that consistent choices exist and that they are all equivalent.

Excellent reviews on the subject of Hamiltonian BRST and BFV methods have been written by Henneaux.[5,8] Most of our notation is taken over from there.

II. BRST QUANTIZATION FOR BOSONIC SYSTEMS

We denote by (q^μ,p_μ), $\mu=1,\dots,n$, the phase-space variables, by ϕ_α, $\alpha=1,\dots,m$, the constraints, and by λ^α the associated Lagrange multipliers. The phase-space variables fulfill the usual Poisson bracket relations. It is convenient sometimes to treat the Lagrange multipliers on the same footing as the canonical variables. To this end we add to the phase space the pairs $(\lambda^\alpha,\pi_\alpha)$ satisfiying $[\pi_\beta,\lambda^\alpha] = -\delta^\alpha_\beta$. In order not to change the dynamical content of the theory we have to impose the additional constraints $\pi_\alpha=0$. We denote the constraints (π_α,ϕ_α) collectively by G_a, $a=1,\dots,2m$. We assume that the constraints depend only on the momenta and thus have vanishing Poisson brackets. To every constraint we add a pair $(\eta^a,\mathscr{P}_a)$, the ghosts and their momenta, of anticommuting variables. They satisfy

$$[\eta^a,\mathscr{P}_b] = [\mathscr{P}_b,\eta^a] = -\delta^a_b, \qquad (\eta^a)^* = \eta^a, \quad (\mathscr{P}_a)^* = -\mathscr{P}_a. \tag{1}$$

Since by assumption $[G_a,G_b]=0$, the classical BRST charge Ω reads

$$\Omega = \eta^a G_a. \tag{2}$$

It satisfies

$$[\Omega,\Omega]=0. \tag{3}$$

To pass to quantum theory we replace the (generalized) Poisson brackets by i times the (anti)commutators. So we have to represent the following operator algebra:

$$[\hat{q}^\mu,\hat{p}_\nu] = i\delta^\mu_\nu, \quad [\hat{\lambda}^\alpha,\hat{\pi}_\beta] = i\delta^\alpha_\beta, \quad [\hat{\mathscr{P}}^a,\hat{\eta}_b] = -i\delta^a_b. \tag{4}$$

Real classical variables must be represented by Hermitian operators, imaginary ones by anti-Hermitian operators. We consider what might be called the (p,π,η) representation on functions $\psi=\psi(p,\pi,\eta)$. Of course, there are also other representations, but this is the most convenient one for our purposes. The operators act as

$$\hat{q}^\mu = i\frac{\partial}{\partial p_\mu}, \quad \hat{p}_\mu = p_\mu, \quad \hat{\lambda}^\alpha = i\frac{\partial}{\partial \pi_\alpha}, \quad \hat{\pi}_\alpha = \pi_\alpha, \qquad \hat{\eta}^a = \eta^a, \quad \hat{\mathscr{P}}_b = -i\frac{\partial^l}{\partial \eta^b}. \tag{5}$$

These operators have the right Hermiticity properties w.r.t. the Hermitian form

$$(\psi,\phi) = i^k \int dp\, d\pi\, d\eta\, \psi^*\phi, \tag{6}$$

where $k=m \pmod 2$. This factor has to be included to have

$$(\psi,\phi)^* = (\phi,\psi).$$

Expression (6) is what is usually given in the literature.[5] So far it is only a formal expression; it will be rigorously defined in Sec. III.

Since all variables occurring in the classical expression for Ω have vanishing Poisson brackets, we have no factor-ordering ambiguities and therefore we may simply substitute the operators (5) for the classical variables in the expression (2) for the classical BRST charge. We thus have

$$\hat{\Omega} = \hat{\eta}^a\hat{G}_a. \tag{7}$$

Nilpotence of $\hat{\Omega}$, i.e.,

$$\hat{\Omega}^2 = 0, \tag{8}$$

is readily checked as well as Hermiticity w.r.t. the Hermitian form (6). The condition for a state to be physical is

$$\hat{\Omega}|\psi\rangle_{\text{phys}} = 0, \tag{9}$$

and one has to identify $\hat{\Omega}|\chi\rangle$ with 0, so

$$|\psi\rangle_{\text{phys}} \sim |\psi\rangle_{\text{phys}} + \hat{\Omega}|\chi\rangle, \tag{10}$$

and thus the physically relevant space is

$$\ker \hat{\Omega}/\sim\, = \ker \hat{\Omega}/\text{Im}\, \hat{\Omega} = H^*_{\text{BRST}}, \tag{11}$$

i.e., the cohomology of the BRST operator. Since $(\psi,\phi) = 0$, $\forall \psi \in \ker \hat{\Omega}$, $\phi \in \text{Im}\, \hat{\Omega}$, the Hermitian form (6) induces a Hermitian form on H^* which we again denote by $(\cdot,\cdot)$. We require it to be nondegenerate, but we do not require it to be positive definite.

Also, the observables are grouped into equivalence classes. An operator $\hat{A}$ is a BRST observable if[5]

(1) it is Hermitian and even,
(2) it has ghost number zero, i.e., commutes with the ghost number operator, (12)
(3) $[\hat{\Omega},\hat{A}] = 0$.

We identify two observables if they differ only by $[\hat{K},\hat{\Omega}]$, i.e.,

$$\hat{A} \sim \hat{A}' = \hat{A} + [\hat{K},\hat{\Omega}], \tag{13}$$

where $\hat{K}$ is any operator that lowers the ghost number by one unit. The action of the BRST observables is well defined on H^* since they map physical states into physical states and map Im $\hat{\Omega}$ into itself as follows from (12) (3). For more details see Ref. 5.

III. THE SCALAR PRODUCT AND H^*_{BRST}

In the expression (6) for the Hermitian form a Berezin integration over the anticommuting variables η occurs. In order to perform this integral we expand the wave functions in the η^a's:

$$\psi(p,\pi,\eta) = \sum_{j=0}^{2m} \sum_{a_i} \psi_{a_1,\ldots,a_j}(p,\pi)\eta^{a_1}\cdots\eta^{a_j}. \tag{14}$$

The coefficient functions are antisymmetric in $a_1,\ldots,a_j$. Putting this in (6) and using $(\eta^a\eta^b)^* = (\eta^b)^*(\eta^a)^* = \eta^b\eta^a$, we get

$$(\psi,\phi) = i^k \int dp^n\, d^m\pi \sum_{j=0}^{2m} \sum_{a_i=1}^{2m} \psi^*_{a_j\cdots a_1}\phi_{a_{j+1}\cdots a_{2m}}\epsilon^{a_1\cdots a_{2m}}, \tag{15}$$

where $\epsilon^{1\cdots 2m} = 1$. We note here a similarity to differential forms. We can associate to ψ a "Hodge dual" defined by

$$(*\psi)^{a_1\cdots a_j} = \frac{1}{[(2m-j)!]}\epsilon^{a_1\cdots a_{2m}}\psi_{a_{j+1}\cdots a_{2m}}. \tag{16}$$

This is exactly the same definition as for differential forms,[9] and as there it holds

$$**\overset{(j)}{\psi} = (-1)^{j(2m-j)}\overset{(j)}{\psi} = (-1)^j\overset{(j)}{\psi}. \tag{17}$$

With this notation the Hermitian form can be written as

$$(\psi,\phi) = \sum_{j=0}^{2m} (\overset{(j)}{\psi},\overset{(j)}{\phi}) = \sum_{j=0}^{2m} \int dp\, d\pi\, \overset{(j)}{\psi}{}^*(*\overset{(j)}{\phi}), \tag{18}$$

i.e., is similar to a sum over scalar products of j-forms. We denote the space of coefficient functions $\psi_{a_1\cdots a_j}(p,\pi)$ by Ω^j, and we may write $\psi \in \Omega^*$, where $\Omega^* = \oplus_j \Omega^j$. This is the same notation as used in the deRham theory.

The expression (15) would be well defined if all the $\psi_{a_1\cdots a_j}$ were square integrable functions. This is not general enough, however. For instance, in the case of the free relativistic particle we should recover the well-known one-particle quantum theory. The states of this theory are $\sim\delta(p^2+m^2)$. Thus we have to find a way to give (15) a proper meaning even when generalized functions are involved. In practice these generalized functions usually are δ functions.

The same problem appears in the Dirac procedure. The physical state condition $(p^2+m^2)\psi = 0$ has only the zero solution if $\psi \in L^2$. Thus one has to modify the original Dirac approach to also allow for generalized functions. Similar problems occur also in quantum mechanics whenever operators with continuous spectrum are considered. As seen below, the BRST formalism is able to master this in an elegant way.

The natural extension of the concept of a Hilbert space when generalized functions are involved is a rigged Hilbert space,[10] which we now proceed to define to the extent we need it.

Definition 1: Let $\mathbb{H}_0$ be a space of fundamental functions (test functions) over some manifold M. In $\mathbb{H}_0$ let there be a positive definite, nondegenerate, continuous, Hermitian inner product $(\cdot,\cdot)$. Let $\mathbb{H}_{1/2}$ be the completion of $\mathbb{H}_0$ w.r.t. this inner product. Let $\mathbb{H}_1$ be the adjoint space to $\mathbb{H}_0$. There exists a continuous linear operator T which maps $\mathbb{H}_0$ 1–1 onto an everywhere dense subset in $\mathbb{H}_{1/2}$ and its (antilinear) adjoint T' maps $\mathbb{H}_{1/2}$ 1–1 onto an everywhere dense subset in $\mathbb{H}_1$. The triple $\mathbb{H}_0 \subset \mathbb{H}_{1/2} \subset \mathbb{H}_1$ is called a *rigged Hilbert space.*

Remark: This $\mathbb{H}_0$ has to be a countably Hilbert nuclear space. This is the case for the usual test function spaces.

We take as M the space spanned by p,π; as $\mathbb{H}_0$ the C^∞ functions from M to $\mathbb{C}$ that decrease together with all their derivatives faster than any power, i.e., $\mathbb{H}_0 = S(M)$. The inner product is $(\psi,\phi) = \int \lambda_M \psi^*\phi$, where λ_M is the Lebesgue measure on M, $\mathbb{H}_{1/2}$ is then $L^2(M)$ and $\mathbb{H}_1 = S'(M)$. We use the notation S, L^2, S' even though the considered functions are complex valued and usually the S, L^2, S' functions in the literature are real valued. (We choose $\mathbb{H}_0 = S$ in order to have it invariant under Fourier transformations.) So the rigged Hilbert space we will work in is

$$\begin{array}{ccccc} S(M) & \subset & L^2(M) & \subset & S'(M) \\ \| & & \| & & \| \\ \mathbb{H}_0 & \subset & \mathbb{H}_{1/2} & \subset & \mathbb{H}_1 \end{array}. \tag{19}$$

To describe in which part of the rigged Hilbert space the coefficient functions lie, we introduce the following notation.

Definition 2: The *character* of a space of coefficient functions Ω^j is defined by

$$\chi(j) = \varkappa\,(\varkappa = 0,\tfrac{1}{2},1) \ \text{ iff } \ \Omega^j = (\mathbb{H}_\varkappa)^{\binom{2m}{j}}.$$

In (15) in every summand an element of Ω^j is multiplied by an element of Ω^{2m-j} and then integrated over. If $\chi(j)$, $\chi(2m-j) \leqslant \frac{1}{2}$, then the $\int$ sign means Lebesgue integration. If $\chi(j) = 1$, $\chi = (2m-j) = 0$ (or vice versa), it means that the linear functional $\psi^* \in S'$ is evaluated on the test function $\phi \in S$ (or vice versa). Although this is not an integral in the usual Lebesgue sense, it is nevertheless common to use the same notation for this operation. In all other cases the expression (15) is ill defined. Summarizing we have the following proposition.

Proposition 1: (a) The Hermitian form (15) is well defined iff $\chi(j) + \chi(2m-j) \leqslant 1$.

(b) For (15) to be well defined it is necessary that $\chi(m) \leqslant \frac{1}{2}$.

For rank 0 systems the action of $\widehat{\Omega} = \widehat{\eta}^a \widehat{G}_a$ on ψ is described by the operation ρ: $\Omega^j \rightarrow \Omega^{j+1}$ defined by

$$(\rho\psi)_{a_1 \cdots a_{j+1}} = G_{[a_{j+1}} \psi_{a_1 \cdots a_j]}. \tag{20}$$

The square bracket means antisymmetrization, and it is obvious that $\rho^2 = 0$. So we have that

$$\begin{array}{ccccccccc} 0 \rightarrow & \Omega^0 \xrightarrow{\rho^1} & \Omega^1 \xrightarrow{\rho^2} \cdots \xrightarrow{\rho^j} & \Omega^j \xrightarrow{\rho^{j+1}} \cdots \xrightarrow{\rho^{2m}} & \Omega^{2m} \rightarrow 0, \\ & | & | & | & | \\ & H^0 & H^1 & H^j & H^{2m} \end{array}, \tag{21}$$

where

$$\Omega^j = \mathbb{H}_{\chi(j)}$$

is a differential complex, the BRST complex, and our task is to calculate its cohomology:

$$H^*_{\mathrm{BRST}} = \bigoplus_j H^j_{\mathrm{BRST}} = \bigoplus_j \ker \rho^{j+1} / \mathrm{Im}\, \rho^j. \tag{22}$$

Let us start with $H^0 = \ker \rho^1 = \{\psi_0 | G_a \psi_0 = 0,\ \forall a\}$. In the representation we have chosen for the quantum operators, these are just algebraic equations for ψ_0. If $\chi(0) = 0, \frac{1}{2}$, the only solution is $\psi_0 = 0$, If $\chi(0) = 1$ then every distribution of the form $\Pi_a\ \delta(G_a) d$, where d is any distribution on the constraint surface $\mathscr{C}$, $\mathscr{C} = \bigcap_a (G_a = 0)$, is a solution, the space of solutions is thus isomorphic to $S'(\mathscr{C})$. This space $S'(\mathscr{C})$ should not be confused with the subspace of $S'(M)$ of all distributions which are concentrated on $\mathscr{C}$. The latter contains elements like δ functions concentrated on $\mathscr{C}$ derived in directions nontangential to $\mathscr{C}$; these are not annihilated by all G_a. Thus we have the following proposition.

Proposition 2:

$$H^0 = \begin{cases} 0, & \text{if } \chi(0) = 0, \frac{1}{2}, \\ S'(\mathscr{C}), & \text{if } \chi(0) = 1. \end{cases}$$

For the general case we have to look at $\cdots \rightarrow \Omega^{j-1} \xrightarrow{\rho^j} \Omega^j \xrightarrow{\rho^{j+1}} \Omega^{j+1} \rightarrow \cdots$. We split it up in several propositions.

Proposition 3: If $\chi(j) \geqslant \frac{1}{2}$ and $\chi(j-1) \geqslant \chi(j)$, then $H^j = 0$, $\forall j$.

Proof: Since $\Omega^j / \mathrm{Im}\, \rho^j \supseteq \ker \rho^{j+1} / \mathrm{Im}\, \rho^j$, it is sufficient to show that $\Omega^j / \mathrm{Im}\, \rho^j = 0$ under the above assumptions. We have to consider three cases.

(a) $\chi(j-1) = \frac{1}{2} = \chi(j)$: The image of $\mathbb{H}_{1/2}$ under ρ are all L^2 functions that vanish on $\mathscr{C}$. But since $\mathscr{C}$ is of measure zero this is ismorphic to L^2, since in L^2 any two functions differing on a set of measure zero are identified. Thus $\Omega^j / \mathrm{Im}\, \rho^j \cong \mathbb{H}_{1/2} / \mathbb{H}_{1/2} = 0$.

(b) $\chi(j-1) = 1, \chi(j) = \frac{1}{2}$: Note that in this case $\mathrm{Dom}\, \rho^j \neq \Omega^{j-1}$. But surely $\mathrm{Dom}\, \rho^j \supset \mathbb{H}_{1/2} \cup \ker \rho^j$, and we can proceed as in (a).

(c) $\chi(j-1) = 1 = \chi(j)$. In this case $\mathrm{Im}\, \rho^j = \mathbb{H}_1$, e.g., $\delta(G)$ can be removed with $\delta'(G)$, etc. So $\Omega^j / \mathrm{Im}\, \rho = \mathbb{H}_1 / \mathbb{H}_1 = 0$.

Proposition 4: If $\chi(j) = 0$, then (a) $H^j = 0 (j < 2m)$ and (b) $H^{2m} = S(\mathscr{C})$.

Proof: Since $\mathbb{H}_0 \subset \mathbb{H}_{1/2} \subset \mathbb{H}_1$, it is sufficient to look at the case when $\chi(j-1) = 0$. We thus have to calculate H^* in the case where all coefficient functions are test functions. We exploit the similarity to differential forms to connect our problem to a theorem proved in Ref. 5. In Ref. 5 an operation δ_2 was defined by

$$(\delta_2 \psi)^{a_1 \cdots a_{j-1}} = \psi^{a_1 \cdots a_j} G_{a_j}.$$

It maps Ω^j to Ω^{j-1}. One can show that

$$\rho = -*\delta_2*, \tag{23}$$

i.e., δ_2 is the adjoint of ρ which can be considered as an exterior derivative. Theorems 3.1 and 3.2 in Ref. 5 state that the cohomology of the differential complex

$$0 \rightarrow \Omega^{2m} \xrightarrow{\delta_2} \cdots \rightarrow \Omega^0 \rightarrow 0$$

is trivial if the functions making up the Ω^j's are sufficiently regular in a neighborhood of $\mathscr{C}$. This is satisfied for the test functions $S(M)$. Using these theorems we can calculate the cohomology of the BRST complex (Ω^*, ρ). $\psi \in \ker \rho$ iff $\rho\psi = 0$ iff, because of (23), $-*\delta_2*\psi = 0$. This holds iff $\delta_2*\psi = 0$. Now applying the theorems of Ref. 5, $*\psi = \delta_2 \widetilde{K}$ for some $\widetilde{K}$ must hold. Taking the Hodge dual of this equation we get, using (17),

$$\begin{aligned} (-1)^j \overset{(j)}{\psi} &= *\delta_2 \overset{(j-1)}{\widetilde{K}} = *\delta_2 (-1)^{j-1} ** \overset{(j-1)}{\widetilde{K}} \\ &= -*\delta_2 * (-1)^j * \overset{(j-1)}{\widetilde{K}}. \end{aligned}$$

Dividing by $(-1)^j$ and using (23), we get $\psi = \rho * \widetilde{K} = \rho K$, i.e., $\psi \in \mathrm{Im}\, \rho$. Thus $\ker \rho = \mathrm{Im}\, \rho$ whenever $\overset{(2m)}{\psi} = 0$ so that $\delta_2 * \psi$ is defined. This proves (a).

(b) For $\psi \in \mathrm{Im}\, \rho^{2m}$ it holds that $\psi|_{\mathscr{C}} = 0$. If $\chi(2m) = 0$, functions such as $\delta'(G)$, etc., are not in the domain of ρ^{2m}, since their image under ρ would be a δ function and thus lie outside $\mathbb{H}_0$. We have $\mathrm{Im}\, \rho^{2m} = \{\psi \in S(M) | \psi|_{\mathscr{C}} = 0\}$. So two functions in Ω^{2m} are equivalent iff they coincide on $\mathscr{C}$. Thus $H^{2m} = \Omega^{2m} / \sim \cong S(\mathscr{C})$.

Remark: The similarity to deRham theory exploited in the proof is only formal, e.g., Poincaré duality does not hold in H^*_{BRST} in general as is seen from the above. Proposition 4 states that when only test functions are allowed then $H^* = H^{2m} = S(\mathscr{C})$. In the literature usually it is stated that $H^* = H^0$. This would be the case if $\chi(j) = 1$, $\forall j$, as seen from Propositions 2 and 3, but in this case the form (15) is ill defined (Proposition 1).

Here we also note that H^* depends on $\chi(j)$. This is an observation similar to the one that H^* depends on the boundary conditions imposed on the coefficient functions made in Ref. 6. Below we will see that for all χ such that H^* contains a subspace on which the induced form is a positive-definite scalar product, we arrive at a unique H^*.

In the remaining cases, $\chi(j-1)<\chi(j)$. We do not need the explicit form of H^j in these cases, since if we require H^* to be such that the induced form on it is nondegenerate and that it contains a subspace on which that form is even positive definite, then they can have only a trivial intersection with that subspace. We must require to have such a subspace, since this will be precisely the physical pre-Hilbert space.

Lemma 1: If $\chi(j-1)<\chi(j)$, then $H^j\neq 0$.

Proof: $\ker\rho^{j+1}\supset\rho(\mathbb{H}_{\chi(j)})$ since $\rho(\mathbb{H}_{\chi(j)})\subset\Omega^j$ and $\rho^2=0$. So $H^j\supset\rho(\mathbb{H}_{\chi(j)})/\mathrm{Im}/\rho^j\neq 0$.

The next proposition shows that $\chi(j)$ can never raise to the value 1.

Proposition 5: For the induced form on H^* to be nondegenerate, it is necessary that whenever $\chi(j)=1$ then it must hold (a) $\chi(j')=1$, $\forall j'\leqslant j$, and (b) $j<m$.

Proof: If $\chi(j)=1$ then by Proposition 1, $\chi(2m-j)=0$. Assume $\chi(j-1)<1$. Then by Lemma 1, $H^j\neq 0$. In order that no degenerate directions occur, H^{2m-j} also has to be nonzero, as is seen from (15). This, however, contradicts Proposition 4. Iterating this gives (a). Then (b) follows from (a) and Proposition 1(b).

This excludes $\chi(j-1)=0,\frac{1}{2}$ and $\chi(j)=1$. The last case to be discussed is $\chi(j-1)=0$ and $\chi(j)=\frac{1}{2}$. To do this we need a result from the theory of linear topological spaces.

Lemma 2: If X is a topological VS, Y a linear subspace, then X/Y is $T_2\Leftrightarrow Y$ closed in X.

This is proved in Ref. 11 (Proposition 1, p. 77). With this we now show the following proposition.

Proposition 6: If $\chi(j-1)=0$, $\chi(j)=\frac{1}{2}$, then H^j does not contain a nontrivial, normable subspace.

Proof: By Proposition 4, $\overline{\ker\rho|_{\mathbb{H}_0}}=\mathrm{Im}\,\rho$ and since $\mathbb{H}_0$ is dense in $\mathbb{H}_{1/2}$, we also have $\overline{\mathrm{Im}\,\rho}=\ker\rho$. We call $\ker\rho=X$, $\mathrm{Im}\,\rho=Y$. We have $Y\neq X$ and $\overline{Y}=X$. We have to look at the linear subspaces $A\subset X/Y$. This A is of the form U/Y with $U=(A\times Y)\subset X$. It is a topological VS (with the subspace topology), and $\overline{Y}=U$ where the closure of Y is taken in U. So the assumptions of Lemma 2 are satisfied, and since Y is not closed in U, A is non-Hausdorff and therefore not normable whenever $A\neq 0$.

Actually we have proved more, namely, that every nonzero subspace of H^j and thus H^j itself is non-Hausdorff. This means that this space can never occur in the true physical subspace of H^*. However, we cannot exclude it completely from H^*. To this end we have to make an additional assumption. Two possiblities are given by Theorems 1 and 1'.

Theorem 1: If H^*, nontrivial, is such that the induced Hermitian form is nondegenerate and there exists a subspace having nontrivial intersection with every $H^j\neq 0$ such that on that subspace the induced form is positive definite, then

(i) χ is monotonically decreasing,

(ii) $\chi(0)=1,\quad \chi(2m)=0$,

(iii) $H^*=H^0\oplus H^{2m}=S'(\mathscr{C})\oplus S(\mathscr{C})$.

Proof: (i) According to Proposition 5, χ can never raise to 1. According to Proposition 6, χ cannot raise to $\frac{1}{2}$, because if it did, there would occur nonzero (Lemma 1) H^j which have trivial intersection with the true physical subspace. Thus χ is monotonic.

(ii) If $\chi(0)\neq 1$, then $\chi(j)\leqslant\frac{1}{2}$, $\forall j$, by (i). But then $H^*=H^{2m}$ by Propositions 3 and 4. So either $H^*=0$, if $H^{2m}=0$, or the induced form is degenerate, if $H^{2m}\neq 0$. Thus $\chi(0)=1$, $\chi(2m)=0$, follows by Proposition 1.

(iii) By (i) and Propositions 3 and 4, $H^j=0$ if $j\neq 0,2m$, so $H^*=H^0\oplus H^{2m}$. By (ii) and Propositions 2 and 4(b) the claim follows.

The additional assumption that only H^j's should occur that contribute to the true physical subspace is reasonable, but as we have seen above, it is not necessary for consistency. An alternative additional assumption, with which one also can prove the claims (i)–(iii) of Theorem 1, is that the inequality in Proposition 1(a) is always satisfied with the equality sign. This is a sensible restriction on χ. One should always choose function spaces as big as allowed by Proposition 1(a) and not artificially restrict oneself to smaller ones. In this case one can prove (i)–(iii) without using Lemma 2.

Theorem 1': If H^* nontrivial is such that the induced form on it is nondegenerate and χ satisfies

$$\chi(j)+\chi(2m-j)=1, \tag{24}$$

then (i)–(iii) of Theorem 1 hold.

Proof: (ii) If $\chi(0)\neq 1$ then $\chi(j)\neq 1$, $\forall j$, by Proposition 5. By (24) it must then hold $\chi(j)=\frac{1}{2}$, which gives $H^*=0$, contradicting the assumptions.

(i) $\chi(j)\neq 0$ for $j\leqslant m$ because otherwise $\chi(2m-j)$ had to be 1 [by (24)], which contradicts Proposition 5. Once χ drops to $\frac{1}{2}$, which occurs at m at the latest, it can never raise to 1 again by Proposition 5. Therefore, if χ is monotonic for $j\leqslant m$, then the fact that it is also monotonic for $j\geqslant m$ follows from (24).

(iii) As in Theorem 1.

The results are summarized in Table I.

Theorem 1' establishes that for reasonable choices of χ, H^* does not depend on χ. For definiteness, in the sequel, we will always assume

$$\chi(0)=1\quad \chi(2m)=0,\quad \chi(j)=\tfrac{1}{2},\quad j\neq 0,2m. \tag{25}$$

We see that (at least) a "doubling" [actually it is more than a doubling since $S'(\mathscr{C})\supset S(\mathscr{C})$] of the degrees of freedom compared to more traditional quantization schemes is a general feature of bosonic systems even when the topology is trivial. Such a doubling is well known in the literature.[4–6,8] It is removed by scalar product arguments, but usually the discussion is rather heuristic. We will systematically treat this problem in Sec. IV.

Since ρ is Hermitian w.r.t. (15) and nilpotent, a Hermitian form is induced on H^* by (15). It is given by

$$(\psi,\phi)=i^k\int_M\lambda_M(\psi_0^*\phi_{1\cdots 2m}+(-1)^m\psi_{2m\cdots 1}^*\phi_0, \tag{26}$$

with $\psi_0,\phi_0\in S'(\mathscr{C})$ and $\psi_{1\cdots 2m},\phi_{1\cdots 2m}\in S(\mathscr{C})$. We introduce the abbreviations

TABLE I. The BRST cohomology groups H^j as a function of $\chi(j-1)$ and $\chi(j)$ $(j\neq 0)$.

$\chi(j-1)$	$\chi(j)$	$H^j(j\neq 2m)$	H^{2m}	Remarks
0	0	0	$S(\mathscr{C})$	
$\frac{1}{2}$	0	0	$S(\mathscr{C})$	by Proposition 4
1	0	0	$S(\mathscr{C})$	
0	$\frac{1}{2}$	non-T_2	non-T_2	by Proposition 6. Excluded under additional assumptions. Theorems 1, 1′
$\frac{1}{2}$	$\frac{1}{2}$	0	0	
1	$\frac{1}{2}$	0	0	by Proposition 3
0	1	$\cdots$	$\cdots$	excluded by Proposition 5
$\frac{1}{2}$	1	$\cdots$	$\cdots$	
1	1	0	0	by Proposition 3

$$\psi_{1\cdots 2m}=\psi_{2m},\quad \psi^*_{2m\cdots 1}=\psi^*_{2m}. \tag{27}$$

Since $\psi_0,\phi_0\in S'(\mathscr{C})$, the "integral" in (26) is effectively only over $\mathscr{C}$ and not over M. We may thus write

$$(\psi,\phi)|_{H^*}=i^k\int_{\mathscr{C}} d\mu(\mathscr{C})(\psi_0^*\phi_{2m}+(-1)^m\psi^*_{2m}\phi_0). \tag{28}$$

The positive measure $d\mu(\mathscr{C})$ on $\mathscr{C}$ is defined by

$$\lambda_M=dG_1\wedge\cdots\wedge dG_{2m}\,d\mu(\mathscr{C}). \tag{29}$$

It is an $(n-m)$-form, the volume form induced on the $(n-m)$-dimensional manifold $\mathscr{C}$. The solution of this equation is not unique, however, if $d\mu$ and $d\tilde{\mu}$ are two solutions, then

$$\int\varphi\,d\mu=\int\varphi\,d\tilde{\mu},\quad\forall\varphi\in\mathrm{H}_0.$$

$d\mu$ can be given an expression in coordinates:

$$d\mu=dx_{2m+1}\cdots dx_n\det{}^{-1}\begin{pmatrix}G_1\cdots G_{2m}\\ x^1\cdots x^{2m}\end{pmatrix} \tag{30}$$

where $(G_1\cdots G_{2m}/x^1\cdots x^{2m})$ is a maximal rank minor of $(\partial G_i/\partial x^k)$ and $x^k=\pi_\alpha,\ p_\mu$. We will never need the explicit form of $d\mu$, so this may suffice. For details the reader may consult Ref. 12, Chap. III, Sec. 1, No. 9. As an example we do the free relativistic particle.

Example: For $G_1=\pi,\ G_2=p^2+m^2,\ x^k=\pi,\ p_\mu,\ \lambda_M=d^4p\,d\pi$,

$$\det\begin{pmatrix}G_1 & G_2\\ x^1 & x^2\end{pmatrix}=\det\begin{pmatrix}\dfrac{\partial G_1}{\partial\pi} & \dfrac{\partial G_1}{\partial p^0}\\[2mm] \dfrac{\partial G_2}{\partial\pi} & \dfrac{\partial G_2}{\partial p^0}\end{pmatrix}=\det\begin{pmatrix}1 & 0\\ 0 & 2p^0\end{pmatrix}=2p^0,$$

therefore

$$d\mu=\frac{d^3p}{2p^0}.$$

This depends, of course, on what coordinates we have chosen to parametrize the mass shell. It is no wonder that the relativistic measure emerges since both λ_M and the contraints are relativistically invariant in this example.

The induced Hermitian form on H^* is still not positive definite:

$$\|\psi\|^2=i^k\int_{\mathscr{C}} d\mu(\psi_0^*\psi_{2m}+(-1)^m\psi^*_{2m}\psi_0)$$
$$=2\int_{\mathscr{C}} d\mu|\psi_0|\,|\psi_{2m}|{\textstyle\frac{\sin}{\cos}}(\arg\psi_0-\arg\psi_{2m})\ \ {\textstyle{m\ \text{odd}\atop m\ \text{even}}}. \tag{31}$$

From this expression we also note that $\|\psi\|^2\in\mathbf{R}$ as it should be.

Let us summarize what we achieved thus far. We have described a formalism that naturally allows for distributions to occur. In this respect it is superior to the original Dirac formalism where the states are represented by square integrable functions only. Often it happens there that the solution of the physical state condition $\widehat{G}_a|\psi\rangle=0$ only allows the zero state (e.g., this is the case for the particle). We proved (under some reasonable assumptions) that H^* is unique and does not depend on the character function χ (Theorems 1 and 1′). However, we have two problems: (1) compared to more traditional quantization schemes we have a doubling of the degrees of freedom; and (2) the induced form on H^* is not positive definite. But, as we will see below, these problems cancel each other so to speak. We were in trouble if we had either only the doubling or only the nonpositive definiteness of the induced form, because if so, the BRST quantization could not be equivalent to other quantization methods for our bosonic systems. In order to get a positive-definite scalar product, we have to choose an appropriate subspace of H^*. By this "truncation" or "selection rule" (both expressions are used in the literature) we also remove the doubling. This is the subject of the next section.

IV. THE SUBSPACE(S) WITH POSITIVE-DEFINITE INNER PRODUCT

We have to choose a subspace $\Pi\subset H^*$ on which the Hermitian form (28) is positive definite. This subspace has to be linear since we are dealing with quantum mechanics, i.e., $\psi,\phi,\in\Pi\Rightarrow\alpha\psi+\beta\phi\in\Pi$. The task is to determine all possible choices for Π. We start by considering a mapping P: $S(\mathscr{C})\Rightarrow 2^{S'(\mathscr{C})}$ defined by

$$P(\psi_{2m}) = \{\psi_0 \in S'(\mathscr{C}) \mid \|\psi\|^2 > 0;$$
$$\psi_0 \in P(\psi_{2m}) \wedge \phi_0 \in P(\phi_{2m})$$
$$\Rightarrow \alpha\psi_0 + \beta\phi_0 \in P(\alpha\psi_{2m} + \beta\phi_{2m}),\ \forall\phi\}. \tag{32}$$

Note that $P(0) = \varnothing$. The linearity requirement turns out to be very restrictive. It implies that there can be only one ψ_0 with every ψ_{2m}. Indeed, if there were two ψ_0 and ψ_0' say, then the difference $(\psi_0, \psi_{2m}) - (\psi_0', \psi_{2m}) = (\psi_0 - \psi_0', 0)$ would be a zero norm state which is not the zero state. So, in fact, $P: S(\mathscr{C}) \to S'(\mathscr{C})$, and since it has to be linear it is of the form

$$P(\psi_{2m}) = f\psi_{2m},\quad f \in S'(\mathscr{C}) \text{ fixed.} \tag{33}$$

But not every $f \in S'(\mathscr{C})$ is allowed in this formula. We have

Proposition 7:

$$\|\psi\|^2 > 0 \Leftrightarrow \begin{cases} \operatorname{Im} f > 0 & (m \text{ odd}), \\ \operatorname{Re} f > 0 & (m \text{ even}). \end{cases}$$

Proof: We do it for m odd. The other case is similar. From the above we have $\psi_0 = f\psi_{2m}$. We put this into the rhs of (31). This turns out to be positive for all $\psi_{2m} \neq 0$ iff $\sin(\arg\psi_0 - \arg\psi_{2m}) = \sin(\arg f)$ is positive on $\mathscr{C}$, i.e., $\operatorname{Im} f > 0$ on all of $\mathscr{C}$.

There are many possible chocies of linear subspaces Π on which the inner product is positive definite. They are indexed by the the set of generalized functions from $\mathscr{C}$ to the upper (m odd) [or right (m even)] complex half-plane. However, $\Pi_f \cong S(\mathscr{C})$, $\forall f$. Actually, we have the following proposition.

Proposition 8: $(\Pi_f;\ (\cdots)|_{\Pi_f})$ and $(\Pi_g;\ (\cdots)|_{\Pi_g})$ are isomorphic scalar product spaces. By this isomorphism F, an isomorphism F_* between the operator algebras over Π_f and Π_g is induced. Thus $\hat{A}$ is Hermitian if and only if $\widehat{F_* A}$ is.

Proof: Let m be even. Then m odd is similar. It is sufficient to prove that all Π_f's are isomorphic to Π_1. The scalar products in Π_1 and Π_f are (28):

$$(\psi,\phi)|_{\Pi_1} = 2\int d\mu\ \psi_{2m}^* \phi_{2m},$$

$$(\psi,\phi)|_{\Pi_f} = 2\int d\mu\ \operatorname{Re} f\, \psi_{2m}^* \phi_{2m}.$$

The isomorphism F between Π_1 and Π_f is given by

$$F(\psi_{2m},\psi_{2m}) = ((f/\sqrt{\operatorname{Re} f})\psi_{2m};\ (1/\sqrt{\operatorname{Re} f})\psi_{2m}).$$

This is well defined and invertible since $\operatorname{Re} f > 0$ according to Proposition 7. Clearly $F: \Pi_1 \to \Pi_f$ is 1–1 and linear. Also F is isometric:

$$(F\psi,F\phi)|_{\Pi_f} = \int d\mu\ 2\operatorname{Re} f \frac{1}{\sqrt{\operatorname{Re} f}}\psi_{2m}^* \frac{1}{\sqrt{\operatorname{Re} f}}\phi_{2m}$$
$$= \int d\mu\ 2\psi_{2m}^*\phi_{2m} = (\psi,\phi)|_{\Pi_1}.$$

The induced map between the operator algebras over Π_1 and Π_f respectively, is defined by $(\widehat{F_* A})\ (F\psi) = F(\hat{A}\psi)$, thus F_* is given by

$$F_*\hat{A} = F\hat{A}F^{-1}.$$

The last claim follows trivially from the definition of F_* and the fact that F is an isometry.

The choice of f is further restricted by the (minimal) requirement that Π_f must be left invariant by the algebra of BRST observables, especially, it must be left invariant by the Hamiltonian. By (12), BRST observables act on H^*, and they do not mix the H^j (since they commute with the ghost number operator). So $\hat{A}$ must be of the form $\hat{A} = \hat{A}_0 \oplus \hat{A}_{2m}$, where $\hat{A}_0$ acts on H^0 and $\hat{A}_{2m}$ on H^{2m}. In the next proposition we show that $\hat{A}_0|_{S(\mathscr{C})} = \hat{A}_{2m}$. We use a matrix notation for $\hat{A}$. The $2^{2m} \times 2^{2m}$ matrix $\hat{\mathrm{A}}$ acts on the space $\Omega^* = \oplus_j \Omega^j$, which is considered as a 2^{2m}-dimensional space. The dimension of Ω^j is $\binom{2m}{j}$ in this notation. We then have the following proposition.

Proposition 9: In every equivalence class $\exists \hat{\mathrm{A}}$ of the form $\hat{A}\cdot\mathbf{1}$ with $\hat{A}$ commuting with $\hat{G}_a$.

Proof: From (12),2, it follows $\hat{\mathrm{A}} = \oplus_j \hat{\mathrm{A}}_j$, where $\hat{\mathrm{A}}_j$ acts on Ω^j, i.e., $\hat{\mathrm{A}}$ transforms every Ω^j into itself. The proof proceeds by induction on j. For $j = 0$ we have $\hat{\mathrm{A}}_0 = \hat{A}\cdot\mathbf{1}$ since Ω^0 is one dimensional and $\hat{A}$ is first class as follows from (12) (3) when considered at $\rho = 0$. [Compare, e.g., Ref. 5 formula (6.1.6).] In every equivalence class there is an $\hat{\mathrm{A}}$ such that even $\hat{A}$ commutes with the $\hat{G}_a$ (see Ref. 5, Sec. 7.3). Then from (12) (3) follows

$$\hat{\mathrm{A}}_{j+1}\rho = \rho\hat{\mathrm{A}}_j = \rho\hat{A} = \hat{A}\rho,$$

and thus $\hat{\mathrm{A}}_{j+1} = \hat{A}\cdot\mathbf{1}$ which completes the proof.

So we must have f such that $f\hat{A}\psi_{2m} = \hat{A} f\psi_{2m}$ for all observables $\hat{A}$. This is the case if one chooses f constant. Such a restriction does not contradict Proposition 8. We illustrate this by an analogous finite-dimensional example.

Example: Let $H = \mathbb{R}^2$ with coordinates x^0, x^1. As symmetric bilinear form, we choose $(x;y) = x^0y^1 + x^1y^0$. It is nondegenerate, x^0 and x^1 axes are null directions. Restricted to the subspace $\Pi_\alpha = \{x \in \mathbb{R}^2 | x^0 = \alpha x^1\}$, the norm induced by (;) is $\|x\|^2 = 2\alpha(x^1)^2$, i.e., $(\ ;\ )|_{\Pi_\alpha}$ is positive (negative) definite iff $\alpha > 0$ ($\alpha < 0$). We have

$$(x;y)|_{\Pi_\alpha} \doteq (x;y)_\alpha = 2\alpha x^1 y^1.$$

The spaces $(\Pi_\alpha,(\cdot,\cdot)_\alpha)$ and $(\Pi_\beta,(\cdot,\cdot)_\beta)$ $(\alpha,\beta > 0)$ are isomorphic. The isomorphism is given by

$$F(\alpha x^1, x^1) = (\sqrt{\alpha\beta}\,x^1; \sqrt{\alpha\beta^{-1}}\,x^1),$$

and we have indeed

$$(Fx,Fy)_\beta = 2\beta(Fx)^1(Fy)^1 = 2\alpha x^1 y^1 = (x,y)_\alpha.$$

Now consider the algebra Γ of linear transformations on $\mathbb{R}^2$ which consists of maps that scale the 45° direction by a factor γ and/or do an overall scaling by σ. They are given by the two-parameter family of matrices

$$\frac{\sigma}{2}\begin{pmatrix} \gamma+1 & \gamma-1 \\ \gamma-1 & \gamma+1 \end{pmatrix},$$

which are self-adjoint w.r.t. $(\cdot,\cdot)$, as is easily checked. When we demand that Π_α be mapped into itself by the algebra Γ, then we must have $\alpha = \pm 1$ since $\Pi_{\pm 1}$ are the only subspaces which are mapped into themselves by all elements of Γ.

Note also that when f commutes with $\hat{A}$, then the mapping F_* introduced in Proposition 8 is the identity mapping. A convenient choice for f is

$$f = i^k/2, \text{ i.e., } \psi_0 = (i^k/2)\psi_{2m}, \tag{34}$$

where k again equals m (mod 2), since this leads directly to the usual Klein–Gordon product in the case of the free relativistic particle. Other choices of $f=$ const are unitarily equivalent to this by Proposition 8. In summary we have the following.

Theorem 2: There exist subspaces $\Pi_f \subset H^*$ isomorphic to the physical space of the more traditional quantization methods (in the cases where the latter work), which are left invariant by the algebra of BRST observables and which carry a positive-definite inner product induced from the Hermitian form on the "big" space of states. The quantum theories corresponding to the different choices of these Π_f's are all equivalent. The completion of Π_f w.r.t. this inner product yields the physical Hilbert space.

ACKNOWLEDGMENTS

The author thanks J. Fisch and M. Henneaux for stimulating discussions and J. Rätz for pointing out Lemma 2 to him. He especially thanks P. Hajicek for valuable discussions and suggestions during all stages of this work.

This work was supported by Schweizerischer Nationalfonds.

[1]M. Kato and K. Ogawa, Nucl. Phys. B **212**, 433 (1983); S. Hwang, Phys. Rev. D **28**, 2614 (1983).
[2]J. Fisch, Quantification BRS des théories de cordes; Mémoire de Licence, Bruxelles (academic year 85–86).
[3]E. S. Fradikin and G. A. Vilkovisky, CERN Report TH:2332, 1977; I. A. Batalin and G. A. Vilkovisky, Phys. Lett. B **69**, 309 (1977); E. S. Fradkin and G. A. Vilkovisky, Phys. Lett. B **55**, 224 (1975).
[4]M. Henneaux, Phys. Lett. B **183**, 59 (1987).
[5]M. Henneaux, Phys. Rep. **126**, 1 (1985).
[6]J. Fisch and M. Henneaux, in *Proceedings of the Florence Workshop on Constrained Dynamics*, 1986 (to appear).
[7]C. Galvao and C. Teitelboim, J. Math. Phys. **21**, 1863 (1980).
[8]M. Henneaux, *Proceedings of the Conference on Quantum Mechanics of Fundamental Systems*, edited by C. Teitelboim (Plenum, New York, to be published).
[9]T. Eguchi, P. Gilkey, and A. Hanson, Phys. Rep. **66**, 213 (1980).
[10]I. Gel'fand and N. Vilenkin, *Generalized Functions* (Academic, London, 1964), Vol. IV.
[11]A. Robertson and W. Robertson, *Topological Vector Spaces* (Cambridge U.P., Cambridge, MA, 1973).
[12]I. Gel'fand and G. Shilov, *Generalized Functions* (Academic, London, 1964), Vol. I.

Yang–Mills equations and solvable groups. II

M. Cahen and S. Gutt[a)]
Mathematics Department, Universite Libre de Bruxelles, Bruxelles, Belgium
C. Kozameh[b)] and E. T. Newman
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 9 December 1987; accepted for publication 16 December 1987)

It is shown that the necessary and sufficient condition for the Yang–Mills equations (associated with an arbitrary group G) to be derivable from a Lagrangian (which is polynonial in the derivatives of the connection) is that the Lie algebra $\mathscr{g}$ of G possesses an invariant nondegenerate quadratic form γ. It is well known that for semisimple groups such a γ exists, namely the Killing form. What is not so well known is that such a γ exists for many other groups and in particular for many solvable and nilpotent groups. Several examples in this class are discussed.

I. INTRODUCTION

The framework of classical Yang–Mills theory is a principal G bundle over a Riemannian (or pseudo-Riemannian) manifold (M,g); in most of the recent developments G is a compact Lie group. It is, however, easily observed that, when G is a *solvable* Lie group, the nonlinear classical Yang–Mills field equations possess an interesting linear structure. They can be divided into ordered sets of equations with the following properties: the first set is a collection of homogeneous linear equations (a set of independent Maxwell equations); each of the following sets is also linear but not homogeneous, the coefficients being determined in terms of the previous sets. This linear structure suggests that in such a situation one could exhibit or study a nontrivial but solvable model for quantization.

Unfortunately the "generic" solvable Yang–Mills theory has field equations that cannot be derived from a variational principle[1]; hence no standard canonical formulation of a quantum theory is available. A simple example of such a situation is provided by the group of affine transformations of the real line, the two-dimensional Borel group (see below).

Nevertheless there are particular solvable and even nilpotent, non-Abelian Lie groups such that the corresponding Yang–Mills equations are indeed the Euler–Lagrange equations of a variational principle.

II. EXISTENCE OF A LAGRANGIAN

Let $\pi: P\to M$ be a principal bundle with structure group G; assume (M,g) is an oriented pseudo-Riemannian manifold of dimension m and that G is a connected Lie group of dimension n. Let γ be a symmetric bilinear form on the Lie algebra $\mathscr{g}$ of G; let ω be a connection one-form on P (which is $\mathscr{g}$ valued) and let Ω be the corresponding curvature two-form. Let $\{U_\alpha;\ \alpha\in A$, an indexing set$\}$ be an open cover of M such that $P|_{U\alpha}$ is trivial and let $\sigma_\alpha: U_\alpha\to P$ be a section of $P|_{U\alpha}$. The curvature field $F_\alpha\equiv\sigma_\alpha{*}\Omega$ can be used to define an m-form on U_α, namely

$$F_\alpha\wedge{}^*F_\alpha$$

(where * is the Hodge dual relative to the pseudo-Riemannian structure g and the given orientation) which has values in $\mathscr{g}\otimes\mathscr{g}$. If one evaluates γ on this element one obtains a real-valued m-form, which we shall denote

$$L_\alpha=\gamma(F_\alpha\wedge\ {}^*F_\alpha). \tag{2.1}$$

If γ is Ad G invariant, L_α is the restriction to U of a global m-form L on M. If M is compact, or if the curvature Ω has a support F such that $\pi(F)$ is compact, one can define the integral

$$S=\int_M L\equiv\int_M \mathscr{L}v_g, \tag{2.2}$$

where v_g is the pseudo-Riemannian volume form associated to the given orientation; the function $\mathscr{L}$ is the Lagrangian.

Assume the open set U is the domain of a chart of (M,g) and denote by x^a $(a\leqslant m)$ the corresponding coordinates: let σ be a section of $P|_U$. Choose a basis X_A $(A\leqslant n)$ of $\mathscr{g}$ and let $C_{AB}{}^C$ be the structure constants of $\mathscr{g}$ in this basis. Write

$$\phi\equiv\sigma{*}\omega,\quad F\equiv\sigma{*}\Omega,$$

and thus components in this coordinate basis read

$$(\sigma{*}\omega)\Big(\frac{\partial}{\partial x^a}\Big)\equiv(\sigma{*}\omega)_a\equiv\phi_a{}^A(x)X_A, \tag{2.3a}$$

$$(\sigma{*}\Omega)\Big(\frac{\partial}{\partial x^a},\frac{\partial}{\partial x^b}\Big)\equiv(\sigma{*}\Omega)_{ab}\equiv F_{ab}{}^A(x)X_A. \tag{2.3b}$$

Then

$$F_{ab}{}^A(x)=\phi_{b,a}{}^A(x)-\phi_{a,b}{}^A(x)+\phi_a{}^B(x)\phi_b{}^C(x)C_{BC}{}^A, \tag{2.4}$$

where $a=\partial/\partial x^a$. Also

$$\mathscr{L}=\tfrac12\gamma_{AB}g^{ij}g^{kl}F_{ik}{}^AF_{jl}{}^B \tag{2.5}$$

with

$$\gamma_{AB}=\gamma(X_A,X_B),\quad g^{ij}=(g^{-1})_{ji},\quad g_{kl}=g\Big(\frac{\partial}{\partial x^k},\frac{\partial}{\partial x^l}\Big).$$

The Euler–Lagrange equations corresponding to the functional S (defined on some space of connections) read

$$\gamma({}^*d{}^*F+{}^*[{}^*F\wedge\phi])=0, \tag{2.6}$$

where $\gamma:\mathscr{g}\to\mathscr{g}^*$ is the linear map associated to γ. Expressing

[a)] Chercheur qualifle au F.N.R., Belgium.
[b)] Present address: FAMAF, Laprida 854, 5000 Cordoba, Argentina.

 0022-2488/88/041022-04$02.50

(2.6) in the chart U one gets

$$\gamma_{AB}\,[\nabla_s F_{tb}{}^B + \phi_s{}^C F_{tb}{}^D C_{CD}{}^B]\,g^{st} = 0, \tag{2.6'}$$

where ∇ is the covariant derivative relative to the Levi-Civita connection of $\mathscr{g}$.

We will consider (by definition) that (2.6′) are the Yang–Mills equations. With γ_{AB} an invariant quadratic form on the Lie algelera we see that they are the Euler–Lagrangian equation of the Lagrangian (2.5). The converse question then arises; given Eq. (2.6′), when are they derivable from a Lagrangian?

Expanding (2.6′) in terms of the derivatives of $\phi_a{}^A$ we have

$$\begin{aligned}
&(g^{lk}g^{lj} - g^{lj}g^{ki})\gamma_{AB}\ \partial^2{}_{ij}\phi_\kappa{}^B\\
&\quad + (g^{lb}g^{jk}C_{BK}{}^D\phi_b{}^K\gamma_{AD} + 2g^{lk}g^{js}C_{GB}{}^D\phi_s{}^G\gamma_{AD}\\
&\quad - g^{lj}g^{kt}C_{GB}{}^D\phi_t{}^G\gamma_{AD} - g^{lk}g^{st}\Gamma_{st}{}^j\gamma_{AB} + g^{lj}g^{st}\Gamma_{st}{}^k\gamma_{AB}\\
&\quad - g^{lb}g^{jt}\Gamma_{tb}{}^k\gamma_{AB} - g^{lb}g^{kt}\Gamma_{tb}{}^j\gamma_{AB})\partial_j\phi_\kappa{}^B\\
&\quad - (g^{lb}g^{st}\gamma_{AD}\Gamma_{ts}{}^m C_{GK}{}^D\phi_m{}^G\phi_b{}^K\\
&\quad + g^{lb}g^{st}\gamma_{AD}\Gamma_{tb}{}^m C_{GK}{}^D\phi_s{}^G\phi_m{}^K\\
&\quad - g^{lb}g^{st}\gamma_{AD}C_{PQ}{}^D C_{GK}{}^Q\phi_s{}^G\phi_b{}^K).
\end{aligned} \tag{2.6''}$$

The $\Gamma_{ts}{}^m$ are the Christoffel symbols of g.

The field equations are now defined to be Eqs. (2.6″) where γ is any nondegenerate quadratic form on $\mathscr{g}$. (They are equivalent (as γ is nondegenerate) to

$$[\nabla_s F_{tb}{}^C + \phi_s{}^A F_{tb}{}^B C_{AB}{}^C]\,g^{st} = 0.) \tag{2.6'''}$$

In what follows we shall denote by ϕ_α ($\alpha \leqslant N = nm$) the components of the connection $\phi_k{}^B$.

We now wish to find under what circumstances (2.6″) are the Euler–Lagrange equations of a Lagrangian. We first state Lemma I, which is proved in the Appendix.

Lemma 1: Let U be a contractible open set of R^m; let E^r be the Banach space of fields ϕ_α, $\alpha \leqslant N$, defined on U, compactly supported and of class C^r ($r \geqslant 3$). If the field equations (2.6″) are the Euler–Lagrange equations associated to a Lagrangian $\mathscr{L}$, defined on $U \times E^r \times E^{r-1}$,

$$\mathscr{L} = \mathscr{L}(x^a, \phi_\alpha, \partial_a\phi_\alpha),$$

which has a polynomial dependence in the last variables and which is at least twice differentiable, then this polynomial may be assumed to be of degree 2 and the coefficients of the terms of order 2 depend on x^a only.

Remarks: (i) The Banach space $E^r \times E^{r-1}$ can be replaced by another "reasonable" Banach space associated to this problem; it is introduced only to be able to use in the proof Poincaré's lemma in an infinite-dimensional framework.

(ii) The Lagrangian (2.5) is of the type described in Lemma I.

Lemma 2: If the field equations (2.6″) are the Euler–Lagrange equations of a Lagrangian $\mathscr{L}$ as in Lemma I, then γ is necessarily an invariant nondegenerate quadratic form on $\mathscr{g}$.

Proof: By Lemma 1, one may assume that

$$\mathscr{L} = \sum a^{ij,k}{}_A{}^l{}_B(x)\partial_i\phi_k{}^A\ \partial_j\phi_l{}^B$$
$$+ \sum b^{i,k}{}_A(x,\phi)\partial_i\phi_k{}^A + c(x,\phi).$$

The corresponding $({}^j{}_A)$ Euler–Lagrange equation, $d/dx^a(\partial\mathscr{L}/\partial[\partial_a\phi_j{}^A] - \partial\mathscr{L}/\partial\phi_j{}^A) = 0$, reads

$$\begin{aligned}
&2a^{ij,k}{}_A{}^l{}_B(x)\partial^2{}_{ij}\phi_k{}^A\\
&\quad + \left(2\partial_i a^{ji,k}{}_A{}^l{}_B(x) - \frac{\partial(b^{j,k}{}_B)}{\partial\phi_l{}^A} + \frac{\partial(b^{j,l}{}_A)}{\partial\phi_l{}^B}\right)\partial_j\phi_k{}^B\\
&\quad + \partial_i b^{j,l}{}_A - \frac{\partial c}{\partial\phi_l{}^A} = 0.
\end{aligned}$$

By comparing the coefficients of the terms linear in $\partial_j\phi_k{}^B$ with the equivalent terms in (2.6″) one sees that

$$\begin{aligned}
&\frac{\partial(b^{j,l}{}_A)}{\partial\phi_l{}^B} - \frac{\partial(b^{j,k}{}_B)}{\partial\phi_l{}^A}\\
&\quad = \gamma_{AD}C_{BK}{}^D\phi_b{}^K(g^{lb}g^{jk} - 2g^{jb}g^{lk} + g^{lj}g^{kb})\\
&\qquad + S^{jkl}{}_{BA}(x).
\end{aligned}$$

By antisymmetry in $({}^k{}_B, {}^l{}_A)$ of the left-hand side we have

$$\gamma_{AD}C_{BK}{}^D + \gamma_{BD}C_{AK}{}^D = 0. \tag{2.7}$$

This is the condition for the invariance of γ by ad $\mathscr{g}$. In order to see this, consider the invariance condition

$$\gamma_{AB}V^AV^B = \gamma_{AB}U^A{}_C U^B{}_D V^C V^D,$$

with V^A in the Lie algebra and $U^A{}_B$ in the adjoint representation of G. Letting $U^A{}_B = \delta^A{}_B + \epsilon^D C_{DB}{}^A$ we have Eq. (2.7) as a consequence.

Lemmas I and 2 prove that the Yang–Mills equations (2.6″) corresponding to a Lie group G are derived from a variational principle if and only if the Lie algebra $\mathscr{g}$ of G admits a nondegenerate invariant quadratic form.

Observe that the invariance condition, which follows from local arguments, in a given trivialization of the bundles, ensures (and it is necessary to have) a global well defined Lagrangian.

[Observe also that the field equations written as (2.6‴) are *not* of the Euler–Lagrange type.]

III. EXAMPLES

In this section we show that nondegenerate solutions of (2.7) do exist for certain solvable groups or algebras (e.g., the extended Heisenberg algebra) and sometimes they do not (e.g., the two-dimensional Borel algebra). Some of these examples have been considered elsewhere in relation with pseudo-Riemannian spaces.[2–4]

(a) Semisimple Lie algebras: We first point out in the well known case of semisimple Lie algebras that, using the Jacobi identity, the Killing form

$$K_{AB} \equiv \gamma_{AB} = C^C{}_{AD}C^D{}_{BC}$$

identically satisfies (2.7). In fact the Killing form can be defined for any group and will always satisfy (2.7). The Killing form, however, is nondegenerate only for semisimple Lie algebras. One is thus forced to seek other nondegenerate quadratic forms if one considers solvable Lie algebras.

(b) Solvable Lie algebras: A solvable Lie algebra $\mathscr{g}_0$ is an algebra such that the chain of subalgebras $\mathscr{g}_0,\mathscr{g}_1,\mathscr{g}_2,\mathscr{g}_3,\ldots,\mathscr{g}N=0$, where $\mathscr{g}_{n+1}$ is the commutator subalgebra of $\mathscr{g}_n$, terminates with the trivial algebra.

(i) The two-dimensional Borel algebra: The algebra of the group of affine transformations of the real line $(x\rightarrow ax+b)$ has no nondegenerate invariant quadratic form. This is seen from the fact that its algebra defined by $[X_1,X_2]=-X_2$ (with $X_1=x\partial_x$ and $X_2=\partial_x$) has the structure constant $C^2{}_{12}=-1$ which when substituted into (2.7) yields

$$\gamma_{12}=\gamma_{22}=0,$$

a *degenerate* quadratic form.

(ii) The extended Heisenberg algebra: This $(2n+2)$-dimensional algebra consists of the generators (z^*,x_i,y_j,z), with $i,j=1,\ldots,n$, satisfying the following commutation relations (with no summations implied):

$$[z^*,x_i]=\alpha_i y_i\rightarrow C^{y_i}{}_{z^*x_i}=\alpha_i,$$

$$[z^*,y_i]=\beta_i x_i\rightarrow C^{x_i}{}_{z^*y_i}=\beta_i,$$

$$[x_i,y_j]=\gamma_i\delta_{ij}z\rightarrow C^z x_i y_j=\gamma_i\delta_{ij},$$

all other commutators and structure constants vanishing. Systematically substituting these structure constants into (2.7) yields the following conditions on the quadratic form:

$$\gamma_{zz}=\gamma_{x_iz}=\gamma_{y_iz}=\gamma_{x_iy_j}=\gamma_z x_i=\gamma_{z^*y_i}=0,$$

$$\gamma_{x_ix_j}=\gamma_{y_iy_j}=0\quad(i\neq j),$$

$$\gamma_{x_ix_i}C^{x_i}{}_{z^*y_i}+\gamma_{zz^*}C^z{}_{x_iy_i}=0,\quad\text{no summation},$$

$$\gamma_{y_iy_i}C^{y_i}{}_{z^*x_i}+\gamma_{zz^*}C^z{}_{y_ix_i}=0,\quad\text{no summation}.$$

Note that the component $\gamma_{z^*z^*}$, though not determined, can be set equal to zero by adding to z an arbitrary multiple of z^*. Since z^* is the center of the algebra this operation does not alter the structure constants. If the γ_{zz^*} is chosen arbitrarily (say equal to one) then the quadratic form is uniquely determined by the above conditions and is easily seen to be nondegenerate. If the α_i, β_i, and γ_i are chosen to be one, then the metric is the Lorentz metric.

(iii) The derived subalgebra of the extended Heisenberg algebra: If we consider the $(2n+1)$-dimensional subalgebra of the extended Heisenberg algebra defined by omitting the generator z^* we have the only nonvanishing commutators

$$[x_i,y_j]=\gamma_i\delta_{ij}z\rightarrow C^z{}_{x_iy_j}=\gamma_i\delta_{ij}$$

making this a nilpotent algebra. It is now easy to check from (2.7) that the associated quadratic form is now degenerate, namely

$$\gamma_{zz}=\gamma_{x_iz}=\gamma_{y_iz}=0.$$

(iv) Nilpotent algebras with nondegenerate invariant forms: Let $\mathscr{g}=\Lambda^2\mathbf{R}^{*n}\oplus(\mathbf{R}^{*n}\oplus\mathbf{R}^n)$, where $n\geqslant 2$, $*$ is a linear isomorphism $\mathbf{R}^n\rightarrow\mathbf{R}^{*n}$, and $\Lambda^2\mathbf{R}^{*n}$ is the exterior product of $\mathbf{R}^{*n}$ with itself. Let R: $\mathbf{R}^{*n}\times\mathbf{R}^{*n}\times\mathbf{R}^{*n}\rightarrow\mathbf{R}^{*n}$ be a trilinear map having all the symmetries of a curvature map. Define a multiplication table by

$$[\mathbf{X}^*,\mathbf{Y}^*]=\mathbf{X}^*\wedge\mathbf{Y}^*,\quad \mathbf{X}^*,\mathbf{Y}^*\in\mathbf{R}^{*n},$$

$$[\mathbf{X}^*\wedge\mathbf{Y}^*,\mathbf{Z}^*]={*}^{-1}\mathbf{R}(\mathbf{X}^*,\mathbf{Y}^*,\mathbf{Z}^*),$$

all other commutators vanishing. The invariant quadratic form becomes

$$\gamma(\mathbf{X}^*,\mathbf{Y})=\langle\mathbf{X},\mathbf{Y}\rangle,$$

$$\gamma(\mathbf{X}^*\wedge\mathbf{Y}^*,\mathbf{Z}^*\wedge\mathbf{U}^*)=\langle X,Z\rangle\langle Y,U\rangle-\langle X,U\rangle\langle Y,Z\rangle,$$

all other components vanishing. Here $\langle\ ,\ \rangle$ is the usual scalar product on $\mathbf{R}^n$. This algebra corresponds to a symmetric space of dimension $2n$ and signature (n,n).

A special case of this is the five-dimensional algebra $(n=2)$ with generators X_1, X_2, $X^*{}_1$, $X^*{}_2$, and $X^*{}_{12}$ given by

$$[X^*{}_1,X^*{}_2]=X^*{}_{12},$$

$$[X^*{}_{12},X^*{}_1]=X_2,\quad (X^*{}_{12},X^*{}_2)=-X_1.$$

The nonvanishing components of the quadratic form are

$$\gamma(X_i,X^*{}_j)=\delta_{ij},\quad \gamma(X_{12},X_{12})=1.$$

IV. DISCUSSION

We have here shown that the necessary and sufficient condition for a set of Yang–Mills equations to be derivable from a Lagrangian is that there exist an invariant (under the action of the associated group) nondegenerate quadratic form. Though a common belief is that this implies that the associated group must be semisimple, we have shown that this is not necessary and that in fact there are many solvable groups with such a quadratic form. Though there is no known classification of solvable groups which describes those possessing an invariant nondegenerate form, it is clear that given the structure constants of a group, Eq. (2.7) permits one to find the form if it exists or show that it does not exist.

Several interesting questions arise from these considerations.

(1) Considering the case of a solvable group with a Lagrangian and hence Hamiltonian, what can one say about the quantization of such a system? As mentioned in the Introduction the underlying classical equations have a simple linear structure, with each linear set behaving as a problem with an external potential (the solutions from the previous set). Does this feature of the equations aid in the quantization and if so how? This is under investigation.

(2) A related question is are there any solvable groups with an invariant quadratic form such that the form is positive definite? Unfortunately the answer to this question is no. This follows from the fact that for a given Lie algebra, a positive definitive form exists if and only if this algebra is that of a compact group.[5] Since a solvable group can never be compact, no such form exists for its Lie algebra. This would seem to prohibit a physical quantization procedure since an indefinite form would introduce an energy operator without a lower bound.

ACKNOWLEDGMENTS

We would like to thank A. Ashtekar, A. Duncan, A. Helfer, M. Hickman, L. Lemaire, and C. Lucey for enlightening comments and discussions.

E. Newman and C. Kozameh thank the National Science Foundation for support.

APPENDIX: PROOF OF LEMMA 1

The Lagrangian $\mathscr{L}$ can be written as

$$\mathscr{L} = \sum A_{a_{11}\cdots a_{nN}} (\partial_1 \phi_1)^{a_{11}} \cdots (\partial_n \phi_N)^{a_{nM}}, \qquad (\mathrm{A1})$$

where the coefficients $A_{a_{11}\cdots a_{nN}}$ are functions of x^a, ϕ_α admitting continuous partial derivatives up to order 2 and where $a_{11} + \cdots + a_{nN} \leqslant M =$ order of the polynomial $\mathscr{L}$. Comparing the coefficient of a second-order derivative $\partial^2{}_{ii}\phi_\alpha$ in the Euler–Lagrange equations of (A1) with (2.6″) one sees that $A_{a_{11}\cdots a_{nN}}$ is zero if any of the $a_{k\lambda} > 2$; also $A_{a_{11}\cdots a_{nN}} = 0$ as soon as one of the $a_{k\lambda}$'s is equal to 2 and one other $a_{k\lambda}$ is not zero. We can thus write the highest-order term (if it exists for $r > 2$) in the polynomial as

$$\mathscr{L} = \sum D_{i_1\cdots i_r,\alpha_1\cdots\alpha_r}(x,\phi)\partial_{i_1}\phi_{\alpha_1}\cdots\partial_{i_r}\phi_{\alpha_r}$$

with no pair $(i_j\alpha_j)$ repeated. If $r > 2$, writing the term in

$$\partial^2{}_{i_k i_l}\phi_{\alpha_k}\partial_{\alpha_1}\phi_{\alpha_1}\cdots(\partial_{i_k}\wedge\phi_{\alpha_k})\cdots\partial_{i_r}\phi_{\alpha_r} \quad (k \neq l)$$

in the equation numbered α_1 one sees $D_{i_1\cdots i_r,\alpha_1\cdots\alpha_r}$ is antisymmetric in i_k, i_l, hence totally antisymmetric in the indices $i_1\cdots i_r$. If $r = 2$, one can write $D_{i_1 i_2,\alpha_1\alpha_2}$ as the sum of a part antisymmetric in $i_1 i_2$ and a part which is symmetric in $i_1 i_2$. Using the form of (2.6″) one sees that the symmetric coefficient is a function of x alone. Hence

$$\mathscr{L} = \sum A_{ij,\alpha\beta}(x)\partial_i\phi_\alpha\partial_j\phi_\beta + \sum B_{i\alpha}(x,\phi)\partial_i\phi_\alpha + C$$
$$+ \sum d_{i_1\cdots i_r,\alpha_1\cdots\alpha_r}(x,\phi)\partial_{i_1}\phi_{\alpha_1}\cdots\partial_{i_r}\phi_{\alpha_r},$$

where $A_{ij,\alpha\beta} = A_{ji,\alpha\beta}$ and d is totally antisymmetric. Consider then the coefficient of the term of order r, in $\partial_{i_1}\phi_{\alpha_1}\cdots\partial_{i_r}\phi_{\alpha_r}$, in the equation numbered α; it is proportional to

$$-\frac{\partial(d_{i_1\cdots i_r,\alpha_1\cdots\alpha_r})}{\partial\phi_\alpha} + \sum_{k=1}^{r}\left(\frac{\partial}{\partial\phi_{\alpha_k}}\right)d_{i_1\cdots i_r,\alpha_1\cdots\alpha_r}.$$

If $r \geqslant 2$ this must be equal to zero. By the Poincaré lemma,[6] there exist functions $e_{i_1\cdots i_r,\alpha_1\cdots\alpha_{r-1}}(x,\phi)$ such that

$$d_{i_1\cdots i_r,\alpha_1\cdots\alpha_r} = \sum_{k=1}^{r}(-1)^{k-1}\left(\frac{\partial}{\partial\phi_{\alpha_k}}\right)e_{i_1\cdots i_r,\alpha_1\cdots\hat{\alpha}_k\cdots\alpha_r}.$$

Modifying the Lagrangian $\mathscr{L}$ by a divergence

$$\sum_{j=1}^{r}(-1)^{j-1}\left(\frac{d}{dx_{i_j}}\right)$$
$$\times[e_{i_1}\cdots i_r,\alpha_1\cdots\alpha_{r-1}\partial_{i_1}\phi_{\alpha_1}\cdots\partial_{i_{j-1}}\phi_{\alpha_{j-1}}$$
$$\times\partial_{i_{j+1}}\phi_{\alpha_j}\cdots\partial_{i_r}\phi_{\alpha_{r-1}}],$$

one gets a new Lagrangian $\mathscr{L}'$ for which the field equations are still the Euler–Lagrange equations and which has no terms of order r. The result follows by induction.

[1]A. Helfer, M. Hickman, C. Kozameh, C. Lucey, and E. Newman, "Yang–Mills equations and solvable groups," Phys. Rev. D. **36**, 1740 (1987).

[2]A. L. Besse, *Einstein Manifolds, Springer Ergebnisse der Mathematik*, Vol. 10 (Springer, Berlin, 1987), p. 207.

[3]M. Cahen and N. Wallach, "Lorenzian symmetric spaces," Bull. Am. Math. Soc. **76**, 585 (1970).

[4]M. Cahen and M. Parker, "Sur des classes d'espaces pseudo-Reimanniens symetriques," Bull. Soc. Math. Belg. **22**, 339 (1970).

[5]G. Hochschild, *The Structure of Lie Groups* (Holden–Day, San Francisco, 1965), p. 142.

[6]S. Lang, *Differential Manifolds* (Addison–Wesley, Reading, MA, 1972), p. 124.

$N=2$ supersymmetric Ward identities on the harmonic superspace

T. Lhallabi[a)]
International Centre for Theoretical Physics, Trieste, Italy

(Received 1 September 1987; accepted for publication 21 October 1987)

The quantization of $N=2$ supersymmetric Yang–Mills theory coupled to a matter hypermultiplet has been done in the harmonic superspace by requiring BRS and anti-BRS invariances. The corresponding Ward identities have also been derived.

I. INTRODUCTION

Supersymmetric field theories have been shown to have remarkable cancellations of ultraviolet divergences. The cancellation is expected to be even more dramatic in the case of extended supersymmetry. These cancellations are most easily seen in the context of superfield perturbation theory, where supersymmetry is manifest. However, superfield perturbation theory has been developed, so far, only for $N=1$ supersymmetry. Indeed, in $N=1$ supersymmetric theories, the quadratic divergences are absent, a fact that gives a key to resolving the hierarchy problem.[1] In supersymmetric models the cancellations of the ultraviolet divergences are due to the supersymmetric Ward identities (WI). The breaking of these identities leads to additional ultraviolet divergences in perturbation theory.[2] For theories with extended supersymmetry, the supersymmetric WI are fulfilled when we use the extended superfield formalism.[3]

On the other hand, the supersymmetric theories need to be regularized. Several approaches are possible. One of these is the dimensional regularization which is not supersymmetric.[4] A remedy for this problem was proposed in the dimensional reduction scheme, which breaks the supersymmetric WI in the $N=1$ supersymmetric Yang–Mills theory, in the Wess–Zumino gauge at the one-loop level.[5] Another possible scheme is the gauge-invariant higher derivative method preserving manifestly the supersymmetry.[6]

In order to handle the ultraviolet divergencies and satisfy the supersymmetric WI of extended supersymmetric theories, it is apparently necessary to develop the superfield formulation.[7] Very recently, Galperin *et al.*[8] have put forward a new idea called "harmonic superspace," and have developed $N=2$ superspace formulations. Their method has an advantage in that their superfields are not constrained. The main features of this formalism is that it contains the zweibeins $U_i^{\pm}$ $(i=1,2)$, which parametrize the coset space SU(2)/U(1), and that any superfield contains an infinite number of component fields.

Superfield perturbation methods for this theory have been developed in two ways: by requiring the BRS and anti-BRS invariance,[9] which is the clue to the proof of renormalizability; and by using the Faddeev–Popov procedure, where the quantum Lagrangian is obtained via covariant gauge.[10,11]

In this paper we complete the construction of the $N=2$ supersymmetric quantum Lagrangian of the gauge and matter hypermultiplets using the BRS and anti-BRS prescription. Furthermore, we construct, in the harmonic superspace, the $N=2$ supersymmetric WI. Following Lee[6] and Zinn-Justin,[12] the renormalized $N=2$ supersymmetric action can be constructed in such a way that it still satisfies the WI.

The paper is organized as follows. In Sec. II we give some preliminaries on an $N=2$ harmonic superspace, while in Sec. III we derive the BRS and anti-BRS transformations for gauge and matter hypermultiplets. We then construct the $N=2$ supersymmetric quantum Lagrangian corresponding to gauge and matter hypermultiplets. In Sec. IV we obtain the $N=2$ supersymmetric WI of BRS and anti-BRS symmetries. At last we construct the algebra of renormalized BRS and anti-BRS operators, which leave the renormalized Lagrangian invariant. Finally, Sec. V is devoted to the conclusion.

II. GENERALITIES ON THE $N=2$ HARMONIC SUPERSPACE

In this section we recall the main features of the $N=2$ harmonic superspace, following the conventions and notations of Refs. 8 and 10. In this space the harmonic variables $U_i^{\pm}$ $(i=1,2)$, called zweibeins, play an important role in the construction of N extended supersymmetric theories. The zweibeins are SU(2) isodoublets satisfying the properties

$$U^{+i}U_i^{-}=1,\quad U^{\pm i}U_i^{\pm}=0. \tag{2.1}$$

The $N=2$ harmonic superspace in the real or central basis (CB) is parametrized by the coordinates

$$(x_\mu,\ \theta_{\alpha i},\ \overline{\theta}_{\dot\alpha}{}^{i},\ U_i^{\pm}), \tag{2.2}$$

where $\theta_{\alpha,i}$ and $\overline{\theta}{}^{i}_{\dot\alpha}$ are Weyl spinors and SU(2) isospinors. From (2.2) one can pass to another basis, the analytic basis (AB), given by

$$Z=(X_A^\mu,\ \theta_\alpha^{+},\ \overline{\theta}{}_{\dot\alpha}^{+},\ \theta_\alpha^{-},\ \overline{\theta}{}_{\dot\alpha}^{-},\ U_i^{\pm}). \tag{2.3}$$

In this new AB the $N=2$ supersymmetry is realized as follows:

$$\begin{aligned}
&\delta X_A^\mu=-2i(\epsilon^i\sigma^\mu\overline{\theta}{}^{+}+\theta^{+}\sigma^\mu\overline{\epsilon}{}^{i})U_i^{-},\\
&\delta\theta_\alpha^{+}=\epsilon_\alpha^iU_i^{+},\quad \delta\theta_\alpha^{-}=\epsilon_\alpha^iU_i^{-},\\
&\delta\overline{\theta}{}_{\dot\alpha}^{+}=\overline{\epsilon}{}_{\dot\alpha}^iU_i^{+},\quad \delta\overline{\theta}{}_{\dot\alpha}^{-}=\overline{\epsilon}{}_{\alpha}^iU_i^{-},\quad \delta U_i^{\pm}=0.
\end{aligned} \tag{2.4}$$

We notice that the variables $(x_A^\mu,\ \theta_\alpha^{+},\ \overline{\theta}{}_{\dot\alpha}^{+},\ U_i^{\pm})$ and $(x_R^\mu,\ \overline{\theta}{}_{\dot\alpha}^{+},\ \overline{\theta}{}_{\dot\alpha}^{-},\ U_i^{\pm})$, where

[a)] Permanent address: L. P. T. Faculté des Sciences, Av. Ibn Batouta, P. O. Box 1014, Rabat, Morocco.

 0022-2488/88/041026-06$02.50

$$X_R^\mu = X^\mu + i\theta^i\sigma^\mu\bar{\theta}_i, \tag{2.5}$$

form closed subsets under the $N=2$ supersymmetric transformations. We call these subspaces analytic and chiral subspaces, respectively. This allows us to define an analytic superfield $\phi^q(X_A,\theta^+,\bar{\theta}^+,U^\pm)$, which satisfies the conditions

$$D^+\phi^q = 0 = \bar{D}^+\phi^q,$$

where q is the $U^c(1)$ charge, and

$$D^+ = \frac{\partial}{\partial\theta^-}, \quad \bar{D}^+ = \frac{\partial}{\partial\bar{\theta}^-}.$$

The $N=2$ gauge hypermultiplet is described by the real analytic superfield V^{++}. The field strength W, which is a chiral superfield independent of $U_i^\pm$ variables, gives the action[10,11]

$$I_1 = \int dX_R\, d^2\bar{\theta}^+\, d^2\bar{\theta}^-\, \mathrm{Tr}\, W^2, \tag{2.6}$$

which can be expanded as

$$I_1 = \int d^4X_A\, d^2\theta^+\, d^2\bar{\theta}^+\, dU \times \mathrm{Tr}\left[V^{++}\frac{1}{D^{++}}\frac{(D^+)^4D^{--}}{16}V^{++}\right] + I_{\mathrm{int}}(V^{++}). \tag{2.7}$$

Therefore there exist two $N=2$ matter hypermultiplets. The first is the Fayet–Sohnius (FS) hypermultiplet, which is described by means of a complex analytic superfield ϕ^+. Its kinetic term is given by

$$I_2 = \int d^4X_A\, d^2\theta^+\, d^2\bar{\theta}^+ dU \overset{*}{\phi}{}^+ D^{++}\phi^+, \tag{2.8}$$

where

$$D^{++} = \partial^{++} - 2i\theta^+\sigma^\mu\bar{\theta}^+\partial_\mu^A,$$

with

$$\partial^{++} = U^{+i}\frac{\partial}{\partial U^{-i}}.$$

The second is the Howe–Stelle–Townsend (HST) hypermultiplet, described by a real analytic superfield Ω. Its free action is

$$I_3 = \int d^4X_A\, d^2\theta^+\, d^2\bar{\theta}^+\, dU(D^{++}\Omega)(D^{++}\Omega). \tag{2.9}$$

Therefore the full action of the FS hypermultiplet, including interactions with the gauge superfield, is

$$I_2^1 = \int d^4X_A\, d^2\theta^+\, d^2\bar{\theta}^+\, dU \overset{*}{\phi}{}^+\mathscr{D}^{++}\phi^+, \tag{2.10}$$

while for the HST hypermultiplet we have

$$I_3^1 = \int d^4X_A\, d^2\theta^+\, d^2\bar{\theta}^+\, dU(\mathscr{D}^{++}\Omega)(\mathscr{D}^{++}\Omega), \tag{2.11}$$

where

$$\mathscr{D}^{++} = D^{++} + iV^{++}$$

is the covariant derivative.

Furthermore, the analytic superfields V^{++}, ϕ^+, and Ω transform under gauge transformation of any gauge group G as

$$V^{++\prime} = e^{i\Lambda}(V^{++} - iD^{++})e^{-i\Lambda}, \tag{2.12}$$

$$\phi'^+ = e^{i\Lambda}\phi^+, \quad \Omega' = e^{i\Lambda}\Omega e^{-i\Lambda}, \tag{2.13}$$

where Λ is an analytic real superfield.

As can be seen from Ref. 9, the action (2.6) is BRS and anti-BRS invariant. In what follows we shall derive the BRS and anti-BRS transformations for the $N=2$ matter hypermultiplets, and see what is the most general form of an $N=2$ supersymmetric, BRS and anti-BRS invariant Lagrangian that includes a matter hypermultiplet.

III. EXTENDED BRS AND ANTI-BRS EQUATIONS

Let G be a compact gauge group, and let V^{++}, C, C' and b be the classical superfields (superconnections). The perturbative gauge theory is generated by $\mathscr{L}_Q$, an $N=2$ supersymmetric, gauge-fixed, but BRS (δ_ξ) and anti-BRS ($\delta_{\xi'}$) invariant Lagrangian. The Lagrangian $\mathscr{L}_Q$ has zero $U^c(1)$ charge, as well as ghost number or Faddeev–Popov charge zero, and is renormalizable by power counting. The mathematical expression for $\mathscr{L}_Q$ is given by[9]

$$\mathscr{L}_Q = \mathscr{L}_{GI} + \frac{1}{2}\int d^2\theta^+\, d^2\bar{\theta}^+\, dU \times \mathrm{Tr}\,\delta_\xi\delta_{\xi'}\{V^{++}V^{++} + 2\alpha C'M^{4+}C\} \tag{3.1}$$

where $\mathscr{L}_{GI}$ is given by (2.6), α is a gauge parameter, and M^{4+} is an arbitrary operator of mass dimension (-2) with a $U^c(1)$ charge of $(+4)$.

The BRS and anti-BRS symmetries are defined by the actions of their generators on all superfields[9] as

$$\begin{aligned}
&\delta_\xi V^{++} = -D^{++}C - i[V^{++},C], && \delta_{\xi'} V^{++} = -D^{++}C' - i[V^{++},C'],\\
&\delta_\xi C = -C\times C, && \delta_{\xi'} C = -b + C\times C',\\
&\delta_\xi C' = b, && \delta_{\xi'} C' = -C'\times C',\\
&\delta_\xi b = 0,
\end{aligned} \tag{3.2}$$

where

$$A\times B = \tfrac{1}{2}[A,B].$$

Equations (3.2) are obtained by extending the analytic subspace to $(X_A^\mu,\theta^+,\bar{\theta}^+,\xi,\xi')$, or by making a special gauge transformation with a gauge parameter $\tilde{\Lambda}$ such that

$$\tilde{\Lambda} = \xi C + \xi' C' + \xi\xi' b. \tag{3.3}$$

Now if we include matter superfields in our formalism, the gauge transformations (2.13) can be extended to

$$\widetilde{\phi}^+ = e^{i\widetilde{\Lambda}}\phi^+, \quad \widetilde{\Omega} = e^{i\widetilde{\Lambda}}\Omega e^{-i\widetilde{\Lambda}}. \tag{3.4}$$

On the other hand, $\widetilde{\phi}^+$ and $\widetilde{\Omega}$ can be expanded in the extended analytic subspace as[9]

$$\widetilde{\phi}^+ = \phi^+ + \xi\delta_\xi\phi^+ + \xi'\delta_{\xi'}\phi^+ + \xi\xi'\delta_\xi\delta_{\xi'}\phi^+,$$
$$\widetilde{\Omega} = \Omega + \xi\delta_\xi\Omega + \xi'\delta_{\xi'}\Omega + \xi\xi'\delta_\xi\delta_{\xi'}\Omega. \tag{3.5}$$

Hence

$$\delta\phi^+ = \xi\delta_\xi\phi^+ + \xi'\delta_{\xi'}\phi^+ + \xi\xi'\delta_\xi\delta_{\xi'}\phi^+,$$
$$\delta\Omega = \xi\delta_\xi\Omega + \xi'\delta_{\xi'}\Omega + \xi\xi'\delta_\xi\delta_{\xi'}\Omega. \tag{3.6}$$

Expanding (3.4) for an infinitesimal transformation, one obtains

$$\delta\phi^+ = i\widetilde{\Lambda}\phi^+, \quad \delta\Omega = i[\widetilde{\Lambda},\Omega]. \tag{3.7}$$

The identification of (3.6) and (3.7) leads to the BRS and anti-BRS transformations of ϕ^+ and Ω:

$$\delta_\xi\phi^+ = iC\phi^+, \quad \delta_{\xi'}\phi^+ = iC'\phi^+,$$
$$\delta_\xi\Omega = i[C,\Omega], \quad \delta_{\xi'}\Omega = i[C',\Omega]. \tag{3.8}$$

It is clear that the actions (2.10) and (2.11) are invariant under BRS and anti-BRS symmetries.

However, in order to construct the most general δ_ξ and $\delta_{\xi'}$, $N=2$ supersymmetric invariant Lagrangian, with zero $U^c(1)$ charge and ghost number conserving, by the Yang–Mills theory coupled to the FS multiplet, consider terms obtained by the action of δ_ξ, $\delta_{\xi'}$, and $\delta_\xi\delta_{\xi'}$ on a polynomial of superfields. One has the following possibilities:

$$\delta_\xi\delta_{\xi'}[\lambda^{2+}\overset{*}{\phi}{}^+\phi^+], \tag{3.9a}$$
$$\delta_\xi\delta_{\xi'}[v^{3+}\phi^+ + \overline{v}^{3+}\overset{*}{\phi}{}^+], \tag{3.9b}$$
$$\delta_\xi[\lambda^{2+}C'\overset{*}{\phi}{}^+\phi^+], \tag{3.9c}$$
$$\delta_{\xi'}[\lambda^{2+}C\overset{*}{\phi}{}^+\phi^+], \tag{3.9d}$$

where λ^{2+} and v^{3+} are coupling constants with mass dimension (-2) and (-1), respectively.

Now while steps (3.9d) and (3.9c) are excluded because they are not BRS and anti-BRS invariant, respectively, steps (3.9a) and (3.9b) are convenient terms. Hence the most general $N=2$ supersymmetric Lagrangian that is invariant with respect to δ_ξ and $\delta_{\xi'}$ and is a function of ϕ^+ and $\overset{*}{\phi}{}^+$ is given by

$$\mathscr{L} = \mathscr{L}_{\text{class}} + \int d^2\theta^+ d^2\overline{\theta}{}^+ dU\,\delta_\xi\delta_{\xi'}$$
$$\times\{\beta_1(v^{3+}\phi^+ + \overline{v}^{3+}\overset{*}{\phi}{}^+) + \beta_2\lambda^{2+}\overset{*}{\phi}{}^+\phi^+\}, \tag{3.10}$$

where $\mathscr{L}_{\text{class}}$ is given by (2.10) and β_1, β_2 are new gauge parameters.

By expanding the two last terms in (3.10) using (3.8) and (3.2), one obtains

$$\delta_\xi\delta_{\xi'}[v^{3+}\phi^+ + \overline{v}^{3+}\overset{*}{\phi}{}^+]$$
$$= [v^{3+}(ib + C'C)\phi^+ + \overline{v}^{3+}\overset{*}{\phi}{}^+(ib - CC')], \tag{3.11}$$
$$\delta_\xi\delta_{\xi'}[\lambda^{2+}\overset{*}{\phi}{}^+\phi^+]$$
$$= [2\lambda^{2+}\overset{*}{\phi}{}^+ib\phi^+ + 2\lambda^{2+}\overset{*}{\phi}{}^+[C',C]\phi^+].$$

On the other hand, the expression (3.1) gives

$$\tfrac{1}{2}\delta_\xi\delta_{\xi'}[V^{++}V^{++} + 2\alpha C'M^{4+}C]$$
$$= [-D^{++}bV^{++} - \alpha bM^{4+}b + D^{++}C'\mathscr{D}^{++}C$$
$$+ [V^{++},C'][V^{++},C]$$
$$-(\alpha/4)[C',C']M^{4+}[C,C]], \tag{3.12}$$

where

$$\mathscr{D}^{++}C = D^{++}C + i[V^{++},C].$$

Eliminating the auxiliary superfield by using its equation of motion, we obtain

$$\mathscr{L}_{\text{Total}} = \mathscr{L}_{GI} + \mathscr{L}_{\text{class}} + \int d^2\theta^+ d^2\overline{\theta}{}^+ dU$$
$$\times\{D^{++}C'\mathscr{D}^{++}C + (1/2\alpha)D^{++}V^{++}$$
$$\times(1/M^{4+})[D^{++}V^{++} + i\beta_1 v^{3+}\phi^+$$
$$+ 2i\beta_2\lambda^{2+}\overset{*}{\phi}{}^+\phi^+ + \text{H.c.}$$
$$+ \text{higher order in } V^{++} \text{ and } \phi^+\}. \tag{3.13}$$

We notice that the gauge condition, given by the equation of motion of the auxiliary superfield, is modified by the introduction of the matter multiplet. We obtain the 't Hooft gauge when $\beta_1 = 1$. Therefore, in such a quantization procedure, we obtain an inconvenient mixing of V^{++} and ϕ^+, which can be overcome if the internal symmetry is spontaneously broken.[13]

In the following section we derive the $N=2$ supersymmetric WI corresponding to BRS and anti-BRS symmetries.

IV. THE $N=2$ SUPERSYMMETRIC WI AND ALGEBRA OF RENORMALIZED BRS AND ANTI-BRS OPERATORS

The $N=2$ supersymmetric theory defined by the Lagrangian (2.10) and (3.1) is renormalizable by power counting. The Lagrangian (3.10) is not renormalizable because the coupling constants λ^{2+} and v^{3+} have dimensions (-2) and (-1), respectively. However, the renormalizable theory can be built by adding local counter terms in all superfields, and the $N=2$ supersymmetric WI must be satisfied at any finite order in the perturbation theory. To obtain these WI, let us start from a gauge-fixed, $N=2$ supersymmetric Lagrangian such as the one given in (3.13):

$$\mathscr{L} = \mathscr{L}_{\text{Total}}(V^{++},\ C,\ C',\ b,\ \phi^+,\ \alpha,\ \beta_1,\ \beta_2).$$

Consider also a system of external superfields ρ^{2-}, γ^-, η, and η' coupled to the composite superfields as follows:

$$I_{\text{eff}} = I_{\text{Total}} + \int d^{12}Z\,dU$$
$$\times\{\rho^{2-}\delta_\xi\delta_{\xi'}V^{++} + \gamma^-\delta_\xi\delta_{\xi'}\phi^+$$
$$+ \overline{\gamma}^-\delta_\xi\delta_{\xi'}\overset{*}{\phi}{}^+ + \eta\delta_\xi\delta_{\xi'}C' + \eta'\delta_\xi\delta_{\xi'}C\}, \tag{4.1}$$

where

$$d^{12}Z = d^4X_A\,d^2\theta^+ d^2\overline{\theta}{}^+ d^2\theta^- d^2\overline{\theta}{}^-.$$

The effective action (4.1) is BRS and anti-BRS invariant. Therefore we introduce external superfields coupled to superfields as

$$I=I_{\rm eff}+\int d^{12}Z\,dU\{J_V^{2-}V^{++}+J_\phi^-\phi^+ +\bar J_\phi^-\bar{\phi}^+ +J_cC+J_{c'}C'+J_bb\}. \tag{4.2}$$

Notice that we may maintain BRS and anti-BRS invariance by prescribing that the external superfields do not transform.

The generating functional of Green's functions is then given by

$$\exp[iW(J,\rho,\eta,\eta',\gamma)] = \int DV^{++}DCDC'D\phi^+D\bar{\phi}^+ \times\exp\Big\{iI_{\rm eff}+i\int d^{12}Z\,dU \times\{J_V^{2-}V^{++}+J_\phi^-\phi^+ +\bar J_\phi^-\bar{\phi}^+ +J_CC +J_{c'}C'+J_bb\}\Big\}. \tag{4.3}$$

We remark that we may express all composite superfields by differentiation with respect to the corresponding external superfields. In fact, the generating functional of a one-particle irreducible is defined by

$$\Gamma[V^{++},\phi^+,C,C',b,J,\rho,\eta,\eta',\gamma] = W[J,\rho,\eta,\eta',\gamma]-\int d^{12}Z\,dU\,[J_V^{2-}V^{++} +J_\phi^-\phi^+ +\bar J_\phi^-\bar{\phi}^+{}_+ +J_CC+J_{C'}C'+J_bb\,]. \tag{4.4}$$

We have

$$\frac{\delta\Gamma}{\delta V^{++}}=-J_V^{2-},\quad \frac{\delta\Gamma}{\delta\phi^+}=-J_\phi^-,\quad \frac{\delta\Gamma}{\delta\bar{\phi}^+}=-\bar J_\phi^-,$$
$$\frac{\delta\Gamma}{\delta C}=-J_C,\quad \frac{\delta\Gamma}{\delta C'}=-J_{C'},\quad \frac{\delta\Gamma}{\delta b}=-J_b, \tag{4.5}$$

and

$$\frac{\delta\Gamma}{\delta\rho^{2-}}=\delta_\xi\delta_{\xi'}V^{++},\quad \frac{\delta\Gamma}{\delta\gamma^-}=\delta_\xi\delta_{\xi'}\phi^+,$$
$$\frac{\delta\Gamma}{\delta\bar\gamma^-}=\delta_\xi\delta_{\xi'}\bar{\phi}^+,\quad \frac{\delta\Gamma}{\delta\eta}=\delta_\xi\delta_{\xi'}C',\quad \frac{\delta\Gamma}{\delta\eta'}=\delta_\xi\delta_{\xi'}C. \tag{4.6}$$

Furthermore, the Γ variation with respect to $\delta_\xi\delta_{\xi'}$ leads to WI:

$$\int d^{12}Z\,dU\Big\{\delta_\xi\delta_{\xi'}V^{++}\frac{\delta\Gamma}{\delta V^{++}}+\delta_\xi\delta_{\xi'}\phi^+\frac{\delta\Gamma}{\delta\phi^+} +\delta_\xi\delta_{\xi'}\bar{\phi}^+\frac{\delta\Gamma}{\delta\bar{\phi}^+}+\delta_\xi\delta_{\xi'}C\frac{\delta\Gamma}{\delta C}+\delta_\xi\delta_{\xi'}C'\frac{\delta\Gamma}{\delta C'}\Big\}=0. \tag{4.7}$$

Using (4.6) one obtains

$$\int d^{12}Z\,dU\Big\{\Big(\frac{\delta\Gamma}{\delta\rho^{2-}}\frac{\delta}{\delta V^{++}}+\frac{\delta\Gamma}{\delta V^{++}}\frac{\delta}{\delta\rho^{2-}}\Big) +\Big(\frac{\delta\Gamma}{\delta\gamma^-}\frac{\delta}{\delta\phi^+}+\frac{\delta\Gamma}{\delta\phi^+}\frac{\delta}{\delta\gamma^-}\Big) +\Big(\frac{\delta\Gamma}{\delta\bar\gamma^-}\frac{\delta}{\delta\bar{\phi}^+}+\frac{\delta\Gamma}{\delta\bar{\phi}^+}\frac{\delta}{\delta\bar\gamma^-}\Big)+\Big(\frac{\delta\Gamma}{\delta\eta'}\frac{\delta}{\delta C}+\frac{\delta\Gamma}{\delta C}\frac{\delta}{\delta\eta'}\Big) +\Big(\frac{\delta\Gamma}{\delta\eta}\frac{\delta}{\delta C'}+\frac{\delta\Gamma}{\delta C'}\frac{\delta}{\delta\eta}\Big)\Big\}\Gamma=0. \tag{4.8}$$

In the above section we have assumed that all parameters involved in the effective action are finite. Furthermore, the generating functional (4.3) contains infinities. We have also assumed that these infinities can be regularized by means of a method that preserves the $\delta_\xi\delta_{\xi'}$ symmetry.

The renormalized $N=2$ supersymmetric theory can now be obtained by adding to the effective action (4.1) local counter terms in the superfields and external superfields. Indeed, the renormalized action I_R leads to a finite renormalized generating functional Γ_R satisfying the WI at any finite order in the loop expansion. The I_R is then given by

$$I_R^{(\text{BRS, anti-BRS})}=\int d^4X_A\,\mathscr{L}^R(V^{++},C,C',b,\phi^+) +\int d^{12}Z\,dU\{\rho^{2-}P_1^{2+}(V^{++}) +\gamma^-P_2^+(\phi^+)+\bar\gamma^-P_2^+(\bar{\phi}^+) +\eta'P_3(C)+\eta P_3(C') +\lambda P_4(b)+\cdots\}, \tag{4.9}$$

where P_1^{2+}, P_2^+, P_3, and P_4 are unknown polynomials in the superfields, and the dots represent quadratic terms in external superfields.

We then introduce the following operator:

$$\delta_R=(\delta_\xi\delta_{\xi'})^R =\int d^{12}Z\,dU\Big\{P_1^{2+}(V^{++})\frac{\delta}{\delta V^{++}} +P_2^+(\phi^+)\frac{\delta}{\delta\phi^+}+P_2^+(\bar{\phi}^+)\frac{\delta}{\delta\bar{\phi}^+} +P_3(C)\frac{\delta}{\delta C}+P_3(C')\frac{\delta}{\delta C'}+P_4(b)\frac{\delta}{\delta b}\Big\}, \tag{4.10}$$

such that

$$(\delta_\xi\delta_{\xi'})^RV^{++}=P_1^{2+}(V^{++}),$$
$$(\delta_\xi\delta_{\xi'})^R\phi^+=P_2^+(\phi^+),\quad (\delta_\xi\delta_{\xi'})^RC=P_3(C), \tag{4.11}$$
$$(\delta_\xi\delta_{\xi'})^RC'=P_3(C'),\quad (\delta_\xi\delta_{\xi'})^Rb=P_4(b).$$

Therefore the WI (4.8) can be rewritten as

$$(\delta_\xi\delta_{\xi'})^RI_R=0. \tag{4.12}$$

Using (4.9) and (4.11) yields the following equations, which are true for any value of external superfields:

$$(\delta_\xi\delta_{\xi'})^R\mathscr{L}_R(V^{++},C,C',b,\phi^+)=0, \tag{4.13}$$
$$(\delta_\xi\delta_{\xi'})^R(\delta_\xi\delta_{\xi'})^R=0, \tag{4.14}$$

on all unknown polynomials.

Expression (4.13) shows that the renormalized $N=2$ supersymmetric Lagrangian must be invariant under a $(\delta_\xi\delta_{\xi'})^R$ symmetry that satisfies (4.14). In what follows we shall see for which conditions the nilpotency of $(\delta_\xi\delta_{\xi'})^R$ is equivalent to that of $(\delta_\xi)^R$ and $(\delta_{\xi'})^R$.

Assuming that the effective action is only BRS invariant,

$$I_{\rm eff}=I_{\rm Total}+\int d^{12}Z\,dU\{\rho_1^{2-}\delta_\xi V^{++}+\gamma_1^-\delta_\xi\phi^+ +\bar{\gamma}_1^-\delta_\xi\bar{\dot{\phi}}^+ +\eta_1\delta_\xi C'+\eta_1'\delta_\xi C\}. \tag{4.15}$$

The variation of the corresponding generating functional Γ with respect to δ_ξ symmetry leads to WI:

$$\int d^{12}Z\,dU\Big\{\Big(\frac{\delta\Gamma}{\delta\rho_1^{2-}}\frac{\delta}{\delta V^{++}}+\frac{\delta\Gamma}{\delta V^{++}}\frac{\delta}{\delta\rho_1^{2-}}\Big)+\Big(\frac{\delta\Gamma}{\delta\gamma_1^-}\frac{\delta}{\delta\phi^+}+\frac{\delta\Gamma}{\delta\phi^+}\frac{\delta}{\delta\gamma_1^-}\Big)+\Big(\frac{\delta\Gamma}{\delta\bar{\gamma}_1^-}\frac{\delta}{\delta\bar{\dot{\phi}}^+}+\frac{\delta\Gamma}{\delta\bar{\dot{\phi}}^+}\frac{\delta}{\delta\bar{\gamma}_1^-}\Big)+\Big(\frac{\delta\Gamma}{\delta\eta_1'}\frac{\delta}{\delta C}+\frac{\delta\Gamma}{\delta C}\frac{\delta}{\delta\eta_1'}\Big)+\Big(\frac{\delta\Gamma}{\delta\eta_2}\frac{\delta}{\delta C'}+\frac{\delta\Gamma}{\delta C'}\frac{\delta}{\delta\eta_2}\Big)\Big\}\Gamma=0. \tag{4.16}$$

The renormalized action is then given by

$$I_R^{\rm BRS}=\int d^4X_A\,\mathscr{L}^R(V^{++},C,C',b,\phi^+)+\int d^{12}Z\{\rho_1^{2-}q_1^{2+}(V^{++})+\gamma_1^-q_2^+(\phi^+)+\bar{\gamma}_1^-q_2^+(\bar{\dot{\phi}}^+)+\eta_1'q_3(C)+\eta_1q_3(C')+\lambda_1q_4(b)+\cdots\}. \tag{4.17}$$

Similarly, the renormalized BRS symmetry is then

$$(\delta_1)^R=(\delta_\xi)^R=\int d^{12}Z\,dU\Big\{q_1^{2+}(V^{++})\frac{\delta}{\delta V^{++}}+q_2^+(\phi^+)\frac{\delta}{\delta\phi^+}+q_2^+(\bar{\dot{\phi}}^+)\frac{\delta}{\delta\bar{\dot{\phi}}^+}+q_3(C)\frac{\delta}{\delta C}+q_3(C')\frac{\delta}{\delta C'}+q_4(b)\frac{\delta}{\delta b}\Big\}. \tag{4.18}$$

Hence

$$(\delta_\xi)^RV^{++}=q_1^{++}(V^{++}),\quad (\delta_\xi)^R\phi^+=q_2^+(\phi^+),\quad (\delta_\xi)^RC=q_3(C),\quad (\delta_\xi)^RC'=q_3(C'),\quad (\delta_\xi)^Rb=q_4(b). \tag{4.19}$$

The BRS WI (4.16) is equivalent to

$$(\delta_\xi)^RI_R^{\rm BRS}=0, \tag{4.20}$$

which implies

$$(\delta_\xi)^R\mathscr{L}_R(V^{++},C,C',b,\phi^+)=0, \tag{4.21}$$

and

$$(\delta_\xi)^R(\delta_\xi)^R=0, \tag{4.22}$$

on all unknown polynomials q_1^{2+}, q_2^+, q_3, and q_4.

Now, if we assume that the effective action is anti-BRS invariant, then

$$I_{\rm eff}=I_{\rm Total}+\int d^{12}Z\,dU\{\rho_2^{2-}\delta_{\xi'}V^{++}+\gamma_2^-\delta_{\xi'}\phi^+ +\bar{\gamma}_2^-\delta_{\xi'}\bar{\dot{\phi}}^+ +\eta_2\delta_{\xi'}C'+\eta_2'\delta_{\xi'}C+\sigma\delta_{\xi'}b\}, \tag{4.23}$$

and the WI is therefore given by

$$\int d^{12}Z\,dU\Big\{\Big(\frac{\delta\Gamma}{\delta\rho_2^{2-}}\frac{\delta}{\delta V^{++}}+\frac{\delta\Gamma}{\delta V^{++}}\frac{\delta}{\delta\rho_2^{2-}}\Big)+\Big(\frac{\delta\Gamma}{\delta\gamma_2^-}\frac{\delta}{\delta\phi^+}+\frac{\delta\Gamma}{\delta\phi^+}\frac{\delta}{\delta\gamma_2^-}\Big)+\Big(\frac{\delta\Gamma}{\delta\bar{\gamma}_2^-}\frac{\delta}{\delta\bar{\dot{\phi}}^+}+\frac{\delta\Gamma}{\delta\bar{\dot{\phi}}^+}\frac{\delta}{\delta\bar{\gamma}_2^-}\Big)+\Big(\frac{\delta\Gamma}{\delta\eta_2'}\frac{\delta}{\delta C}+\frac{\delta\Gamma}{\delta C}\frac{\delta}{\delta\eta_2'}\Big)+\Big(\frac{\delta\Gamma}{\delta\eta_2}\frac{\delta}{\delta C'}+\frac{\delta\Gamma}{\delta C'}\frac{\delta}{\delta\eta_2}\Big)+\Big(\frac{\delta\Gamma}{\delta\sigma}\frac{\delta}{\delta b}+\frac{\delta\Gamma}{\delta b}\frac{\delta}{\delta\sigma}\Big)\Big\}\Gamma=0. \tag{4.24}$$

The renormalized anti-BRS action, given in terms of unknown polynomials R_1^{2+}, R_2^+, R_3, and R_4, is

$$I_R^{\rm anti\text{-}BRS}=\int d^4x_A\,\mathscr{L}_R(V^{++},C,C',b,\phi^+)+\int d^{12}Z\,dU\{\rho_2^{2-}R_1^{2+}(V^{++})+\gamma_2^-R_2^+(\phi^+)+\bar{\gamma}_2^-R_2^+(\bar{\dot{\phi}}^+)+\eta_2'R_3(C)+\eta_2R_3(C')+\sigma R_4(b)+\cdots\}. \tag{4.25}$$

The renormalized operator $(\delta_{\xi'})^R$ is such that

$$(\delta_{\xi'})^RV^{++}=R_1^{2+}(V^{++}),\quad (\delta_{\xi'})^R\phi^+=R_2^+(\phi^+),\quad (\delta_{\xi'})^RC=R_3(C),\quad (\delta_{\xi'})^RC'=R_3(C'),\quad (\delta_{\xi'})^Rb=R_4(b). \tag{4.26}$$

As before, the anti-BRS WI (4.24) leads to

$$(\delta_{\xi'})^R\mathscr{L}_R(V^{++},C,C',b,\phi^+)=0, \tag{4.27}$$

$$(\delta_{\xi'})^R(\delta_{\xi'})^R=0. \tag{4.28}$$

Therefore, acting by $(\delta_\xi)^R$ on (4.26) one obtains

$$\begin{aligned}&(\delta_\xi)^R(\delta_{\xi'})^RV^{++}=(\delta_\xi)^RR_1^{2+}(V^{++}),\\ &(\delta_\xi)^R(\delta_{\xi'})^R\phi^+=(\delta_\xi)^RR_2^+(\phi^+),\\ &(\delta_\xi)^R(\delta_{\xi'})^RC=(\delta_\xi)^RR_3(C),\\ &(\delta_\xi)^R(\delta_{\xi'})^RC'=(\delta_\xi)^RR_3(C'),\\ &(\delta_\xi)^R(\delta_{\xi'})^Rb=(\delta_\xi)^RR_4(b).\end{aligned} \tag{4.29}$$

Now in order to satisfy the equation

$$(\delta_\xi\delta_{\xi'})^R=(\delta_\xi)^R(\delta_{\xi'})^R, \tag{4.30}$$

one must have

$$\begin{aligned}&(\delta_\xi)^RR_1^{2+}(V^{++})=P_1^{2+}(V^{++}),\\ &(\delta_\xi)^RR_2^+(\phi^+)=P_2^+(\phi^+),\\ &(\delta_\xi)^RR_3(C)=P_3(C),\\ &(\delta_\xi)^RR_3(C')=P_3(C'),\quad (\delta_\xi)^RR_4(b)=P_4(b).\end{aligned} \tag{4.31}$$

Finally, to obtain the complete BRS algebra for δ_ξ^R and $\delta_{\xi'}^R$, one imposes

$$(\delta_\xi)^R(\delta_{\xi'})^R + (\delta_{\xi'})^R(\delta_\xi)^R = 0. \tag{4.32}$$

In fact, on applying $(\delta_{\xi'})^R$ to (4.19) and identifying with (4.19), we find

$$\begin{aligned}
&(\delta_{\xi'})^R q_1^{2+}(V^{++}) + (\delta_\xi)^R R_1^{2+}(V^{++}) = 0,\\
&(\delta_{\xi'})^R q_2^{+}(\phi^{+}) + (\delta_\xi)^R R_2^{+}(\phi^{+}) = 0,\\
&(\delta_{\xi'})^R q_3(C) + (\delta_\xi)^R R_3(C) = 0,\\
&(\delta_{\xi'})^R q_3(C') + (\delta_\xi)^R R_3(C') = 0,\\
&(\delta_{\xi'})^R q_4(b) + (\delta_\xi)^R R_4(b) = 0.
\end{aligned} \tag{4.33}$$

Note that (4.31) and (4.33) are the necessary conditions on all the counterterms in order to satisfy the algebraic properties of the renormalized BRS and anti-BRS symmetries. However, (4.13), (4.21), and (4.27) are sufficient to build the renormalized $N=2$ supersymmetric Lagrangian, which contains all necessary counter terms to make the theory finite, as the most general $(\delta_\xi)^R$ and $(\delta_{\xi'})^R$ invariant.

V. CONCLUSION

In this paper we have first derived the BRS and anti-BRS transformations of the FS and HST hypermultiplets. We have seen how the matter hypermultiplet is included in the harmonic superspace quantization procedure of $N=2$ supersymmetric Yang–Mills theory, by requiring BRS and anti-BRS invariances. The part of the $N=2$ supersymmetric, BRS and anti-BRS invariant Lagrangian corresponding to a matter hypermultiplet contains dimensional coupling constants λ^{2+} and v^{3+}. This leads to a nonrenormalized ϕ^+ interaction and, consequently, to additional divergences in the theory. In (3.13) Lagrangian, one recovers the 't Hooft gauge for $\beta_1 = 1$, which is very useful when the internal symmetry is spontaneously broken. Second, we have derived the BRS and anti-BRS WI by using the functional formalism. We notice that, at this level, the effective action is invariant under $(\delta_\xi\delta_{\xi'})$ symmetry. The renormalized action I_R is then given by adding local counter terms in all superfields to the renormalized Lagrangian. Such action also satisfies the WI. Furthermore, we have proved in a very simple way that the renormalized Lagrangian must be invariant under a symmetry $(\delta_\xi\delta_{\xi'})^R$ satisfying (4.14). In order to complete the usual BRS algebra, we have constructed two effective actions, one invariant under BRS, the other invariant under anti-BRS symmetries. Then we have defined two renormalized BRS and anti-BRS operators $(\delta_\xi)^R$ and $(\delta_{\xi'})^R$, respectively.

One of the futures of this work is to construct the general $(\delta_\xi)^R$ and $(\delta_{\xi'})^R$ forms compatible with power counting, $U^c(1)$ charge, and ghost number conservations, and then to build the renormalized $N=2$ supersymmetric Lagrangian as the most general $(\delta_\xi\delta_{\xi'})^R$ or $(\delta_\xi)^R$ and $(\delta_{\xi'})^R$ invariant.

ACKNOWLEDGMENTS

The author would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste, where this work was done. She is also grateful to Professor J. Strathdee for reading the manuscript.

[1]M. Veltman, University of Michigan preprint, 1980.

[2]O. Piguet and K. Sibold, Nucl. Phys. B **247**, 484 (1984).

[3]S. Mandelstam, Nucl. Phys. B **213**, 149 (1983); S. Rajpoot, J. G. Taylor, and M. Zaimi, Phys. Lett. B **127**, 347 (1983).

[4]G. 't Hooft and M. Veltman, Nucl. Phys. B **44**, 189 (1972); J. Ashmore, Nuovo Cimento B **4**, 289 (1972).

[5]W. Siegel, Phys. Lett. **84**, 193 (1979).

[6]B. W. Lee and J. Zinn-Justin, Phys. Rev. D **5**, 3121 (1972); S. Ferrara and O. Piguet, Nucl. Phys. B **93**, 261 (1975).

[7]P. S. Howe, K. S. Stelle, and P. K. Townsend, Nucl. Phys. B **236**, 125 (1984).

[8]A. Galperin, E. Ivanov, S. Kalitzin, V. Ogievetsky, and E. Sokatchev, Class. Quant. Grav. **1**, 469 (1984).

[9]T. Lhallabi and E. H. Saidi, ICTP, Trieste, preprint IC/86/193, 1986.

[10]A. Galperin, E. Ivanov, V. Ogievetsky, and E. Sokatchev, Class. Quant. Grav. **2**, 601 (1985).

[11]N. Ohta and H. Yamaguchi, Phys. Rev. D **32**, 1954 (1985).

[12]J. Zinn-Justin, Bonn Lectures, 1974.

[13]A. Ovrut and J. Wess, Phys. Rev. D **25**, 409 (1982).

The 2ω-dimensional light-cone integrals with momentum shift

A. T. Suzuki
Instituto de Física Teórica, Universidade Estadual Paulista, Rua Pamplona, 145, 01405-São Paulo, Brazil

(Received 22 July 1987; accepted for publication 16 September 1987)

A class of light-cone integrals typical to one-loop calculations in the two-component formalism is considered. For the particular cases considered, convergence is verified though the results cannot be expressed as a finite sum of elementary functions.

I. INTRODUCTION

In the evaluation of one-loop diagrams (such as the "swordfish" diagrams) in the two-component formalism of the light-cone gauge, one finds integrals of the type

$$K(p,q)=\int \frac{dr}{r^2(r-q)^2}\,\frac{1}{(p^+ + r^+)} \qquad (1.1)$$

and

$$K^l(p,q)=\int \frac{dr}{r^2(r-q)^2}\,\frac{(p^l+r^l)}{(p^+ + r^+)},\quad l=1,2, \quad (1.2)$$

where the measure dr according to the dimensional regularization technique is defined over an analytically continued space-time of 2ω dimensions.

Right from the start one should be aware that naive shifts of integration variable are not permissible in light-cone integrals that are linearly divergent by power counting assessment.[1] Happily none of the above integrals falls into this category and for convenience I consider the shifted versions

$$\widetilde{K}(p,q)=\int \frac{dr}{(r-p)^2(r-p-q)^2}\,\frac{1}{r^+} \qquad (1.1')$$

and

$$\widetilde{K}^l(p,q)=\int \frac{dr}{(r-p)^2(r-p-q)^2}\,\frac{r^l}{r^+} \qquad (1.2')$$

instead of Eqs. (1.1) and (1.2). Here I treat the singularity at $r^+=0$ according to the prescription first suggested by Mandelstam,[2] namely,

$$1/r^+ = \lim_{\epsilon\to 0^+}[1/(r^+ + i\epsilon r^-)]\,. \qquad (1.3)$$

II. EVALUATION OF $\widetilde{K}(p,q)$

Using the standard procedure of exponentiating propagators, Eq. (1.1′) becomes

$$\widetilde{K}(p,q) = -\int_0^\infty d\alpha\, d\beta\, e^{i(\alpha p^2+\beta q^2)}\int \frac{dr}{r^+}\,e^{ix(r^2+2r\cdot R)}\,, \qquad (2.1)$$

where

$$x=\alpha+\beta, \qquad (2.2)$$

$$xR=\beta q-\alpha p. \qquad (2.3)$$

Resolving the momentum integral through the employment of the Mandelstam prescription [Eq. (1.3)], one obtains[3]

$$\widetilde{K}(p,q)=\frac{(-\pi)^\omega\Gamma(2-\omega)}{(p^+ + q^+)}\times\left((q^2)^{\omega-2}\int_0^1 dy\,\frac{y^{\omega-2}(1-y)^{\omega-2}}{(1-\sigma y)} - (\hat{q}^2)^{\omega-2}\int_0^1 dy\,\frac{(y-\xi)^{\omega-2}(y-\bar{\xi})^{\omega-2}}{(1-\sigma y)}\right), \qquad (2.4)$$

where

$$y=\alpha/(\alpha+\beta)\,, \qquad (2.5)$$

$$\sigma=q^+/(p^+ + q^+)\,, \qquad (2.6)$$

$$\hat{q}^2=q^1q^1+q^2q^2=2q^+q^- - q^2, \qquad (2.7)$$

$$\xi=\tfrac{1}{2}[(1+\nu-\rho)+\sqrt{(1+\nu-\rho)^2-4\nu}\,]\,, \qquad (2.8)$$

$$\bar{\xi}=\tfrac{1}{2}[(1+\nu-\rho)-\sqrt{(1+\nu-\rho)^2-4\nu}\,]\,, \qquad (2.9)$$

$$\nu=[2(p^+ + q^+)(p^- + q^-)]/\hat{q}^2\,, \qquad (2.10)$$

$$\rho=2p^+p^-/\hat{q}^2\,, \qquad (2.11)$$

$$p^\pm=(p^0\pm p^3)/\sqrt{2}\,. \qquad (2.12)$$

In order to carry out the y-integration, first I expand the denominator of the integrands in power series

$$\frac{1}{(1-\sigma y)}=\sum_{k=1}^{\infty}(\sigma y)^{k-1} \qquad (2.13)$$

so that now, the y-integrations can be expressed in terms of beta functions and hypergeometric functions of two variables[4]

$$\widetilde{K}(p,q)=\frac{(-\pi)^\omega\Gamma(2-\omega)}{(p^+ + q^+)}\times\sum_{k=1}^{\infty}\sigma^{k-1}[(q^2)^{\omega-2}B(k+\omega-2,\omega-1) - (\nu\hat{q}^2)^{\omega-2}B(1,k)\times F_1(k,2-\omega,2-\omega;k+1;\xi^{-1},\bar{\xi}^{-1})]\,. \qquad (2.14)$$

Isolating the $k=1$ term of the sum and employing the following functional relations for the hypergeometric functions[5]:

$$F_1(\alpha,\beta,\beta';\gamma;x,y) = F_1(\alpha+1,\beta,\beta';\gamma;x,y) - (\beta/\gamma)xF_1(\alpha+1,\beta+1,\beta';\gamma+1;x,y) - (\beta'/\gamma)yF_1(\alpha+1,\beta,\beta'+1;\gamma+1;x,y), \qquad (2.15)$$

 0022-2488/88/041032-02$02.50

$$F_1(\alpha,\beta,\beta';\alpha;x,y) = (1-x)^{-\beta}(1-y)^{-\beta'}, \tag{2.16}$$

$$F_1(\alpha,\beta,0;\gamma;x,y) = {}_2F_1(\alpha,\beta;\gamma;x), \tag{2.17}$$

$$F_1(\alpha,0,\beta';\gamma;x,y) = {}_2F_1(\alpha,\beta';\gamma;y), \tag{2.18}$$

and expanding, wherever necessary, for $w \to 2$, one gets

$$\begin{aligned}\widetilde{K}(p,q) = {} & \frac{\pi^2}{(p^+ + q^+)}\left\{\ln\left(\frac{2p^+p^-}{q^2}\right) - \xi\ln\left(\frac{\xi-1}{\xi}\right)\right. \\ & \left. - \bar{\xi}\ln\left(\frac{\bar{\xi}-1}{\bar{\xi}}\right)\right\} + \frac{\pi^2}{(p^+ + q^+)} \\ & \times \sum_{k=1}^{\infty} \frac{\sigma^k}{(k+1)}\left\{\frac{2}{(k+1)} + \ln\left(\frac{2p^+p^-}{q^2}\right)\right. \\ & + \sum_{m=1}^{k}\frac{1}{m} + \frac{{}_2F_1(1,k+2;k+3;\xi^{-1})}{(k+2)\xi} \\ & \left. + \frac{{}_2F_1(1,k+2;k+3;\bar{\xi}^{-1})}{(k+2)\bar{\xi}}\right\} + O(2-\omega).\end{aligned} \tag{2.19}$$

Furthermore, using the expansion[4]

$$\begin{aligned}& \frac{{}_2F_1(1,k+2;k+3;z)}{(k+2)} \\ & = -\frac{\ln(1-z)}{z^{(k+2)}} - \sum_{m=0}^{k}\frac{z^{m-k-1}}{(1+m)},\end{aligned} \tag{2.20}$$

the final expression for $\widetilde{K}$ is written as

$$\begin{aligned}\widetilde{K}(p,q) = {} & \frac{\pi^2}{(p^+ + q^+)}\left\{\ln\left(\frac{2p^+p^-}{q^2}\right) - \xi\ln\left(\frac{\xi-1}{\xi}\right)\right. \\ & \left. - \bar{\xi}\ln\left(\frac{\bar{\xi}-1}{\bar{\xi}}\right)\right\} + \frac{\pi^2}{(p^+ + q^+)}\sum_{k=1}^{\infty}\frac{\sigma^k}{(k+1)} \\ & \times\left\{\frac{2}{(k+1)} + \ln\left(\frac{2p^+p^-}{q^2}\right)\right. \\ & - \xi^{k+1}\ln\left(\frac{\xi-1}{\xi}\right) - \bar{\xi}^{k+1}\ln\left(\frac{\bar{\xi}-1}{\bar{\xi}}\right) \\ & \left. + \sum_{m=1}^{k}\frac{1}{m} - \sum_{m=0}^{k}\frac{(\xi^{k-m} + \bar{\xi}^{k-m})}{(1+m)}\right\} \\ & + O(2-\omega).\end{aligned} \tag{2.21}$$

III. EVALUATION OF $\widetilde{K}^l(p,q)$

The evaluation of $\widetilde{K}^l(p,q)$ follows in a completely analogous way. After the y-integrations are performed the result is

$$\begin{aligned}\widetilde{K}^l(p,q) = {} & (p^l + q^l)\widetilde{K}(p,q) - \pi^2\frac{q^l}{(p^+ + q^+)}\Gamma(2-\omega) \\ & \times\sum_{k=1}^{\infty}\sigma^{k-1}[(-\pi q^2)^{\omega-2} \\ & \times B(k+\omega-1,\omega-1) \\ & - (-\pi v\hat{q}^2)^{\omega-2}B(1,k+1) \\ & \times F_1(k+1,2-\omega,2-\omega;k+2;\xi^{-1},\bar{\xi}^{-1})],\end{aligned} \tag{3.1}$$

which yields

$$\begin{aligned}\widetilde{K}^l(p,q) = {} & (p^l + q^l)\widetilde{K}(p,q) - \pi^2\frac{q^l}{(p^+ + q^+)} \\ & \times\sum_{k=1}^{\infty}\frac{\sigma^{k-1}}{(k+1)}\left\{\frac{2}{(k+1)} + \ln\left(\frac{2p^+p^-}{q^2}\right)\right. \\ & - \xi^{k+1}\ln\left(\frac{\xi-1}{\xi}\right) - \bar{\xi}\ln\left(\frac{\bar{\xi}-1}{\bar{\xi}}\right) \\ & \left. + \sum_{m=1}^{k}\frac{1}{m} - \sum_{m=0}^{k}\frac{(\xi^{k-m} + \bar{\xi}^{k-m})}{(1+m)}\right\} \\ & + O(2-\omega).\end{aligned} \tag{3.2}$$

IV. CONCLUDING REMARKS

The explicit calculation showed that the integral defined in Eq. (1.2) has its pole part canceled out. It suggests that in the Mandelstam prescription the transversal components of the vector momentum, r^l, over the longitudinal component, r^+, yield $n < 0$ as far as power counting is concerned. There remains to be seen whether the lower limit for n can be determined accurately. In any case, it is interesting to note that this pattern of convergence for momentum integrals in the light-cone gauge *à la* Mandelstam is characteristic of this gauge.

ACKNOWLEDGMENT

This work was supported by FAPESP-São Paulo, under Contract No. Proc. 86/2.376-9.

[1]R. B. Mann and L. V. Tarasov, Mod. Phys. Lett. A **1**, 525 (1986).
[2]S. Mandelstam, Nucl. Phys. B **213**, 149 (1983).
[3]D. M. Capper, J. J. Dulwich, and M. J. Litvak, Nucl. Phys. B **241**, 463 (1984).
[4]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1980).
[5]P. Appell and J. Kampé de Fériet, *Fonctions Hypergéométriques, et Hypersphériques—Polynomes D'Hermite* (Gauthier-Villars, Paris, 1926).

On the Gribov ambiguity in the Polyakov string

Z. Jaskolski[a)]
International School for Advanced Studies (ISAS), Trieste, Italy

(Received 5 June 1987; accepted for publication 18 November 1987)

The global aspects of the gauge fixing in the Polyakov path integral for the bosonic string are considered within the Ebin–Fischer–Mareden approach to the geometry of spaces of Riemannian metrics and conformal structures. It is shown that for surfaces of higher genus, the existence of local conformal gauges is sufficient to derive the globally defined integral over the Teichmüller space. The generalized Faddeev–Popov procedure for incomplete gauges is formulated and used to derive the global expression for the Polyakov path integral in the cases of torus and sphere. The Gribov ambiguity in the functional integral over surfaces without boundary can be successfully overcome for arbitrary genus.

I. INTRODUCTION

The Faddeev–Popov procedure applied to the Polyakov path integral for the bosonic string[1] is the basic tool for obtaining the expression for multiloop amplitudes in the form of a finite-dimensional integral over moduli space.[2–6] It also gives the complete expression for the integrand in terms of functional determinants, which is the starting point of recent investigations of the structure of multiloop amplitudes.[7,8] It turns out that an essential simplification of the bosonic string integrand is possible, if one takes into account the natural complex structure of the moduli space. The factorization theorem[9–11] asserts that the integrand is the square of a holomorphic form on the moduli space. Moreover, there is only one (up to constant) holomorphic volume form on the moduli space. All of this provides a very powerful framework to analyze string perturbation theory.

The first step on the long way from the Polyakov path integral to the unique holomorphic form on the moduli space is the Faddeev–Popov procedure. In the commonly used prescription,[2–6] the functional measure is treated as a formal volume form related to some Riemannian structure on the space of fields. The Faddeev–Popov determinant appears as a Jacobian factor due to the change of variables. Such an approach is essentially local, and it is not obvious that it is valid globally. The situation is more complicated in the case of a functional integral over surfaces of genus $h = 1,0$. One uses a mixture of global (dividing by the volume of the conformal group) and local (modified Jacobian factor) arguments.[2–6]

In order to investigate the global geometrical aspects of the Faddeev–Popov procedure in our previous paper, a slightly modified prescription has been introduced.[12] Following the ideas of Singer,[13] Schwartz,[14] and Babelon and Viallet[15] in the case of Yang–Mills theories, the functional measure is treated as a Riemannian volume form, but instead of the change of variables, the appropriate formal generalization of the Fubini theorem on the principal fiber bundle is used. Roughly speaking, the basic idea of this approach is to pass from functional measure (which is not well-defined) to the infinite-dimensional Riemannian geometry, which has a precise mathematical meaning. The Fubini theorem for finite-dimensional principal bundles has the form of the equality of two integrals, each of which is uniquely determined by the specific set of geometrical data. In the infinite-dimensional case such integrals have no meaning, but one can still consider the corresponding sets of geometrical objects. This geometrical transcription of the Faddeev–Popov procedure has been discussed in Ref. 12 for the Yang–Mills theory and for the point particle. The case of the bosonic string for higher genus ($h > 1$) has also been considered under the simplified assumption that the global smooth gauge fixing exists.

Global aspects of the gauge fixing in the Polyakov path integral for the bosonic string were previously considered by Killingback.[16] It was shown that the fibration

$$\begin{array}{ccc} \mathscr{D}_h^0 \to & \bar{\mathscr{M}}_h & \\ & \downarrow & \\ & \bar{\mathscr{M}}_h/\mathscr{D}_h^0 & \end{array} \qquad (1.1)$$

is nontrivial for genus $h = 1,0$ and topologically trivial in other cases. ($\mathscr{D}_h^0$ denotes the identity component of the diffeomorphism manifold group of genus h, and $\bar{\mathscr{M}}_h$ denotes the space of Riemannian metrics with a trivial isometry group.) More recently, the topology of this fibration has been discussed in the context of the Yang–Mills theory.[17] As follows from the fiber bundle description of the Teichmüller theory developed by Earle and Eells,[18] the fibration (1.1) is topologically reducible to the principal fiber bundle with the structure group $SL(2C)/Z_2$ in the case of a sphere and $SO(2)\times SO(2)$ in the case of a torus. These reductions can be used as global generalized gauge fixings (incomplete gauges). In the Yang–Mills theory the situation is worse—there are no reductions with a finite-dimensional fiber.[17]

The aim of the present paper is to justify, from the global point of view, our previous[12] geometrical interpretation of the Faddeev–Popov procedure and extend it to the cases of sphere and torus. It requires a more detailed knowledge of the geometry of fibration (1.1) than the topological observations mentioned above. We shall refer to the Ebin–Fischer–Marsden approach[19–23] to the geometry of the space of Riemannian metrics and of conformal structures on manifolds

---

[a)] On leave of absence from Institute of Theoretical Physics, University of Wrocław, Wrocław, Poland.

 0022-2488/88/041034-10$02.50

of arbitrary dimension. Taking into account the peculiarities of two dimensions and connections with the Teichmüller theory,[18] the appropriate theorems are formulated in Sec. II. In Sec. III, using local conformal gauges, the global gauge-independent form of the integral over Teichmüller space is derived. The underlying geometry of the d'Hoker–Phong[4] approach is also discussed. In Sec. IV it is shown that in the case of a torus, the fibration (1.1) has smooth reduction, and the reduced bundle is explicitly constructed. The global geometrical interpretation of the Faddeev–Popov procedure for generalized (incomplete) gauges is achieved by the formal generalization of the Fubini theorem for finite-dimensional reducible principal fiber bundles, which is formulated in the Appendix. In Sec. V the case of a sphere is discussed along similar lines. The results of Sec. III–V complete our geometrical interpretation of the Faddeev–Popov procedure for the Polyakov path integral over closed surfaces, and provide the global geometrical justification for the results derived by the "change of variables."

II. INFINITE-DIMENSIONAL GEOMETRY

Let us briefly describe some basic results of the geometry of the space of Riemannian metrics and conformal structures on two-dimensional manifolds in the form adapted to our further considerations. These results will be formulated in terms of Fréchet manifolds and Fréchet–Lie groups.[24,25] The structure of a Fréchet manifold naturally arises if one considers a space of smooth (C^∞) sections of some fiber bundle. The topology of uniform convergence of all derivatives is most natural in this case. Then the structure of a Fréchet manifold can be introduced by the "vector bundle neighborhoods" construction.[24,26] However, for Fréchet manifolds the implicit function theorem and the inverse function theorem break down.[24] To overcome these difficulties one commonly regards the Fréchet manifold as the inverse limit of the family of Hilbert manifolds modeled on appropriate Sobolev spaces H^s (s denotes a Sobolev class).[27] Such construction is possible for a large class of Fréchet manifolds appearing in physical applications.[19–23] It turns out that many theorems have easier proofs in the case of Hilbert manifolds for arbitrary s (both theorems mentioned above work in this case). If the s dependence is sufficiently regular, then it is possible to derive the information about the smooth case.[19–23] The results listed below are derived within this framework, as a consequence of the Ebin slice theorem.[19] For all details and proofs we refer to the original papers.

Let M_h be a compact orientable smooth two-dimensional manifold without boundary, of genus h. Let $\mathscr{M}_h$ denote the space of C^∞ Riemannian metrics on M_h. So $\mathscr{M}_h$ is a Fréchet manifold modeled on the Fréchet space $C^\infty(S_2) = C^\infty(T^*M_h \otimes_s T^*M_h)$ of smooth symmetric tensor fields of the type (0,2) on M_h with the C^∞ topology (i.e., the topology of uniform convergence of all derivatives on compact subsets of M_h). Let $\mathscr{D}_h$ denote the group of smooth diffeomorphisms of M_h with C^∞ topology. It is the Fréchet–Lie group[25] under the operation of composition of mappings. The Fréchet–Lie group is a smooth Fréchet manifold with a group structure such that the multiplication map and the inverse map are smooth. In fact, $\mathscr{D}_h$ has the finer structure of the inverse limit Hilbert (ILH)–Lie group.[27,28] As a Fréchet manifold, $\mathscr{D}_h$ is modeled on the space $C^\infty(TM_h)$ of smooth vector fields. The space tangent to $\mathscr{D}_h$ at $f\in\mathscr{D}_h$ is naturally isomorphic to the space

$$C^\infty(f^*TM_h) = \{\delta f\colon M_h \to TM_h,\ \delta f \text{ is smooth and } \pi\circ\delta f = f\},$$

where $\pi\colon TM_h \to M_h$ is the bundle projection. We denote by $\mathscr{W}_h = C^\infty(M_h)$ the Fréchet space of smooth real functions on M_h with C^∞ topology. With the pointwise addition of functions it is, of course, the Abelian Fréchet–Lie group. We introduce the semidirect product $\mathscr{D}_h \odot \mathscr{W}_h$ as a space $\mathscr{D}_h \times \mathscr{W}_h$ with the product Fréchet manifold structure and with the group operation defined by

$$(f',\varphi')\cdot(f,\varphi) \equiv (f'\circ f, \varphi + \varphi'\circ f).$$

So $\mathscr{D}_h \odot \mathscr{W}_h$ is a Fréchet–Lie group.[23] As a Fréchet manifold it is modeled on the Fréchet space $C^\infty(TM_h) \times C^\infty(M_h)$.

We have the following natural right actions of the groups listed above, on the space of matrices $\mathscr{M}_h$:

$$A^D\colon \mathscr{M}_h \times \mathscr{D}_h \ni (g,f) \to f^*g \in \mathscr{M}_h, \tag{2.1}$$

$$A^W\colon \mathscr{M}_h \times \mathscr{W}_h \ni (g,\varphi) \to e^\varphi g \in \mathscr{M}_h, \tag{2.2}$$

$$A^{DW}\colon \mathscr{M}_h \times (\mathscr{D}_h \odot \mathscr{W}_h) \ni (g,(f,\varphi)) \to e^\varphi f^* g \in \mathscr{M}_h. \tag{2.3}$$

All of these actions are smooth.[19,23] We will briefly describe the geometry of these actions. The actions A^D and A^{DW} are not free. We define the isotropy group I^g_h of g in $\mathscr{D}_h$,

$$I^g_h \equiv \{f\in\mathscr{D}_h\colon f^*g = g\},$$

and the isotropy group C^g_h of g in $\mathscr{D}_h \odot \mathscr{W}_h$,

$$C^g_h \equiv \{(f,\varphi)\in\mathscr{D}_h\odot\mathscr{W}_h\colon e^\varphi f^*g = g\}.$$

Clearly C^g_h is isomorphic to the conformal group of g:

$$\{f\in\mathscr{D}_h\colon f^*g = e^\varphi g,\ \varphi\in\mathscr{W}_h\}.$$

For this reason we will use the symbol C^g_h for both groups. With this identification, I^g_h is a subgroup of C^g_h.

In order to obtain the free $\mathscr{D}_h$ action let us restrict our attention to the space $\bar{\mathscr{M}}_h$ of metrics with trivial isotropy groups:

$$\bar{\mathscr{M}}_h \equiv \{g\in\mathscr{M}_h\colon I^g_h = \{\mathrm{id}\}\}.$$

This space is open and dense in $\mathscr{M}_h$.[19] Moreover, from the Ebin slice theorem there follows this proposition.

Proposition 1[19,20]: The restriction of the action A^D to the submanifold $\bar{\mathscr{M}}_h \subset \mathscr{M}_h$ defines a smooth principal fiber bundle

$$\begin{array}{c} \mathscr{D}_h \to \bar{\mathscr{M}}_h \\ \downarrow \\ \bar{\mathscr{M}}_h/\mathscr{D}_h. \end{array} \tag{2.4}$$

The topology of (2.4) is described by the following proposition.

Proposition 2: For all h, the principal fiber bundle (2.4) is nontrivial.

Proof: The space $\mathscr{M}_h$ is convex and hence contractible. The space $\bar{\mathscr{M}}_h$ is also contractible [see at Singer's proof that the space of irreducible connections on a principal SU(n) bundle over a compact manifold is contractible,[13] and see also the remark that follows]. Therefore if our bundle is trivial, then

$$0 = \pi_q(\overline{\mathscr{M}}_h) = \pi_q(\overline{\mathscr{M}}/\mathscr{D}_h) \oplus \pi_q(\mathscr{D}_h),$$

and, for all q, the homotopy groups $\pi_q(\mathscr{D}_h)$ must be trivial. But it is not true for all h. In fact, for $h \geqslant 1$, $\pi_0(\mathscr{D}_h) \neq 0$, and for $h = 0$, $\pi_2(\mathscr{D}_0) = \pi_2(\mathrm{SO}(3))$.[29]

The situation improves if we consider the subgroup $\mathscr{D}_h^0 \subset \mathscr{D}_h$ of diffeomorphisms homotopic to identity. We define

$$\overline{C}_h^g \equiv C_h^g \cap \mathscr{D}_h^0, \quad \overline{I}_h^g \equiv I_h^g \cap \mathscr{D}_h^0,$$

$$\overline{\mathscr{M}}_h \equiv \{g \in \mathscr{M}_h : \overline{I}_h^g = \{\mathrm{id}\}\}.$$

Proposition 3 [30]*:* (a) For all $h > 1$ and all $g \in \mathscr{M}_h$, $\overline{C}_h^g = \{\mathrm{id}\}$.

(b) For $h = 1$ and all $g \in \mathscr{M}_1$, $\overline{C}_1^g \approx \mathrm{SO}(2) \times \mathrm{SO}(2)$.

(c) For $h = 0$ and all $g \in \mathscr{M}_0$, $\overline{C}_0^g \approx \mathrm{SL}(2C)/Z_2$.

As follows from Proposition 3(a), in the case of higher genus we have $\overline{\mathscr{M}}_h = \mathscr{M}_h$.

Note that $\mathscr{D}_h^0$ is a closed subgroup, locally diffeomorphic to $\mathscr{D}_h$. Ebin's derivation[19] of the slice theorem can be immediately applied to the restricted action. This yields the following.

Proposition 4 [19]*:* The space $\overline{\mathscr{M}}_h$ is an open and dense subset in $\mathscr{M}_h$. The restriction of the action A^D to $\overline{\mathscr{M}}_h \times \mathscr{D}_h^0$ defines the smooth principal fiber bundle

$$\begin{array}{ccc} \mathscr{D}_h^0 \to & \overline{\mathscr{M}}_h & \\ & \downarrow & \\ & \overline{\mathscr{M}}_h/\mathscr{D}_h^0. & \end{array} \tag{2.5}$$

We have also the result about triviality derived by standard topological arguments by Killingbeck[16] and slightly extended by Solov'ev.[17]

Proposition 5 [16, 17]*:* For $h > 1$, the bundle (2.5) is topologically trivial. In other cases it is nontrivial but topologically reducible to the $\mathrm{SO}(2) \times \mathrm{SO}(2)$ bundle for $h = 1$ and to the SO(3) bundle for $h = 0$.

The sharpened version of this result in the smooth case will be derived in Secs. IV and V. Now let us consider the action of $\mathscr{W}_h$ on $\mathscr{M}_h$. This action is free, and we have the following result of Fischer and Marsden.

Proposition 6 [23]*:* The action $A^W: \mathscr{M}_h \times \mathscr{W}_h \to \mathscr{M}_h$ defines a smooth principal trivial $\mathscr{W}_h$ bundle,

$$\begin{array}{ccc} \mathscr{W}_h \to & \mathscr{M}_h & \\ & \downarrow & \\ & \mathscr{M}_h/\mathscr{W}_h. & \end{array} \tag{2.6}$$

Let us note that the triviality of (2.6) is stronger than the topological triviality. Proposition 6 asserts that the global smooth section of the bundle (2.6) exists.

As follows from the Proposition 6, the quotient space $\mathscr{C}_h \equiv \mathscr{M}_h/\mathscr{W}_h$ has a well-defined Fréchet manifold structure. The space $\mathscr{C}_h$ consists of smooth conformal structures on M_h.[18,23] The tangent space $\mathscr{T}_{[g]}\mathscr{C}_h$ at $[g] \in \mathscr{C}_h$ is naturally isomorphic with the space of smooth, traceless (with respect to $g' \in [g]$) symmetric tensor fields of the type (0.2). On the space $\mathscr{C}_h$ one can define the action of the group $\mathscr{D}_h^0$,[18]

$$A: \mathscr{C}_h \times \mathscr{D}_h^0 \ni ([g], f) \to [f^*g] \in \mathscr{C}_h. \tag{2.7}$$

This action is free for $h > 1$, as follows from Proposition 3. In order to obtain the free action in the cases of torus and sphere, one can proceed, following the idea of Fischer's extension theorem for superspace,[20] to the appropriate subgroups of $\mathscr{D}_h^0$. Our case is more regular. All isotropy groups are isomorphic [Proposition 3(b),(c)], so we have only one type of orbit. The suitable subgroups are

$$\mathscr{D}_1^0(x) \equiv \{f \in \mathscr{D}_1^0 : f(x) = x\} \quad (x \in M_1)$$

in the case of a torus, and

$$\mathscr{D}_0(x_1,x_2,x_3) \equiv \{f \in \mathscr{D}_0 : f(x_i) = x_i,\ i = 1,2,3\}$$

(x_1,x_2,x_3 are three distinct points on M_0) in the case of a sphere. They are closed smooth Fréchet–Lie subgroups of $\mathscr{D}_h^0$.[21]

The action (2.7) has been studied in the context of the Teichmüller theory by Earle and Eells.[18] Their results can be summarized in the following.

Proposition 7 [18]*:* (a) For $h > 1$, the action (2.7) defines a smooth, topologically trivial principal $\mathscr{D}_h^0$ bundle,

$$\begin{array}{ccc} \mathscr{D}_h^0 \to & \mathscr{C}_h & \\ & \downarrow & \\ & \mathscr{C}_h/\mathscr{D}_h^0 = \mathscr{T}_h & \end{array} \tag{2.8}$$

over the Teichmüller space $\mathscr{T}_h$ of the surface of genus h.

(b) For $h = 1$, the action (2.7) restricted to the $\mathscr{C}_1 \times \mathscr{D}_1^0(x)$ defines a smooth principal $\mathscr{D}_1^0(x)$ bundle,

$$\begin{array}{ccc} \mathscr{D}_1^0(x) \to & \mathscr{C}_1 & \\ & \downarrow & \\ & \mathscr{C}_1/\mathscr{D}_1^0(x) = \mathscr{T}_1 & \end{array} \tag{2.9}$$

over the Teichmüller space of the torus.

(c) For $h = 0$, we have a diffeomorphism

$$\mathscr{C}_0 \approx \mathscr{D}_0(x_1,x_2,x_3).$$

Proof: In the original paper[18] it was shown that (2.8) and (2.9) are topological bundles. The smooth bundle structure can be obtained by construction of smooth bundle isomorphisms of (2.8) and (2.9) onto some smooth principal fiber bundles.

(a) Our general idea is to identify $\mathscr{C}_h$ with the space $\mathscr{M}_h^{-4}$ of metrics with constant scalar curvature $R(g) = -4$. The space $\mathscr{M}_h^{-4}$ is a closed smooth submanifold of $\mathscr{M}_h$,[22] and the restriction of the action A^D to $\mathscr{M}_h^{-4} \times \mathscr{D}_h^0$ defines on $\mathscr{M}_h^{-4}$ the structure of a smooth principal $\mathscr{D}_h^0$ bundle,

$$\begin{array}{ccc} \mathscr{D}_h^0 \to & \mathscr{M}_h^{-4} & \\ & \downarrow & \\ & \mathscr{M}_h^{-4}/\mathscr{D}_h^0. & \end{array} \tag{2.10}$$

By the results of Kazdan and Warner,[31] for every metric $g \in \mathscr{M}_h$ there exists a uniquely determined function $\varphi \in C^\infty(M_h)$ for which $e^\varphi g \in \mathscr{M}_h^{-4}$. Then from the form of the Berger–Ebin splitting of $\mathscr{T}_g\mathscr{M}_h$ for metrics with constant scalar curvature[32] [see also (3.15) and (3.16) in the next section], it follows that $\mathscr{M}_h^{-4}$ defines a smooth global section of the fibration (2.6). This section is the required bundle isomorphism of (2.8) onto (2.10).

(b) The following considerations are based essentially on the results of Ref. 22. In the case of a torus, the space of metrics with zero scalar curvature coincides with the space $\mathscr{F}$ of flat metrics. Now $\mathscr{F}$ is a smooth closed submanifold of $\mathscr{M}_1$. The restriction of the action A^D (2.1) to $\mathscr{F}\times\mathscr{D}_1^0(x)$ defines on $\mathscr{F}$ the structure of a smooth principal fiber bundle,

$$\begin{array}{ccc}\mathscr{D}_1^0(x) \rightarrow & \mathscr{F} & \\ & \downarrow & \\ & \mathscr{F}/\mathscr{D}_1^0(x). & \end{array} \tag{2.11}$$

Let $\mathscr{F}_\Gamma \subset \mathscr{F}$ be the space of metrics with a fixed Levi-Civita connection Γ. By the results of Ref. 22, for every $g\in\mathscr{F}$ we have a splitting

$$\mathscr{T}_g\mathscr{F} = \mathscr{T}_g\mathscr{O}_g \oplus \mathscr{T}_g\mathscr{F}_{\Gamma(g)},$$

where $\mathscr{O}_g$ is the $\mathscr{D}_1^0$ orbit of g. Moreover, all metrics from $\mathscr{F}_\Gamma$ have the same isotropy groups $\overline{I}_1^g$. It follows that $\mathscr{F}_\Gamma$ defines the global smooth section of the bundle (2.11). Let us consider the intersection $\mathscr{F}^0\equiv\mathscr{F}\cap\mathscr{V}_1$, where

$$\mathscr{V}_1 \equiv \left\{g\in\mathscr{M}_1:\ \int_{M_1}\sqrt{g}\,d^2z = 1\right\}.$$

Now $\mathscr{V}_1$ is a closed smooth submanifold of $\mathscr{M}_1$,[19] invariant under the action of the group $\mathscr{D}_1$. The restriction of A^D to $\mathscr{F}_\Gamma^0\times\mathscr{D}_1^0(x)$ defines the principal fiber bundle

$$\begin{array}{ccc}\mathscr{D}_1^0(x) \rightarrow & \mathscr{F}^0 & \\ & \downarrow & \\ & \mathscr{F}^0/\mathscr{D}_1^0(x). & \end{array} \tag{2.12}$$

The submanifold $\mathscr{F}_\Gamma^0 \equiv \mathscr{F}_\Gamma \cap \mathscr{V}_1$ provides a smooth global section of this bundle. It is a so-called Teichmüller section (see Ref. 18 for discussion of this section in terms of covering the space of the torus). As in the case $h>1$, one can prove that $\mathscr{F}^0$ determines the smooth global section of the bundle (2.6), which is a bundle isomorphism of (2.9) onto (2.12).

(c) In the case of a sphere, $\mathscr{D}_0=\mathscr{D}_0^0$. The space of metrics with constant scalar curvature $R(g)=1$ coincides with the $\mathscr{D}_0$ orbit $\mathscr{O}_{g_s}$ of the standard metric g_s.[22] Using the Ebin technique[19] one can show that $\mathscr{O}_{g_s}$ is diffeomorphic to $\mathscr{D}_0/I_0^{g_s}\approx\mathscr{D}_0(x_1,x_2,x_3)$. On the sphere, every metric is conformally equivalent to a metric with constant scalar curvature.[31] Therefore it follows from the Berger–Ebin splitting[32] that $\mathscr{O}_{g_s}$ determines a smooth global section of the bundle (2.6), which is the required diffeomorphism.

As a simple consequence of Propositions 3, 6, and 7, we have the following (compare the Fischer–Marsden proof of the slice theorem for the action of $\mathscr{D}\odot\mathscr{W}$ on $\mathscr{M}$ for higher dimensions[23]).

Proposition 8: (a) For $h>1$, the action A^{DW} (2.3) defines the smooth, principal, topologically trivial $\mathscr{D}_h^0\odot\mathscr{W}_h$ bundle

$$\begin{array}{ccc}\mathscr{D}_h^0\odot\mathscr{W}_h \rightarrow & \mathscr{M}_h & \\ & \downarrow & \\ & \mathscr{T}_h. & \end{array} \tag{2.13}$$

(b) For $h=1$, the action A^{DW} restricted to the subgroup $\mathscr{D}_1^0(x)\odot\mathscr{W}_1$ defines the smooth principal trivial $\mathscr{D}_1^0(x)\times\mathscr{W}_1$ bundle

$$\begin{array}{ccc}\mathscr{D}_1^0(x)\odot\mathscr{W}_1 \rightarrow & \mathscr{M}_1 & \\ & \downarrow & \\ & \mathscr{T}_1. & \end{array} \tag{2.14}$$

(c) For $h=0$, we have a diffeomorphism $\mathscr{M}_0\approx\mathscr{D}_0(x_1,x_2,x_3)\times\mathscr{W}_0$.

Let us also note that the space of conformal structures $\mathscr{C}_h$ has a natural complex structure (it is a complex Fréchet manifold[18]). Taking the quotient of this structure by the canonical projection of the bundle (2.8), we obtain the natural complex analytic structure on $\mathscr{T}_h$. For these complex structures, the global holomorphic section exists for the torus only.[18] Since in this paper we are interested in the connection between the geometry of the space of metrices (rather than the space of conformal structures) and the Polyakov path integral expressed as an integral over Teichmüller space, the complex structure of $\mathscr{T}_h$ will not be discussed. It is, however, crucial for the holomorphic factorization theorem.[9–11]

III. $h>1$

In this section we will discuss the global aspects of the conformal gauge fixing in the Polyakov path integral over surfaces of higher genus. For the geometrical analysis of this path integral we refer to our previous paper.[12] Here we will start with the following form of the vacuum-to-vacuum amplitude:

$$Z_h^{26} = \int_{\mathscr{M}_h}\left(\int_{\mathscr{D}_h} d\Omega^{H^g}\right)^{-1}\left(\int_{\mathscr{W}_h} d\Omega^{W^g}\right)^{-1} \times\left(\frac{\det'\mathscr{L}_g}{\int_{M_h}\sqrt{g}\,d^2z}\right)^{-13} d\Omega^M, \tag{3.1}$$

obtained after integration over embeddings of the model world sheet manifold M_h into a 26-dimensional Euclidean target space.[2–6,12] In the above expression, $d\Omega^M, d\Omega^{H^g}, d\Omega^{W^g}$ denote the formal volume forms related to the weak Riemannian structures $M(\ |\ )$ on M_h, $H^g(\ |\ )$ on $\mathscr{D}_h^0$, and $W^g(\ |\ )$ on $\mathscr{W}_h$, respectively. They are defined as follows:

$$M_g(\ |\ ):\ C^\infty(S_2)\times C^\infty(S_2)\rightarrow\mathbb{R},$$

$$M_g(\delta g|\delta g') \equiv \int_{M_h}\sqrt{g}\,d^2z\,g^{ac}g^{bd}\delta g_{ab}\delta g'_{cd}; \tag{3.2}$$

$$W_\varphi^g(\ |\ ):\ C^\infty(M_h)\times C^\infty(M_h)\rightarrow\mathbb{R},$$

$$W_\varphi^g(\delta\varphi|\delta\varphi') \equiv \int_{M_h}\sqrt{g}\,d^2z\,\delta\varphi\delta\varphi'; \tag{3.3}$$

$$H_{\rm id}^g(\ |\ ):\ C^\infty(TM_h)\times C^\infty(TM_h)\rightarrow\mathbb{R},$$

$$H_{\rm id}^g(\delta f|\delta f') \equiv \int_{M_h}\sqrt{g}\,d^2z\,g_{ab}\delta f^a\delta f'^b. \tag{3.4}$$

The above Riemannian structures are not invariant under conformal deformation of a metric, so we can not apply the Faddeev–Popov procedure (in its standard form) to the integral (3.1) using the gauge fixing for the whole gauge group $\mathscr{D}_h\odot\mathscr{W}_h$. The integrand and the measure in (3.1) are $\mathscr{D}_h$ invariant, so according to the commonly used approach[1–6] one can fix the gauge only for the diffeomorphism group. The most convenient is the conformal gauge introduced by Polyakov.[1]

In our global geometrical formulation of the Faddeev–Popov procedure, it is necessary to have the global smooth section. Then the topological triviality of the bundle (2.5) (Proposition 5) is not sufficient. In fact, the splitting of the space tangent to the bundle, determined by the subspace tangent to the gauge fixing slice, is crucial in the construction of the Faddeev–Popov determinant.[12] Leaving aside the problem of the existence of a global smooth section of the bundle (2.5), we will show that it is sufficient to work with local smooth sections (which, by Proposition 4, always exist).

Let us choose the family $\{\bar{\sigma}_\alpha\}_{\alpha\in I}$ of local smooth sections

$$\bar{\sigma}_\alpha:\ \mathscr{T}_h \supset \bar{U}_\alpha \ni t \to \bar{g}_\alpha \in \mathscr{M}_h$$

of the bundle (2.13), such that the related family of open sets $\{\bar{U}_\alpha\}_{\alpha\in I}$ is a locally finite covering of the Teichmüller space. This is always possible since, by the Teichmüller theorem,[33] $\mathscr{T}_h$ is homeomorphic to $\mathbb{R}^{6h-6}$ and, therefore, paracompact. We introduce the partition of unity on $\mathscr{T}_h$,

$$1 = \sum_{\alpha\in I} \Phi_\alpha, \tag{3.5}$$

related to the coverning $\{\bar{U}_\alpha\}_{\alpha\in I}$. Let $\bar{\Sigma}_\alpha \equiv \bar{\sigma}_\alpha(\bar{U}_\alpha) \subset \mathscr{M}_h$. Then $\Sigma_\alpha \equiv A^W(\bar{\Sigma}_\alpha \times \mathscr{W}_h)$ is a smooth submanifold of $\mathscr{M}_h$. We define local sections of the bundle (2.5) (local conformal gauges) by

$$\begin{aligned} &\sigma_\alpha:\ \mathscr{M}_h/\mathscr{D}_h^0 \supset U_\alpha \ni u \to g_\alpha \in \mathscr{M}_h, \\ &\{g_\alpha\} = \Sigma_\alpha \cap \Pi_D^{-1}(u), \end{aligned} \tag{3.6}$$

where $U_\alpha \equiv \Pi_D(\Sigma_\alpha)$, and $\Pi_D:\ \mathscr{M}_h \to \mathscr{M}_h/\mathscr{D}_h^0$ denotes the canonical projection of the bundle (2.5). Note that, for all $g\in\Sigma_\alpha$, $\mathscr{T}_g\Sigma_\alpha \cap \ker \Pi_* = 0$, and the local section (3.6) is smooth. The family $\{U_\alpha\}_{\alpha\in I}$ is a locally finite covering of $\mathscr{M}_h/\mathscr{D}_h^0$ for which the partition of unity (3.5) can be easily extended.

For every local trivialization

$$\begin{array}{ccc} \mathscr{D}_h^0 \to & \mathscr{M}_h^\alpha \equiv \Pi_D^{-1}(U_\alpha) & \\ & \downarrow & \\ & \mathscr{M}_h^\alpha/\mathscr{D}_h^0 = U_\alpha, & \end{array}$$

the geometrical description of the Faddeev–Popov procedure given in Ref. 12 is fully justified. It yields (at the critical dimension)

$$\begin{aligned} Z_{h,\alpha}^{26} = &\int_{U_\alpha} \Phi_\alpha\, d^{6h-6}t \det M_{\bar{g}_\alpha}(\delta\psi_i|\delta\chi_j) \\ &\times\left(\frac{\det P_{\bar{g}_\alpha}^+ P_{\bar{g}_\alpha}}{\det H(P_{\bar{g}_\alpha}^+)}\right)^{1/2} \left(\frac{\det' \mathscr{L}_{\bar{g}_\alpha}}{\int_{\mathscr{M}_h}\sqrt{\bar{g}_\alpha}\, d^2z}\right)^{-13}, \end{aligned} \tag{3.7}$$

where

$$\delta\chi_j \equiv \frac{\partial}{\partial t^j}\bar{g}_\alpha - \frac{1}{2}\operatorname{Tr}\left(\frac{\partial}{\partial t^j}\bar{g}_\alpha\right)\cdot \bar{g}_\alpha.$$

The expression $\{\delta\psi_i\}_{i=1}^{6h-6}$ is an arbitrary basis of $\ker P_{\bar{g}_\alpha}^+ \subset \mathscr{T}_{\bar{g}_\alpha}\mathscr{M}_h$, and

$$(H(P_g^+))_{ij} = M_g(\delta\psi_i|\delta\psi_j).$$

In order to clarify the global geometrical meaning of the integrand of (3.7), we introduce a few natural orthogonal splittings of the tangent space $\mathscr{T}_g\mathscr{M}_h$. First, we consider the splitting determined by the action A^W (2.2):

$$\begin{aligned} &\mathscr{T}_g\mathscr{M}_h = \mathscr{H}_g \oplus \mathscr{K}_g, \\ &\mathscr{K}_g \equiv \{\delta g\in C^\infty(S_2):\ \delta g_{ab} = \delta\varphi g_{ab},\ \delta\varphi\in C^\infty(M_h)\}, \\ &\mathscr{H}_g \equiv \{\delta g\in C^\infty(S_2):\ g^{ab}\delta g_{ab} = 0\}. \end{aligned} \tag{3.8}$$

This decomposition is globally integrable.[23] Here $\mathscr{K}_g$ is the subspace tangent to the $\mathscr{W}_h$ orbit of g, and $\mathscr{H}_g$ is tangent to the submanifold

$$\mathscr{N}_{\mu(g)} \equiv \{g'\in\mathscr{M}_h:\ \mu(g') = \mu(g)\}$$

[$\mu(g)$ denotes the volume form related to g]. Introducing the conformal Lie derivative

$$\begin{aligned} &P_g:\ C^\infty(TM_h) \to C^\infty(S_2), \\ &(P_g\delta f)_{ab} \equiv (g_{ac}\nabla_b + g_{bc}\nabla_a - g_{ab}\nabla_c)\delta f^c, \\ &P_g^+:\ C^\infty(S_2) \to C^\infty(TM_h), \\ &(P_g^+\delta g)^c \equiv -g^{ca}\nabla^b\delta g_{ab}, \end{aligned} \tag{3.9}$$

we have the splitting[23]

$$\mathscr{H}_g = \operatorname{Im} P_g \oplus \ker P_g^+,$$

and the York decomposition[34]

$$\mathscr{T}_g\mathscr{M}_h = \operatorname{Im} P_g \oplus \ker P_g^+ \oplus \mathscr{K}_g. \tag{3.10}$$

This decomposition can be obtained by the intersection of the splitting (3.8) with the following one related to the action A^{DW} (2.3)[23]:

$$\begin{aligned} &\mathscr{T}_g\mathscr{M}_h = \operatorname{Im}\tau_g \oplus \ker\tau_g^+, \\ &\tau_g:\ C^\infty(TM_h)\times C^\infty(M_h) \to C^\infty(S_2), \\ &\tau_g(\delta f,\delta\varphi)_{ab} \equiv (g_{ac}\nabla_b + g_{bc}\nabla_a)\delta f^c + g_{ab}\delta\varphi, \\ &\tau_g^+:\ C^\infty(S_2) \to C^\infty(TM_h)\times C^\infty(M_h), \\ &\tau_g^+\delta g \equiv (-g^{ac}\nabla^b\delta g_{ab},\ g^{ab}\delta g_{ab}). \end{aligned} \tag{3.11}$$

Note that $\ker P_g^+ = \ker\tau_g^+$. Let $\Pi_g^{P^+}:\ \mathscr{T}_g\mathscr{M}_h \to \ker P_g^+$ be the projection operator related to the decomposition (3.11). We have

$$M_g(\delta\psi_i|\delta\chi_j) = M_g\left(\delta\psi_i\Big|\Pi_g^{P^+}\left(\frac{\partial}{\partial t^j}\bar{g}_\alpha\right)\right) = M_g(\delta\psi_i|\tilde{\partial}_j^g).$$

Here $\tilde{\partial}_j^g \equiv \Pi_g^{P^+}\circ\bar{\sigma}_{\alpha*}(\partial_j)$ denotes the horizontal lift of the vector $\partial_j \equiv \partial/\partial t^j$ tangent to the Teichmüller space at $t\in\mathscr{T}_h$, i.e., the unique solution of the conditions

$$\Pi_{DW*}\tilde{\partial}_j^g = \partial_j,\quad \tilde{\partial}_j^g \in \ker P_g^+ = (\mathscr{T}_g\Pi_{DW}^{-1}(g))^\perp,$$

where $g\in\Pi_{DW}^{-1}(t)$, and $\Pi_{DW}:\ \mathscr{M}_h \to \mathscr{T}_h$ denotes the canonical projection of the bundle (2.13). The whole integrand of (3.7) is independent of the basis $\{\delta\psi_i\}_{i=1}^{6h-6}$ and $\mathscr{D}_h^0$ invariant. In order to show the $\mathscr{W}_h$ invariance, it must be decomposed into two pieces: $\det M_g(\delta\psi_i|\tilde{\partial}_j^g)$ and

$$\left(\frac{\det P_g^+ P_g}{\det H(P_g^+)}\right)^{1/2} \left(\frac{\det' \mathscr{L}_g}{\int_{M_h}\sqrt{g}\, d^2z}\right)^{-13}. \tag{3.12}$$

The $\mathscr{W}_h$ invariance of the first term is a consequence of the relation

$\tilde{\partial}_j^{e^{\varphi}g} = e^{\varphi}\tilde{\partial}_j^g.$

The $\mathscr{W}_h$ invariance of (3.12) has been shown by Alvarez.[2] It follows that the integrand of (3.7) is a well-defined $(6h-6)$-form on the Teichmüller space independent of a special choice of local sections. So for global coordinates $\{t^i\}_{i=1}^{6h-6}$ on $\mathscr{T}_h$, we can drop out the partition of unity and arrive at the following formula:

$$Z_h^{26} = \int_{\mathscr{T}_h} d^{6h-6}t \det M_g(\delta\psi_i|\tilde{\partial}_j^g) \times \left(\frac{\det P_g^+ P_g}{\det H(P_g^+)}\right)^{1/2} \left(\frac{\det' \mathscr{L}_g}{\int_{M_h}\sqrt{g}\, d^2z}\right)^{-13}. \qquad (3.13)$$

In order to obtain a somewhat more elegant form of (3.13), one can use the freedom in the choice of bases $\{\delta\psi_i\}$ of $\ker P_g^+ \subset \mathscr{T}_g\mathscr{M}_h$. Let us consider the set of vector fields $\{\delta v_i\}_{i=1}^{6h-6}$ on $\mathscr{T}_h$ such that, at every point $t\in\mathscr{T}_h$, $\{\delta v_i(t)\}_{i=1}^{6h-6}$ is a basis of the space tangent to $\mathscr{T}_h$ at t. Then the horizontal lifts $\{\delta\tilde{v}_i\}$ of $\{\delta v_i\}$ [in the bundle (2.13) with respect to the metric $M_g(\ |\ )$ (3.2)] form a basis of $\ker P_g^+ \subset \mathscr{T}_g\mathscr{M}_h$ at every metric $g\in\mathscr{M}_h$. Choosing in (3.13) $\{\delta\psi_i\}_{i=1}^{6h-6}$ obtained in this way, and using the same symbol for the vector fields on $\mathscr{T}_h$ and their lifts on $\mathscr{M}_h$, one can rewrite (3.13) in a form independent of a special choice of coordinates on $\mathscr{T}_h$:

$$Z_h^{26} = \int_{\mathscr{T}_h} d\psi^1 \wedge \cdots \wedge d\psi^{6h-6} \times \left(\frac{\det P_g^+ P_g}{\det H(P_g^+)}\right)^{1/2} \left(\frac{\det' \mathscr{L}_g}{\int_{M_h}\sqrt{g}\, d^2z}\right)^{-13}, \qquad (3.14)$$

where $\{d\psi^i\}_{i=1}^{6h-6}$ are one-forms on $\mathscr{T}_h$ dual to the vector fields $\{\delta\psi_i\}_{i=1}^{6h-6}$. The above expressions (3.13) and (3.14) are the starting point for the holomorphic factorization.[9-11]

A slightly different geometrical interpretation of (3.7) has been done in terms of the Weil–Peterson metric on the Teichmüller space by d'Hoker and Phong.[4] The geometry underlying their approach is described by the Berger–Ebin splitting[32]:

$$\begin{aligned}
&\mathscr{T}_g\mathscr{M}_h = \ker\gamma_g \oplus \operatorname{Im}\gamma_g^+,\\
&\gamma_g\colon C^\infty(S_2)\to C^\infty(M_h),\\
&\gamma_g\delta g \equiv -\nabla_c\nabla^c g^{ab}\delta g_{ab} + \nabla^a\nabla^b\delta g_{ab} - R(g)^{ab}\delta g_{ab}, \qquad (3.15)\\
&\gamma_g^+\colon C^\infty(M_h)\to C^\infty(S_2),\\
&(\gamma_g^+ f)_{ab} \equiv -\delta g_{ab}\nabla_c\nabla^c f + \nabla_a\nabla_b f - fR(g)_{ab}.
\end{aligned}$$

The summand $\ker\gamma_g$ is the space tangent to the submanifold

$$\mathscr{M}_h^{R(g)} \equiv \{g'\in\mathscr{M}_h\colon R(g') = R(g)\}$$

$[R(g)_{ab}$ denotes the Ricci tensor of g; $R(g)\equiv g^{ab}R(g)_{ab}$ is the scalar curvature of $g]$.

For metric $\hat{g}$ with constant scalar curvature, the intersection of the decompositions (3.15) and (3.11) gives a finer splitting:

$$\mathscr{T}_{\hat{g}}\mathscr{M}_h = \operatorname{Im}\gamma_{\hat{g}}^+ \oplus \ker P_{\hat{g}}^+ \oplus \mathscr{T}_{\hat{g}}\mathscr{O}_{\hat{g}}, \qquad (3.16)$$

where $\mathscr{O}_{\hat{g}} \equiv A^D(\hat{g},\mathscr{D}_h^0)$ is the $\mathscr{D}_h^0$ orbit of $\hat{g}$. Therefore, if one restricts oneself to the submanifold $\mathscr{M}_h^{-4}\subset\mathscr{M}_h$ of metrics with constant scalar curvature $R=-4$, then that part of the integrand (3.13),

$$d^{6h-6}t\,\frac{\det M_{\hat{g}}(\delta\psi_i|\tilde{\partial}_j^{\hat{g}})}{\det H(P_{\hat{g}}^+)},$$

can be interpreted as a volume form $d\omega^{WD}$ related to the Weil–Peterson metric on $\mathscr{T}_h$. It is defined as the projection of the $\mathscr{D}_h^0$-invariant Riemannian structure $M_{\hat{g}}(\ |\ )$ (3.2) on $\mathscr{M}_h^{-4}$, onto $\mathscr{T}_h$. The remaining $\mathscr{D}_h^0$-invariant (but not $\mathscr{W}_h$-invariant) part must be evaluated on $\mathscr{M}_h^{-4}$. It is a well-defined function on the Teichmüller space. The d'Hoker–Phong expression of Z_h^{26} has the following form:

$$Z_h^{26} = \int_{\mathscr{T}_h} d\omega^{WP}(\det P_{\hat{g}}^+ P_{\hat{g}})^{1/2}\left(\frac{\det'\mathscr{L}_{\hat{g}}}{\int_{M_h}\sqrt{\hat{g}}\, d^2z}\right)^{-13}.$$

IV. $h=1$

We will start the discussion of the global aspects of the Faddeev–Popov procedure in the case of a torus with the derivation of the sharper version of Proposition 5. Let us denote by $\mathscr{F}_\Gamma^0$ the space of metrics on M_1 with fixed flat Levi-Civita connection Γ, such that $\int_{M_1}\sqrt{g}\, d^2z = 1$. Let Σ_Γ be a space obtained from $\mathscr{F}_\Gamma^0$ by the action of the group $\mathscr{W}_1$.

Proposition 9: For $h=1$, the smooth principal $\mathscr{D}_1^0$ bundle

$$\begin{array}{ccc} \mathscr{D}_1^0 \to & \overline{\mathscr{M}}_1 & \\ & \downarrow & \\ & \overline{\mathscr{M}}_1/\mathscr{D}_1^0, & \end{array} \qquad (4.1)$$

is reducible to the smooth principal $SO(2)\times SO(2)$ bundle

$$\begin{array}{ccc} C_\Gamma \to & \overline{\Sigma}_\Gamma & \\ & \downarrow & \\ & \overline{\Sigma}_\Gamma/C_\Gamma = \overline{\mathscr{M}}_1/\mathscr{D}_1^0, & \end{array} \qquad (4.2)$$

where $C_\Gamma = \overline{C}_1^g$, $g\in\Sigma_\Gamma$; $\overline{\Sigma}_\Gamma \equiv \Sigma_\Gamma \cap \overline{\mathscr{M}}_1$.

Proof: From the proof of Proposition 7 we have that $\mathscr{F}_\Gamma^0$ is a smooth closed submanifold of $\mathscr{M}_h$. Then, by Proposition 6, the orbit $\Sigma_\Gamma = A^W(\mathscr{F}_\Gamma^0,\mathscr{W}_1)$ is a smooth closed submanifold of $\mathscr{M}_1$. For $g_1,g_2\in\mathscr{F}_\Gamma^0$, the $\mathscr{D}_1$ isotropy groups $I_1^{g_1}, I_1^{g_2}$ have the same component of identity: $\overline{I}_1^{g_1} = \overline{I}_1^{g_2}$.[22] For flat metrics on the torus, $\overline{I}_1^g = \overline{C}_1^g$ and $\overline{C}_1^{e^{\varphi}g} = \overline{C}_1^g$, so all metrics from Σ_Γ have the same $\mathscr{D}_1^0\odot\mathscr{W}_1$ isotropy group, which we will denote by C_Γ. The intersection $\overline{\Sigma}_\Gamma = \Sigma_\Gamma\cap\overline{\mathscr{M}}_1$ is a closed smooth submanifold of $\overline{\mathscr{M}}_1$, and the restriction of the action $\overline{A}^D$: $\overline{\mathscr{M}}_1\times\mathscr{D}_1^0\to\overline{\mathscr{M}}_1$ to $\overline{\Sigma}_\Gamma\times C_\Gamma^0$ provides the structure of the principal smooth C_Γ bundle (4.2). By Proposition 3, $C_\Gamma \approx SO(2)\times SO(2)$. Finally, since every metric on the torus is $\mathscr{W}_1$ equivalent to the flat metric, we have $\overline{\Sigma}_\Gamma/C_\Gamma = \overline{\mathscr{M}}_1/\mathscr{D}_1^0$, and the bundle (4.2) is a reduction of (4.1).

The reduction of the bundle is a natural generalization of the gauge fixing—the incomplete gauge fixing. In fact, the global smooth section can be viewed as a reduction to the bundle with a trivial structure group $\{\mathrm{id}\}$. We will formulate the Faddeev–Popov procedure for this generalized gauge fixing. Roughly speaking, our aim is to express the functional integral over the bundle (4.1) in terms of the integral over the reduced bundle (4.2). In the finite-dimensional case we have an appropriate version of the Fubini theorem (see Appendix). According to the ideas of Ref. 12 mentioned in the Introduction, the Faddeev–Popov procedure can be seen as a formal generalization of the finite-dimensional Fubini theorem.

The infinite-dimensional counterpart of the lhs of (A8) is determined by the following set of geometrical objects: the smooth principal fiber bundle (4.1), the weak Riemannian $\mathscr{D}_1^0$-invariant structures (3.2)–(3.4), and the $\mathscr{D}_1^0$-invariant functional on $\mathscr{M}_1$ formally defined by

$$F[g]\equiv\left(\int_{\mathscr{W}_1} d\Omega^{W^g}\right)^{-1}\left(\frac{\det'\mathscr{L}_g}{\int_{M_1}\sqrt{g}\,d^2z}\right)^{-13}.$$

The infinite-dimensional counterpart of the rhs of (A8) is determined by the smooth reduction (4.2) of the bundle (4.1), by the induced Riemannian structures M^Σ on $\overline{\Sigma}_\Gamma$ and $\overline{H}^g$ on C_Γ, by $\overline{F}[g]\equiv F[g]|_{\overline{\Sigma}_\Gamma}$, and by the functional $D[g]$ on $\overline{\Sigma}_\Gamma$ (the Faddeev–Popov determinant). In order to find a form of $D[g]$, we must construct the infinite-dimensional counterparts of the operators Δ_u, Δ_u^+ defined by the formulas (A6) and (A7). Let us consider the canonical decomposition of the space tangent to $\mathscr{M}_1$ at $g\in\Sigma_\Gamma$, related to the action A^D (2.1)[19]:

$$\begin{aligned}
&\mathscr{T}_g\mathscr{M}_1=\operatorname{Im}\alpha_g\oplus\ker\alpha_g^+,\\
&\alpha_g\colon C^\infty(TM_h)\to C^\infty(S_2),\\
&(\alpha_g\delta f)_{ab}\equiv(g_{ac}\nabla_b+g_{bc}\nabla_a)\delta f^c,\\
&\alpha_g^+\colon C^\infty(S_2)\to C^\infty(TM_h),\\
&(\alpha_g^+\delta g)^c\equiv-g^{ac}\nabla^b\delta g_{ab}.
\end{aligned}\tag{4.3}$$

The summand $\operatorname{Im}\alpha_g$ is the space tangent to the $\mathscr{D}_1^0$ orbit of g.

We have also another splitting determined by the subspace $\mathscr{T}_g\Sigma_\Gamma$ tangent to Σ_Γ at $g\in\Sigma_\Gamma$:

$$\mathscr{T}_g\mathscr{M}_1=\mathscr{T}_g\Sigma_\Gamma\oplus(\mathscr{T}_g\Sigma_\Gamma)^\perp.\tag{4.4}$$

For $g\in\mathscr{F}_\Gamma^0$, we have a decomposition[22]

$$\mathscr{F}_g\Sigma_\Gamma=\ker P_g^+\oplus\mathscr{K}_g,\tag{4.5}$$

where the first summand is tangent to $\mathscr{F}_\Gamma^0$ and the second to the $\mathscr{W}_1$ orbit of g. This splitting is $\mathscr{W}_1$ invariant, so the decomposition (4.5) is valid for all $g\in\Sigma_\Gamma$. It follows that (4.4) agrees with the York splitting (3.10), and $(\mathscr{T}_g\Sigma_\Gamma)^\perp=\operatorname{Im}P_g$.

Let $\Pi_g^P\colon\mathscr{T}_g\mathscr{M}_1\to\operatorname{Im}P_g$ be the orthogonal projection related to the York splitting (3.10). The infinite-dimensional counterpart of the operator Δ_u (A6) is given by

$$\Pi_g^P\circ\alpha_g\colon\ C^\infty(TM_1)\to\operatorname{Im}P_g,$$

and coincides with the conformal Lie derivative P_g: $C^\infty(TM_1)\to\mathscr{H}_g$ (3.9). Therefore the Faddeev–Popov determinant in our case has the form

$$D[g]=(\det'P_g^+P_g)^{1/2},$$

where P_g^+ is given by (3.9), and the symbol det′ for "determinant" means that the zero eigenvalues are omitted.

Resumming the Faddeev–Popov procedure applied to the functional integral (3.1) for $h=1$ yields the following expression:

$$\begin{aligned}
Z_1^{26}=&\int_{\Sigma_\Gamma}d\Omega^\Sigma\left(\int_{\mathscr{W}_1}d\Omega^{W^g}\right)^{-1}\left(\int_{C_\Gamma}d\omega^g\right)^{-1}\\
&\times(\det'P_g^+P_g)^{1/2}\left(\frac{\det'\mathscr{L}_g}{\int_{M_1}\sqrt{g}\,d^2z}\right)^{-13},
\end{aligned}\tag{4.6}$$

where $d\Omega^\Sigma$ and $d\omega^g$ are volume forms related to the induced Riemannian structures M^Σ on Σ_Γ and $\overline{H}^g$ on C_Γ, respectively. Note that in (4.6) the submanifold $\overline{\Sigma}_\Gamma$ is replaced by Σ_Γ. This is partly justified since $\overline{\Sigma}_\Gamma$ is open and dense in Σ_Γ, so we can assume that $\Sigma_\Gamma\setminus\overline{\Sigma}_\Gamma$ has "zero measure." Moreover, $\overline{\Sigma}_\Gamma/\mathscr{W}_1=\Sigma_\Gamma/\mathscr{W}_1\approx\mathscr{T}_1$.

The next step is to perform the formal Fubini theorem in the case of the trivial fibration:

$$\begin{aligned}
\mathscr{W}_1\to\ &\Sigma_\Gamma\\
&\downarrow\Pi_\Gamma\\
&\mathscr{F}_\Gamma^0.
\end{aligned}$$

It is straightforward, since the manifolds $A^W(\mathscr{F}_\Gamma^0,\varphi)$ and $A^W(\hat{g},\mathscr{W}_1)$, for $\varphi\in\mathscr{W}_1$, $\hat{g}\in\mathscr{F}_\Gamma^0$, are orthogonal whenever they intersect, and their tangent spaces split $\mathscr{T}_g\Sigma_\Gamma$ (4.5). It leads to the expression

$$\begin{aligned}
Z_1^{26}=&\int_{\mathscr{F}_\Gamma^0}d\omega^\Gamma\int_{\Pi_\Omega^{-1}(\hat{g})}d\Omega^{\hat{g}}\\
&\times\left(\int_{\mathscr{W}_1}d\Omega^{W^g}\right)^{-1}\left(\int_{C_\Gamma}d\omega^g\right)^{-1}\\
&\times\left(\frac{\det'P_g^+P_g}{\det M_g(\delta\tilde{\psi}_i|\delta\tilde{\psi}_j)}\right)^{1/2}\left(\frac{\det'\mathscr{L}_g}{\int_{M_1}\sqrt{g}\,d^2z}\right)^{-13},
\end{aligned}$$

where $\hat{g}\in\mathscr{F}_\Gamma^0$, $g\in\Pi_\Gamma^{-1}(\hat{g})$. Since the metric induced on Σ_Γ is not $\mathscr{W}_1$ invariant, the "Jacobian factor"

$$(\det M_g(\delta\tilde{\psi}_i|\delta\tilde{\psi}_j))^{-1/2}\tag{4.7}$$

appears, where $\{\delta\tilde{\psi}_1,\delta\tilde{\psi}_2\}$ is an orthonormal basis of $\ker P_g^+$ with respect to the scalar product $M_{\hat{g}}(\ |\ )$, $\hat{g}=\Pi_\Gamma(g)$.

In order to extract the g-dependent part of the volume of C_Γ, we introduce, for an arbitrary basis $\{\delta\varphi_1,\delta\varphi_2\}$ of $\ker P_{\hat{g}}$, the matrix

$$H(P_g)_{ij}\equiv M_g(\delta\varphi_i|\delta\varphi_j),\quad g\in\Sigma_\Gamma.$$

But $\ker P_g$ forms a Lie algebra of C_Γ, so $\delta\varphi_1,\delta\varphi_2$ can be interpreted as linear independent right-invariant vector fields on C_Γ. Taking the dual basis $\{d\varphi^1,d\varphi^2\}$ of right-invariant one-forms on C_Γ, we have (for $g\in\Sigma_\Gamma$)

$$\int_{C_\Gamma}d\omega^g=(\det H(P_g))^{1/2}\times\int_{C_\Gamma}d\varphi^1\wedge d\varphi^2.$$

Note that the integral on the rhs is independent of the metric g. Changing variables by the Teichmüller section σ_Γ: $\mathscr{T}_1\to\mathscr{F}_\Gamma^0$ and by $\mathscr{W}_1\ni\varphi\to e^\varphi\hat{g}\in\Pi_\Gamma^{-1}(\hat{g})$, and proceeding as in the case of $h>1$,[12] we arrive at the d'Hoker–Phong form of Z_1^{26},[4]

$$\begin{aligned}
Z_1^{26}=&\left(\int_{C_\Gamma}d\varphi^1\wedge d\varphi^2\right)^{-1}\times\int_{\mathscr{T}_1}d\omega^{WP}\\
&\times\left(\frac{\det'P_{\hat{g}}^+P_{\hat{g}}}{\det H(P_{\hat{g}})}\right)^{1/2}\left(\frac{\det'\mathscr{L}_{\hat{g}}}{\int_{M_1}\sqrt{\hat{g}}\,d^2z}\right)^{-13},
\end{aligned}$$

where $d\omega^{WP} \equiv \sigma_\Gamma^* d\omega^\Gamma$ is a Weil–Petersson volume form on the Teichmüller space. Note that factor (4.7) equals 1 for $g = \hat{g}$.

The form of Z_1^{26} suitable for a holomorphic factorization[9–11] can be obtained as in the previous section by the appropriate choice of $\{\delta\varphi_1, \delta\varphi_2\}$, $\{\delta\psi_1, \delta\psi_2\}$:

$$Z_1^{26} = \left(\int_{C_1^g} d\varphi^1 \wedge d\varphi^2\right)^{-1} \times \int_{\mathscr{T}_1} d\psi^1 \wedge d\psi^2 \times \left(\frac{\det' P_g^+ P_g}{\det H(P_g) \det H(P_g^+)}\right)^{1/2} \left(\frac{\det' \mathscr{L}_g}{\int_{M_1} \sqrt{g}\, d^2 z}\right)^{-13}. \tag{4.8}$$

In the above formula, $d\psi^1, d\psi^2$ are dual to the linear independent vector fields $\delta\psi_1, \delta\psi_2$ on $\mathscr{T}_1$, and $H(P_g^+)$ is evaluated with respect to the horizontal lifts of $\delta\psi_1, \delta\psi_2$ at $g \in \mathscr{M}_1$. Recall that these lifts are understood in terms of fibration (2.14) and the Riemannian structure (3.2). The choice of $\{\delta\varphi_1, \delta\varphi_2\}$ is restricted by the following condition. Let $\{\delta\varphi_1, \delta\varphi_2\}$, $\{\delta\varphi_1', \delta\varphi_2'\}$ be the bases in $\ker P_g \subset \mathscr{T}_g \mathscr{M}_1$ and $\ker P_{f_* g} \subset \mathscr{T}_{f_* g} \mathscr{M}_1$, respectively. Then

$$\delta\varphi_i' = f_* \delta\varphi_i \circ f^{-1}.$$

With this choice the integrand of (4.8) is a $\mathscr{D}_1^0 \odot \mathscr{W}_1$-invariant functional on $\mathscr{M}_1$, and $\int_{C_1^g} d\varphi^1 \wedge d\varphi^2$ is g independent.

V. $h=0$

As follows from Proposition 8(c), the geometry of the space of Riemannian metrics on the sphere is very simple. In order to obtain the smooth version of Proposition 5 in this case, let us fix a certain metric $\hat{g} \in \mathscr{M}_0$. Let $\Sigma_{\hat{g}}$ be the $\mathscr{W}_0$ orbit of this metric. By Proposition 6 it is a smooth submanifold of $\mathscr{M}_0$. For every metric $g \in \Sigma_{\hat{g}}$, $C_0^g = C_0^{\hat{g}}$. Then the restriction of the action A^D to $\bar{\Sigma}_{\hat{g}} \times C_0^{\hat{g}}$, where $\bar{\Sigma}_{\hat{g}} \equiv \Sigma_{\hat{g}} \cap \bar{\mathscr{M}}_0$, determines the structure of the principal fiber bundle with the structure group $C_0^{\hat{g}}$. From the proof of Proposition 8, it follows that each metric $g \in \mathscr{M}_0$ is $\mathscr{D}_0 \odot \mathscr{W}_0$ equivalent to $\hat{g}$, so $\bar{\Sigma}_{\hat{g}} / C_0^{\hat{g}} = \mathscr{M}_0 / \mathscr{D}_0$. In consequence we have the following.

Proposition 10: For $h = 0$, the smooth principal $\mathscr{D}_0$ bundle

$$\begin{array}{l} \mathscr{D}_0 \to \bar{\mathscr{M}}_0 \\ \qquad\quad \downarrow \\ \qquad \bar{\mathscr{M}}_0 / \mathscr{D}_0 \end{array} \tag{5.1}$$

reduces to the smooth principal $SL(2C)/Z_2$ bundle

$$\begin{array}{l} C_0^{\hat{g}} \to \bar{\Sigma}_{\hat{g}} \\ \qquad\quad \downarrow \\ \qquad \bar{\Sigma}_{\hat{g}} / C_0^{\hat{g}} = \bar{\mathscr{M}}_0 / \mathscr{D}_0 . \end{array} \tag{5.2}$$

The reduced bundle (5.2) can be interpreted as a generalized (incomplete) gauge fixing. One can apply the geometrical Faddeev–Popov procedure described in the previous section. In the case of a sphere it yields

$$Z_0^{26} = \int_{\bar{\Sigma}_{\hat{g}}} d\Omega^\Sigma \left(\int_{\mathscr{W}_0} d\Omega^{W^g}\right)^{-1} \left(\int_{C_0^g} d\omega^g\right)^{-1} \times (\det' P_g^+ P_g)^{1/2} \left(\frac{\det' \mathscr{L}_g}{\int_{\mathscr{M}_0} \sqrt{g}\, d^2 z}\right)^{-13},$$

where $d\Omega^\Sigma, d\omega^g$ are volume forms related to the induced Riemannian structures M^Σ on $\Sigma_{\hat{g}}$ and $\bar{H}^g$ on C_0^g, respectively. Introducing an arbitrary basis $\{\delta\varphi_i\}_{i=1}^6$ in $\ker P_g \approx T_{\rm id} C_0^g$ and changing variables $\mathscr{W}_0 \ni \varphi \to e^\varphi \hat{g} \in \Sigma_{\hat{g}}$, we have

$$Z_0^{26} = \left(\int_{C_0^{\hat{g}}} d\varphi^1 \wedge \cdots \wedge d\varphi^6\right)^{-1} \times \left(\frac{\det' P_{\hat{g}}^+ P_{\hat{g}}}{\det H(P_{\hat{g}})}\right)^{1/2} \left(\frac{\det' \mathscr{L}_{\hat{g}}}{\int_{M_0} \sqrt{\hat{g}}\, d^2 z}\right)^{-13}.$$

Since the group $C_0^{\hat{g}} \approx SL(2C)/Z_2$ is noncompact, one can argue that Z_0^{26} vanishes.[6]

ACKNOWLEDGMENT

I would like to express my warm thanks to Professor F. Strocchi for hospitality at SISSA, where the present paper was written.

APPENDIX: THE FINITE-DIMENSIONAL INTEGRATION

In this appendix we will consider an integration of the G-invariant function on a finite-dimensional nontrivial (but reducible) principal fiber bundle.

Let $P(B,G,\pi)$ be the principal fiber bundle over B, with the compact Lie group G as a structure group and with the projection π: $P \to B \approx P/G$. The expression

$$R_a:\ P \ni p \to p \cdot a \in P \quad (a \in G)$$

denotes the right action of G on P. We assume that P is reducible to the bundle $\bar{P}(B,K,\bar{\pi})$, where K is the nontrivial subgroup of G. It means[35] that there exists a homomorphism (σ', σ'') of the bundle $\bar{P}(B,K,\bar{\pi})$ into $P(B,G,\pi)$ such that σ': $K \to G$ is an embedding, and the mapping σ'' induces the identity diffeomorphism of the base B. Identifying $\bar{P}$ with $\sigma''(\bar{P})$, one can treat $\bar{P}$ as a submanifold of P. The right action $\bar{R}_a$: $\bar{P} \to \bar{P}$ of K on $\bar{P}$ is the restriction of the right action R_a of G on P; similarly, $\bar{\pi} = \pi|_{\bar{P}}$. We consider a G-invariant Riemannian metric g on P ($\forall a \in G$: $R_a^* g = g$). This metric, if restricted to the submanifold $\bar{P}$, provides the K-invariant (induced) Riemannian structure $\bar{g}$ on $\bar{P}$. All objects considered here are assumed to be sufficiently smooth.

We will consider the integral

$$I[f] = \int_P \left(\int_G d\omega^{h_p}\right)^{-1} f\, d\omega^g,$$

where f is a G-invariant function on P, $d\omega^g$ denotes the volume form related to the metric g on P, and $d\omega^{h_p}$ is the volume form related to the right-invariant metric h_p on G. The family of Riemannian structures $\{h_p\}_{p \in P}$ is such that $f_0(p) = \int_G d\omega^{h_p}$ is a G-invariant function on P. Our aim is to express the integral $I[f]$ as an integral over the reduced bundle $\bar{P}$ with the volume form related to the metric $\bar{g}$ on $\bar{P}$. For this purpose we will use two versions of the Fubini theorem on the trivial principal bundles.[12]

For each G-invariant Riemannian structure g on $P(B,G,\pi)$, there exists a uniquely defined Riemannian structure $\tilde{g}$ on B as well as the related volume form $d\tilde{\omega}^g$ on B. Let us denote by $\tilde{\bar{g}}$ the Riemannian structure on B determined by the metric $\bar{g}$ on $\bar{P}$, and by $d\tilde{\bar{\omega}}^g$ the related volume form. Using the partition of unity on B and Theorems 1 and 3 of Ref. 12 the following relation can be derived:

$$I[f]=\int_B f\cdot(\det\tau_u^+\tau_u)^{1/2}d\tilde{\omega}^g\,, \tag{A1}$$

where $\{\tau_p\}_{p\in P}$ is the family of mappings defined by

$$\tau_p\colon\ G'=T_eG\to T_pP\,,\quad \tau_p\equiv\beta_{p*|T_eG}\,,$$

where

$$\beta_p\colon\ G\to\pi^{-1}(\pi(p))\,,\quad R_ap=p\cdot a\,.$$

The family of adjoint operators $\{\tau_p^+\}_{p\in P}$ is determined by

$$h_p(\delta a,\tau_p^+\delta p)=g_p(\tau_p\delta a,\delta p)\,,$$

where $\delta a\in T_eG$, $\delta p\in T_pP$. (Let us note that the function f is G invariant and can be seen as a function on B.) Similarly, for the reduced bundle $\bar{P}$ we have

$$\int_{\bar{P}}\left(\int_K d\omega^{\bar{h}_u}\right)^{-1}f'\,d\bar{\omega}^g=\int_B f'(\det\bar{\tau}_u^+\bar{\tau}_u)^{1/2}d\bar{\tilde{\omega}}^g\,, \tag{A2}$$

where f' is the K-invariant function on $\bar{P}$, while $d\omega^{\bar{h}_u}$ denotes the volume form on K related to the induced metric $\bar{h}_u=h_{u|TK}$. The families $\{\bar{\tau}_u\}_{u\in\bar{P}},\{\bar{\tau}_u^+\}_{u\in\bar{P}}$ are defined by

$$\bar{\tau}_u\colon\ K'=T_eK\to T_u\bar{P}\,,\quad \bar{\tau}_u\equiv\bar{\beta}_{u*|T_eK}\,,$$

where

$$\bar{\beta}_u\colon\ K\to\bar{\pi}^{-1}(\bar{\pi}(u))\,,\quad \bar{\beta}_u(c)\equiv R_cu=u\cdot c\quad(u\in\bar{P}\subset P)\,,$$

and

$$\bar{h}_u(\delta c,\bar{\tau}_u^+\delta u)=g_u(\bar{\tau}_u\delta c,\delta u)\qquad(\delta c\in K'=T_eK,\ \delta u\in T_u\bar{P})\,.$$

Now we will compare the volume forms $d\tilde{\omega}^g$ and $d\bar{\tilde{\omega}}^g$. Let us introduce the following orthogonal decompositions of the tangent spaces $T_uP,T_u\bar{P}$ at the point $u\in\bar{P}\subset P$:

$$T_uP=V_u\oplus V_u^\perp\,, \tag{A3}$$

$$W_u\equiv T_u\bar{P}=\bar{V}_u\oplus\bar{V}_u^\perp\,,\quad T_uP=W_u\oplus W_u^\perp\,, \tag{A4}$$

where V_u and $\bar{V}_u$ denote the spaces tangent to the fibers of the bundles P and $\bar{P}$, respectively. We choose two orthonormal (with respect to the metric g_u) bases $\{\delta p_i\}_{i=1}^n$ in $V_u^\perp$ and $\{\delta u_i\}_{i=1}^n$ in $\bar{V}_u^\perp$ $[n=\dim(B)]$. From the definition of the metric induced on the base manifold B,[12] it follows that the bases in $T_{\pi(u)}B$,

$$\delta q_i\equiv\pi_*\delta p_i\,,\quad \delta\bar{q}_i\equiv\pi_*\delta u_i\,,$$

are orthonormal in the metrics $\tilde{g}$ and $\bar{\tilde{g}}$, respectively. The transition matrix A between these bases can be easily evaluated:

$$\delta\bar{q}_i=\sum_{j=1}^n\tilde{g}_{\pi(u)}(\delta\bar{q}_i,\delta q_j)\delta q_j=\sum_{j=1}^n g_u(\mathscr{P}_u^{V^\perp}\delta u_i,\delta p_j)\delta q_j=\sum_{j=1}^n A_{ij}\delta q_j\,,$$

where

$$\mathscr{P}_u^{\bar{V}^\perp}\colon\bar{V}_u^\perp\to V_u^\perp\,,\quad \mathscr{P}_u^{V^\perp}\equiv\Pi_u^{V^\perp}|_{\bar{V}^\perp}\,,$$

and $\Pi_u^{V^\perp}\colon\ T_uP\to V_u^\perp$ is the projection operator related to the decomposition (A3). Passing to the dual bases $\{dq_i\}_{i=1}^n,\{d\bar{q}_i\}_{i=1}^n$, we have

$$d\tilde{\omega}^g=\bigwedge_{i=1}^n dq^i=\det A\bigwedge_{i=1}^n d\bar{q}^i=\det A\,d\bar{\tilde{\omega}}^g\,. \tag{A5}$$

Let us note that $\det A$ can be seen as a determinant of the operator $\mathscr{P}_u^{V^\perp}$ evaluated in the orthonormal bases of $V_u^\perp$ and $\bar{V}_u^\perp$. So det $\mathscr{P}_u^{V^\perp}$ defined in this way is the K-invariant function on $\bar{P}$. Now, from (A1), (A2), and (A5) we have

$$I[f]=\int_{\bar{P}}\left(\int_K d\omega^{\bar{h}_u}\right)^{-1}f(u)\frac{(\det\tau_u^+\tau_u)^{1/2}}{(\det\bar{\tau}_u^+\bar{\tau}_u)^{1/2}}\times(\det\mathscr{P}_u^{V^\perp})d\bar{\omega}^g\,.$$

In order to simplify this expression, it is convenient to introduce the following family of operators $\{\Delta_u\}_{u\in\bar{P}}$:

$$\Delta_u\colon\ G'=T_eG\to W_u^\perp\,,\quad \Delta_u\equiv\Pi_u^{W^\perp}\circ\tau_u\,, \tag{A6}$$

where $\Pi_u^{W^\perp}\colon\ T_uP\to W_u^\perp$ is the operator related to the decomposition (A4). The adjoint operators $\{\Delta_u^+\}_{u\in\bar{P}}$ are defined by

$$h_u(\delta a,\Delta_u^+\delta u)=g_u(\Delta_u\delta a,\delta p)\,, \tag{A7}$$

$\delta a\in G'=T_eG$, $\delta u\in T_uP$. By an explicit calculation one can verify that

$$\left(\frac{\det\tau_u^+\tau_u}{\det\bar{\tau}_u^+\bar{\tau}_u}\right)^{1/2}\times\det\mathscr{P}_u^{V^\perp}=(\det{}'\,\Delta_u^+\Delta_u)^{1/2}\,,$$

where the symbol det′ for "determinant" means that the zero eigenvalues are omitted. Let us summarize our discussion with the following theorem.

Theorem 1: If $\bar{P}(B,K,\bar{\pi})$ is the reduced bundle of the principal fiber bundle $P(B,G,\pi)$, then for the G-invariant metric g on P and the G-invariant function on P, there holds

$$\int_P\left(\int_G d\omega^{h_p}\right)^{-1}f\,d\omega^g=\int_{\bar{P}}\left(\int_K d\omega^{\bar{h}_u}\right)^{-1}\det{}'\,\Delta_u^+\Delta_u\,d\bar{\omega}^g\,, \tag{A8}$$

where the operators $\{\Delta_u\}_{u\in\bar{P}},\{\Delta_u^+\}_{u\in\bar{P}}$ are defined by (A6) and (A7), respectively.

In the special case of the trivial reduction provided by the global section of P, this theorem reduces to Theorem 3 of Ref. 12 for the trivial bundles (with the convention that $\int_{\{e\}}d\omega^{\bar{h}_u}=1$).

[1]A. Polyakov, Phys. Lett. B **103**, 204 (1981).
[2]O. Alvarez, Nucl. Phys. B **216**, 125 (1983).
[3]G. Gilbert, Nucl. Phys. B **277**, 102 (1986).
[4]E. D'Hoker and D. H. Phong, Nucl. Phys. B **269**, 205 (1986).
[5]J. Polchinski, Commun. Math. Phys. **104**, 37 (1986).
[6]G. Moore and P. Nelson, Nucl. Phys. B **266**, 58 (1986).
[7]E. Gava, R. Iengo, T. Jayaraman, and R. Ramachandran, Phys. Lett. B **168**, 207 (1986).
[8]M. A. Namazi and S. Rajeev, Nucl. Phys. B **277**, 332 (1986).
[9]A. A. Belavin and V. G. Knizhnik, Phys. Lett. B **168**, 201 (1986).
[10]J. B. Bost and T. Jolicoeur, Phys. Lett. B **174**, 273 (1986).
[11]R. Catenacci, M. Cornalba, M. Martinelli, and C. Reina, Phys. Lett. B **174**, 328 (1986).
[12]Z. Jaskolski, Commun. Math. Phys. **111**, 439 (1987).
[13]I. M. Singer, Commun. Math. Phys. **60**, 7 (1978).

[14]A. S. Schwarz, Commun. Math. Phys. **64**, 233 (1979).
[15]O. Babelon and C. M. Viallet, Phys. Lett. B **85**, 246 (1979).
[16]T. P. Killingback, Commun. Math. Phys. **100**, 267 (1985).
[17]M. A. Solov'ev, Pis'ma Zh. Eksp. Teor. Fiz. **44**, 366 (1986).
[18]C. J. Earle and J. Eells, J. Diff. Geom. **3**, 19 (1969).
[19]D. Ebin, *Proceedings of Symposia in Pure Mathematics*, Vol. 15 (Am. Math. Soc., Providence, RI, 1970), p. 11.
[20]A. Fischer, in *Relativity*, edited by M. Carmeli, S. Fickler, and L. Witten (Plenum, New York, 1967).
[21]D. Ebin and J. E. Marsden, Ann. Math. **92**, 102 (1970).
[22]A. Fischer and J. E. Marsden, Duke Math. J. **42**, 519 (1975).
[23]A. Fischer and J. E. Marsden, Can. J. Math. **29**, 193 (1977).
[24]R. Hamilton, Bull. Am. Math. Soc. 7, 65 (1982).
[25]J. Milnor, in *Relativity Groups and Topology*, edited by B. S. De Witt and R. Stora (Elsevier, Amsterdam, 1984).
[26]R. Palais, *Foundation of Global Non-linear Analysis* (Benjamin, New York, 1968).
[27]H. Omori, *Lecture Notes in Mathematics*, Vol. 427 (Springer, Berlin, 1974).
[28]M. Adams, T. Ratiu, and R. Schmid, in *Infinite Dimensional Groups with Applications*, edited by V. Kac (Springer, New York, 1985).
[29]W. Balcerak and B. Hajduk, Coll. Math. **45**, 1 (1981).
[30]S. Goldberg, *Curvature and Homology* (Academic, New York, 1962).
[31]J. Kazdan and F. Warner, Ann. Math. **99**, 14,203 (1974).
[32]M. Berger and D. Ebin, J. Diff. Geom. **3**, 379 (1969).
[33]W. Abikoff, *Lecture Notes in Mathematics*, Vol. 820 (Springer, Berlin, 1980).
[34]J. W. York, J. Math. Phys. **14**, 456 (1973).
[35]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963).

Conservation laws for nonlinear evolution equations

J. A. Cavalcante and K. Tenenblat
Departmento de Matemática, Instituto de Ciencias Exatas, Caixa Postal 15301, 70910 Brasilia D.F., Brazil

(Received 24 April 1986; accepted for publication 28 October 1987)

A method to derive conservation laws for evolution equations that describe pseudospherical surfaces is introduced based on a geometrical property of these surfaces. A new third-order evolution equation is obtained as a first example for a nongeneric case in the classification given by Chern and Tenenblat [Stud. Appl. Math. **74**, 1 (1986)].

I. INTRODUCTION

Chern and Tenenblat[1] introduced the notion of a differential equation for a function $u(x,t)$ that describes a pseudospherical surface (p.s.s.), and they obtained a classification for such equations of type $u_t = F(u,u_x,\ldots,\partial^k u/\partial x^k)$. These results provide a systematic procedure to obtain a linear eigenvalue problem associated to nonlinear equations of this type.

In this paper we introduce a new method to derive an infinite set of conservation laws for equations that describe a p.s.s., based on a geometrical property of these surfaces. Moreover, we exhibit a new third-order evolution equation,

$$u_t = (u_x^{-1/2})_{xx} + u_x^{3/2}\,, \tag{1}$$

which is the first example for a nongeneric case in the classification mentioned above (see Theorem 2.5 in Ref. 1). The traveling wave solutions of (1) are obtained in Sec. III D.

In Sec. II we present a general result (Theorem 2.1), which is the basis of our method, and its main consequence (Corollary 2.2). In Sec. III we apply these results to specific equations obtaining conserved densities which are of local or nonlocal type, depending on the equation being considered.

II. GENERAL RESULT

We recall[1] the definition of a differential equation that describes a p.s.s. Let M^2 be a two-dimensional differentiable manifold with coordinates (x,t). A differential equation for a real function $u(x,t)$ describes a p.s.s. if it is a necessary and sufficient condition for the existence of differentiable functions

$$f_{ij},\quad 1\leqslant i\leqslant 3,\quad 1\leqslant j\leqslant 2,$$

depending on u and its derivatives such that the one-forms

$$\omega_i = f_{i1}\,dx + f_{i2}\,dt \tag{2}$$

satisfy the structure equations of a p.s.s., i.e.,

$$d\omega_1 = \omega_3\wedge\omega_2\,,\quad dw_2 = \omega_1\wedge\omega_3\,,\quad d\omega_3 = \omega_1\wedge\omega_2\,. \tag{3}$$

As a consequence, each solution of the differential equation provides a metric on M^2, whose Gaussian curvature is constant, equal to -1. Moreover, the above definition is equivalent to saying that the differential equation for u is the integrability condition for the problem

$$dv = \Omega v\,, \tag{4}$$

where v is a vector and Ω is a traceless 2×2 matrix of one-forms given by

$$\Omega = \frac{1}{2}\begin{pmatrix} \omega_2 & \omega_1-\omega_3 \\ \omega_1+\omega_3 & -\omega_2 \end{pmatrix}. \tag{5}$$

Whenever $f_{21}=\eta$ is a parameter and the functions f_{11} and f_{31} do not depend on the parameter η, (4) is the eigenvalue problem considered by Ablowitz *et al.*,[2] as was observed by Sasaki.[3]

If M^2 is a two-dimensional Riemannian manifold with Gaussian curvature -1, then there exist orthonormal vector fields whose integral curves are geodesics and horocycles of M^2 (see Proposition 4.1 in Ref. 1). The analytic interpretation of this result for differential equations which describes a p.s.s. is contained in the following theorem which is a generalization of Proposition 4.2 in Ref. 1.

Theorem 2.1: Let f_{ij}, $1\leqslant i\leqslant 3$, $1\leqslant j\leqslant 2$, be differentiable functions of x,t such that

$$\begin{aligned} -f_{11,t}+f_{12,x} &= f_{31}f_{22}-f_{21}f_{32}\,,\\ -f_{21,t}+f_{22,x} &= f_{11}f_{32}-f_{12}f_{31}\,,\\ -f_{31,t}+f_{32,x} &= f_{11}f_{22}-f_{12}f_{21}\,. \end{aligned} \tag{6}$$

Then the following statements are valid.

(i) The following system is completely integrable for ϕ:

$$\begin{aligned} \phi_x &= f_{31}+f_{11}\sin\phi+f_{21}\cos\phi\,,\\ \phi_t &= f_{32}+f_{12}\sin\phi+f_{22}\cos\phi\,. \end{aligned} \tag{7}$$

(ii) For any solution ϕ of (7),

$$\omega = (f_{11}\cos\phi - f_{21}\sin\phi)dx + (f_{12}\cos\phi - f_{22}\sin\phi)dt \tag{8}$$

is a closed one-form.

(iii) If f_{ij} are analytic functions of a parameter η at zero, then the solutions $\phi(x,t,\eta)$ of (7) and the one-form ω are also analytical in η at zero.

Proof: Point (i) follows from the Frobenius theorem. In fact, a straightforward computation shows that (6) implies $\phi_{xt}=\phi_{tx}$.

Point (ii) is proved by showing that the systems (6) and (7) imply that the exterior differentiation of ω is zero.

As for point (iii), suppose f_{ij} are analytic functions of a parameter η. Each equation of the system (7) can be considered as an ordinary differential equation whose right-hand side is an analytic function of (ϕ,η). The solutions $\phi(x,t,\eta)$ of this equation exist as defined by (i). It follows from a theorem of ordinary differential equations (Ref. 4, p. 36), on the dependence of solutions upon parameters, that $\phi(x,t,\eta)$

 0022-2488/88/041044-06$02.50

is an analytic function of η, for η in an appropriate neighborhood of zero. Q.E.D.

Before considering the main consequence of the above result, we point out that condition (6) says that the forms ω_i given by (2) satisfy (3).

In the following corollary, we consider the functions f_{ij} to be analytic in η, and describe the solutions ϕ of (7) as a power series of η. Moreover, from (8) we obtain a sequence of closed one-forms.

In order to state the result we need to fix our notation. We suppose

$$f_{ij}(x,t,\eta) = \sum_{k=0}^{\infty} f_{ij}^{k}(x,t)\eta^k . \tag{9}$$

Then the solutions ϕ of (7) are of the form

$$\phi(x,t,\eta) = \sum_{j=0}^{\infty} \phi_j(x,t)\eta^j . \tag{10}$$

We consider the following functions of η, for fixed x,t:

$$\begin{aligned} C(\eta) &= \cos(\phi) = \cos\Big(\sum_{j=0}^{\infty} \phi_j\eta^j\Big), \\ S(\eta) &= \sin(\phi) = \sin\Big(\sum_{j=0}^{\infty} \phi_j\eta^j\Big). \end{aligned} \tag{11}$$

It follows from (11) that

$$\begin{aligned} &C(0) = \cos\phi_0, \quad S(0) = \sin\phi_0, \\ &\frac{d^kC}{d\eta^k}(0) = -(k-1)! \sum_{i=0}^{k-1} \frac{k-i}{i!} \frac{d^iS}{d\eta^i}(0)\phi_{k-i}, \\ &\frac{d^kS}{d\eta^k}(0) = (k-1)! \sum_{i=0}^{k-1} \frac{k-i}{i!} \frac{d^iC}{d\eta^i}(0)\phi_{k-i}, \end{aligned} \tag{12}$$

for $k\geqslant 1$.

Finally, we define the functions of x,t:

$$\begin{aligned} H_k^{ij} &= f_{1k}^i \frac{d^{j-i}C}{d\eta^{j-i}}(0) - f_{2k}^i \frac{d^{j-i}S}{d\eta^{j-i}}(0), \\ L_k^{ij} &= f_{1k}^i \frac{d^{j-i}S}{d\eta^{j-i}}(0) + f_{2k}^i \frac{d^{j-i}C}{d\eta^{j-i}}(0), \\ F_{1k} &= f_{3k}^1 + L_k^{11}, \\ F_{lk} &= f_{3k}^l + \sum_{r=1}^{l-1} \frac{l-r}{r!} H_k^{0r}\phi_{l-r} + \sum_{r=1}^{l} \frac{1}{(l-r)!} L_k^{rl}, \end{aligned} \tag{13}$$

where i,j,l are non-negative integers such that $j\geqslant i$, $l\geqslant 2$, and $k=1,2$. We observe that the functions H_k^{ij} and L_k^{ij} defined above depend on $\phi_0, \phi_1,\ldots,\phi_{j-i}$; and the functions F_{1k} and F_{lk} depend on ϕ_0 and $\phi_0,\ldots,\phi_{l-1}$, respectively.

As an immediate consequence of Thoerem 2.1 we obtain the following corollary.

Corollary 2.2: Let $f_{ij}(x,t,\eta)$, $1\leqslant i\leqslant 3$, $1\leqslant j\leqslant 2$, be differentiable functions of x,t, analytic at $\eta=0$, that satisfy (6). Then, with the above notation, the following statements hold.

(a) The solutions ϕ of (7) are analytic at $\eta=0$; ϕ_0 is determined by

$$\phi_{0,x} = f_{31}^0 + L_1^{00}, \quad \phi_{0,t} = f_{32}^0 + L_2^{00}; \tag{14}$$

and, for $j\geqslant 1$, ϕ_j are recursively determined by the system

$$\phi_{j,x} = H_1^{00}\phi_j + F_{j1}, \quad \phi_{j,t} = H_2^{00}\phi_j + F_{j2}. \tag{15}$$

(b) For any such solution ϕ and any integer $j\geqslant 0$,

$$\omega^j = \sum_{i=0}^{j} \frac{1}{(j-i)!}(H_1^{ij}\,dx + H_2^{ij}\,dt) \tag{16}$$

is a closed one-form.

Now we consider a nonlinear evolution equation for $u(x,t)$ which describes a p.s.s. There exist functions f_{ij}, $1\leqslant i\leqslant 3$, $1\leqslant j\leqslant 2$, which depend on $u(x,t)$ and its derivatives such that, for any solution u of the evolution equation, f_{ij} satisfy (6). Then it follows from Theorem 2.1 that (7) is completely integrable for ϕ. If we suppose f_{ij} to be analytic functions of a parameter η, then the solutions ϕ of (7) and the one-form ω, given by (8), are analytic in η. Their coefficients ϕ_j and ω^j, as functions of u, are determined by (14)–(16). Therefore the closed one-forms ω^j provide a sequence of conservation laws for the evolution equation, with conserved density and flux given, respectively, by

$$\begin{aligned} \mathscr{D}_j &= \sum_{i=0}^{j} \frac{1}{(j-i)!} H_1^{ij}, \\ \mathscr{F}_j &= -\sum_{i=0}^{j} \frac{1}{(j-i)!} H_2^{ij}, \quad j\geqslant 0. \end{aligned} \tag{17}$$

Remark 2.3: If an evolution equation is the integrability condition of Eq. (4), where Ω (hence all functions f_{ij}) is analytic with respect to η, then it follows from Corollary 2.2 that a sequence of conservation laws is obtained from the analyticity of the form ω. As we shall see in Secs. III E and III F, the sufficient condition—all f_{ij} to be analytic in η—is not necessary in order to obtain a sequence of conservation laws.

III. APPLICATIONS

We will apply the method described in the preceding section to obtain an infinite number of conserved densities for the following equations: Burgers, modified Korteweg–de Vries (MKdV), KdV, Eq. (1), sine–Gordon, and sinh–Gordon. To each equation we associate functions f_{ij} satisfying (6), as in Ref. 1.

A. Burgers equation

This equation has the form

$$u_t = u_{xx} + uu_x . \tag{18}$$

For any solution u of Burgers equation, the functions

$$\begin{aligned} &f_{11} = u/2, \quad f_{12} = u^2/4 + u_x/2, \\ &f_{21} = \eta, \quad f_{22} = \eta(u/2), \\ &f_{31} = -\eta, \quad f_{32} = -\eta(u/2), \end{aligned} \tag{19}$$

satisfy (6). Therefore, applying Corollary 2.2, we obtain a sequence of functions ϕ_j of u, which are determined by (14) and (15). It follows from (13) and (19) that we may consider

$$\phi_0 = 2\arctan\exp\Big(\frac{1}{2}\int u\,dx\Big) \tag{20}$$

and the ϕ_l are recursively defined by

$$\phi_l = e^h\Big(1 + \int F_{l1}e^{-h}\,dx\Big), \quad l\geqslant 1, \tag{21}$$

where

$$h=\int \frac{u}{2}\cos\phi_0\,dx,\quad F_{11}=-1+\cos\phi_0,$$

and

$$F_{l1}=\sum_{r=1}^{l-1}\frac{l-r}{r!l}\frac{u}{2}\frac{d^rC}{d\eta^r}(0)\phi_{l-r}+\frac{1}{(l-1)!}\frac{d^{l-1}C}{d\eta^{l-1}}(0),\quad l\geqslant 2.$$

Using (12) in the above expressions, we obtain ϕ_l in terms of u. We display only the first terms of the series:

$$\phi_0=2\arctan\exp\Big(\frac{1}{2}\int u\,dx\Big),$$

$$\phi_1=e^h\Big(1-\int\frac{2\exp(\int u\,dx)}{1+\exp(\int u\,dx)}e^{-h}\,dx\Big),$$

$$\phi_2=e^h\Big(1-\int e^{-h}\phi_1\sin\phi_0\Big(1+\frac{u}{4}\phi_1\Big)dx\Big),$$

$$\vdots\,.$$

The conserved densities of Eq. (18) are given by (17), namely

$$\frac{u}{2}\cos\phi_0,$$

$$\frac{u}{2j!}\frac{d^jC}{d\eta^j}(0)-\frac{1}{(j-1)!}\frac{d^{j-1}S}{d\eta^{j-1}}(0),\quad j\geqslant 1.$$

Using (12) in the above expression, we obtain the first terms

$$\frac{u}{2}\cos\phi_0,$$

$$\text{i.e., } \frac{u}{2}\Big(1-\exp\Big(\int u\,dx\Big)\Big)\Big(1+\exp\Big(\int u\,dx\Big)\Big)^{-1},$$

$$-\Big(1+\frac{u}{2}\phi_1\Big)\sin\phi_0,$$

$$-\frac{u}{2}\phi_2\sin\phi_0-\frac{u}{4}\phi_1^2\cos\phi_0-\phi_1\cos\phi_0,$$

$$\vdots\,,$$

where the functions ϕ_j are given by (20) and (21).

B. MKdV equation

The MKdV equation is expressed as

$$u_t=u_{xxx}+\tfrac{3}{2}u^2u_x\,. \tag{22}$$

For any solution u of the MKdV, the functions

$$f_{11}=0,\quad f_{12}=-\eta u_x\,,$$
$$f_{21}=\eta,\quad f_{22}=\tfrac{1}{2}\eta u^2+\eta^3\,, \tag{23}$$
$$f_{31}=u,\quad f_{32}=u_{xx}+\tfrac{1}{2}u^3+\eta^2u$$

satisfy (6). Applying Corollary 2.2, we obtain $\phi_j, j\geqslant 0$, defined by

$$\phi_0=\int u\,dx,$$
$$\phi_j=\frac{1}{(j-1)!}\int\frac{d^{j-1}C}{d\eta^{j-1}}(0)dx,\quad j\geqslant 1\,. \tag{24}$$

Using (12) in the above expressions we obtain ϕ_j. The first terms are

$$\phi_0=\int u\,dx,$$

$$\phi_1=\int\cos\phi_0\,dx,$$

$$\phi_2=-\int\phi_1\sin\phi_0\,dx,$$

$$\vdots\,.$$

The conserved densities are given by

$$\frac{d^{j-1}S}{d\eta^{j-1}}(0),\quad j\geqslant 1\,.$$

Using (12), we obtain the first terms

$$\sin\phi_0,$$
$$\phi_1\cos\phi_0,$$
$$2\phi_2\cos\phi_0-\phi_1^2\sin\phi_0,$$
$$\vdots\,,$$

where the ϕ_j are given by (24).

C. KdV equation

This is the equation

$$u_t=u_{xxx}+6uu_x\,. \tag{25}$$

We consider the functions of $u(x,t)$ defined by

$$f_{11}=1-u,$$
$$f_{12}=-u_{xx}-2u^2+2u+u_x\eta+(-u+1)\eta^2,$$
$$f_{21}=\eta,\quad f_{22}=-2u_x+2u\eta+\eta^3, \tag{26}$$
$$f_{31}=-1-u,$$
$$f_{32}=-u_{xx}-2u^2-2u+u_x\eta+(-u-1)\eta^2.$$

For any solution u of Eq. (25), the above functions f_{ij} satisfy (2). Applying Corollary 2.2, we have a sequence of functions ϕ_j determined by (11) and (12). It follows from (26) that (11) reduces to

$$\phi_{0,x}=-1-u+(1-u)\sin\phi_0,$$
$$\phi_{0,t}=-u_{xx}-2u^2-2u+(-u_{xx}-2u^2+2u)\sin\phi_0-2u_x\cos\phi_0, \tag{27}$$

and from (12) we obtain recursively

$$\phi_j=e^h\Big(1+\int F_{j1}e^{-h}\,dx\Big),\quad j\geqslant 1, \tag{28}$$

where

$$h=\int(1-u)\cos\phi_0\,dx$$

and

$$F_{j1}=\frac{1}{j}\sum_{i=1}^{j-1}\frac{j-i}{i!}(1-u)\frac{d^iC}{d\eta^i}(0)\phi_{j-i}+\frac{1}{(j-1)!}\frac{d^{j-1}C}{d\eta^{j-1}}(0)\,.$$

The sequence of conserved densities for KdV is given by

$(1-u)\cos\phi_0$,

$$\frac{(1-u)}{j!}\frac{d^jC}{d\eta^j}(0)-\frac{1}{(j-1)!}\frac{d^{j-1}}{d\eta^{j-1}}(0),\quad j\geqslant 1.$$

Solving the integrable system of Eq. (27), from ϕ_0 we obtain ϕ_j, $j\geqslant 1$, defined by (28).

D. Equation (1)

We now consider Eq. (1),

$$u_t=(u_x^{-1/2})_{xx}+u_x^{3/2}.$$

The functions of $u(x,t)$,

$$\begin{aligned}
&f_{11}=\eta\sinh u,\\
&f_{12}=\eta(u_x^{-(1/2)})_x\cosh u\\
&\qquad+(u_x^{(1/2)}-\eta u_x^{-(1/2)}\eta\sinh u,\\
&f_{21}=\eta,\quad f_{22}=-\eta^2u_x^{-(1/2)},\quad f_{31}=\eta\cosh u,\\
&f_{32}=\eta(u_x^{-(1/2)})_x\sinh u\\
&\qquad+(u_x^{1/2}-\eta u_x^{-(1/2)})\eta\cosh u,
\end{aligned}\tag{29}$$

satisfy (6) whenever u satisfies (1).

Since the functions defined in (29) are analytic with respect to η, we can apply Corollary 2.2. Therefore, by solving (14) and (15), we obtain

$$\phi_0=a\quad(\text{constant}),\quad \phi_j=\int F_{j1}\,dx,\quad j\geqslant 1,\tag{30}$$

where

$$F_{j1}=\cosh u\delta_{j1}+\frac{1}{(j-1)!}\times\left(\sinh u\frac{d^{j-1}S}{d\eta^{j-1}}(0)+\frac{d^{j-1}C}{d\eta^{j-1}}(0)\right).$$

We show the first elements in the series in terms of u:

$$\begin{aligned}
&\phi_0=a,\\
&\phi_1=\int(\sin a\sinh u+\cos a+\cosh u)dx,\\
&\phi_2=\int\phi_1(\cos a\sinh u-\sin a)dx,\\
&\phi_3=\int[\phi_2(\cos a\sinh u-\sin a)\\
&\qquad-\tfrac{1}{2}\phi_1^2(\sin a\sinh u+\cos a)]dx,\\
&\vdots
\end{aligned}\tag{31}$$

The conserved densities are given by

$$\sinh u\frac{d^{j-1}C}{d\eta^{j-1}}(0)-\frac{d^{j-1}S}{d\eta^{j-1}}(0),\quad j\geqslant 1.$$

Using (12), we list the first terms

$\cos a\sinh u-\sin a$,

$-\phi_1(\sin a\sinh u+\cos a)$,

$-\phi_2(\sin a\sinh u+\cos a)-\frac{1}{2}\phi_1^2(\cos a\sinh u-\sin a)$,

$\vdots$,

where $\phi_1,\phi_2,\ldots$ are given by (31).

Before going on to the next example, we point out that the traveling wave solutions of Eq. (1) are given as follows: Define

$$\psi_k(\xi)=\int_0^\xi\frac{1}{g_k^2(\tau)}d\tau,\quad k=1,2,$$

where $\xi=x-ct$, c is a positive constant, and $g_1(\tau)$ [resp. $g_2(\tau)$] is the unique positive (resp. negative) valued real function, defined for $\tau\geqslant 0$, which satisfies the equation

$$cg-\log(cg+1)=c^2\tau.$$

Then $u(x,t)=\psi_k(\xi)$ is a traveling wave solution of Eq. (1). Observe that $f_{31}^2-f_{11}^2=\eta^2$. Therefore (1) is an equation that describes a p.s.s. as in Theorem 2.5 of Ref. 1.

E. Sine–Gordon equation

This equation has the form

$$u_{xt}=\sin u.\tag{32}$$

Consider the functions defined by

$$\begin{aligned}
&f_{11}=0,\quad f_{12}=(1/\eta)\sin u,\\
&f_{21}=\eta,\quad f_{22}=(1/\eta)\cos u,\\
&f_{31}=u_x,\quad f_{32}=0.
\end{aligned}\tag{33}$$

For any solution u of Eq. (32), the above functions satisfy (6). Since they are not analytic with respect to η, we cannot apply Corollary 2.2. However, we will obtain a sequence of conserved densities for the sine–Gordon equation by showing that the one-form ω of Theorem 2.1 provides another closed one-form which *is* analytic with respect to η.

From Theorem 2.1 (i), we have the completely integrable system for ϕ,

$$\phi_x=u_x+\eta\cos\phi,\quad \phi_t=(1/\eta)\cos(u-\phi),\tag{34}$$

whenever u is a solution of (32).

We consider the first equation of (34) as an ordinary differential equation. Since the right-hand side of it is an analytic function of (ϕ,η), it follows that the solutions of (34) are analytic with respect to η. Therefore, we can consider

$$\phi=\sum_{j=0}^{\infty}\phi_j(x,t)\eta^j.$$

Hence (34) reduces to

$$\phi_{0,x}=u_x,\quad \cos(u-\phi_0)=0,\tag{35}$$

and, for $j\geqslant 1$,

$$\begin{aligned}
&\phi_{j,x}=\frac{1}{(j-1)!}\frac{d^{j-1}C}{d\eta^{j-1}}(0),\\
&\phi_{j-1,t}=\frac{1}{j!}\left(\sin u\frac{d^jS}{d\eta^j}(0)+\cos u\frac{d^jC}{d\eta^j}(0)\right).
\end{aligned}\tag{36}$$

From (35) we may consider

$$u-\phi_0=\pi/2,\tag{37}$$

and from the second equation of (36) we obtain, recursively, for $j\geqslant 1$,

$$\phi_j=\phi_{j-1,t}-\frac{1}{j}\sum_{i=1}^{j-1}\frac{j-i}{i!}\times\left(\sin u\frac{d^iC}{d\eta^i}(0)-\cos u\frac{d^iS}{d\eta^i}(0)\right)\phi_{j-i}.$$

It is not difficult to show that such ϕ_j satisfy the first equation of (36). We display explicitly the first terms:

$$\begin{aligned}
\phi_0 &= u - \pi/2\,,\\
\phi_1 &= u_t\,,\\
\phi_2 &= u_{tt}\,,\\
\phi_3 &= u_{ttt} + \tfrac{1}{6}(u_t)^3\,,\\
&\vdots\,.
\end{aligned}\tag{38}$$

Using (33) we obtain from Theorem 2.1 (ii) a closed one-form given by

$$\omega = -\eta \sin\phi\, dx + (1/\eta)\sin(u-\phi)dt\,.$$

Since ϕ is analytic with respect to η, then using (11) and (37) we reduce the above one-form to

$$\omega = \frac{1}{\eta}\,dt + \sum_{j=1}^{\infty} \Omega_j \eta^j\,, \tag{39}$$

where

$$\Omega_j = -\frac{1}{(j-1)!}\frac{d^{j-1}S}{d\eta^{j-1}}(0)dx + \frac{1}{(j+1)!} \times\left(\sin u\,\frac{d^{j+1}C}{d\eta^{j+1}}(0) - \cos u\,\frac{d^{j+1}S}{d\eta^{j+1}}(0)\right)dt\,.$$

The one-form ω being closed implies that $\Omega = \Sigma_{j=1}^{\infty}\Omega_j\eta^j$ is closed. Therefore, from (39) we obtain a sequence of conserved densities given by

$$\frac{d^jS}{d\eta^j}(0),\quad j\geqslant 0\,.$$

Substituting (12) into the above expression and using ϕ_j given by (38), we obtain

$$\begin{aligned}
&\cos u\,,\\
&u_t \sin u\,,\\
&u_{tt}\sin u + \tfrac{1}{2}u_t^2 \cos u\,,\\
&u_{ttt}\sin u + u_t u_{tt}\cos u\,,\\
&\vdots\,.
\end{aligned}$$

We conclude the discussion of the sine–Gordon equation by observing that by changing the independent variables, Eq. (32) is equivalent to

$$u_{xx} - u_{tt} = \sin u\,. \tag{40}$$

This is the integrability condition of an equation $dv = \Omega v$, where Ω is a 2×2 matrix of one-forms given by $\Omega = M\,dx + N\,dt$, where

$$\begin{aligned}
M &= \tfrac{1}{4}[\eta A + (1/\eta)Q + (u_x + u_t)P]\,,\\
N &= \tfrac{1}{4}[\eta A - (1/\eta)Q + (u_x + u_t)P]\,,
\end{aligned}$$

with

$$A = \begin{pmatrix} 1 & 0\\ 0 & -1\end{pmatrix},\quad Q = \begin{pmatrix}\cos u & \sin u\\ \sin u & -\cos u\end{pmatrix},$$
$$P = \begin{pmatrix} 0 & -1\\ 1 & 0\end{pmatrix}.$$

Applying Theorem 2.1 we can also obtain a sequence of conserved densities for the sine–Gordon equation (40), where the functions f_{ij} are determined by the matrices M and N.

F. Sinh–Gordon equation

This equation is expressed as

$$u_{xt} = \sinh u\,. \tag{41}$$

Consider the functions defined by

$$\begin{aligned}
&f_{11} = u_x,\quad f_{12} = 0\,,\\
&f_{21} = n,\quad f_{22} = (1/\eta)\cosh u\,,\\
&f_{31} = 0,\quad f_{32} = (1/\eta)\sinh u\,.
\end{aligned}\tag{42}$$

For any solution u of Eq. (41), the above functions satisfy (6). As in the preceeding example, we will obtain a sequence of conserved densities for (41) by using Theorem 2.1. Substituting (42) into (7), we obtain the system of equations

$$\begin{aligned}
\phi_x &= u_x \sin\phi + \eta\cos\phi\,,\\
\phi_t &= (1/\eta)\sinh u + (1/\eta)\cosh u\cos\phi\,,
\end{aligned}\tag{43}$$

which is completely integrable whenever u is a solution of (41).

From the first equation of (43) we conclude that ϕ is analytic with respect to η. Therefore, considering

$$\phi = \sum_{j=0}^{\infty}\phi_j(x,t)\eta^j\,,$$

(43) reduces to

$$\phi_{0,x} = u_x\sin\phi_0\,,\quad \cos\phi_0\cosh u + \sinh u = 0\,, \tag{44}$$

and, for $j\geqslant 1$,

$$\begin{aligned}
\phi_{j,x} &= u_x\frac{1}{j!}\frac{d^jS}{d\eta^j}(0) + \frac{1}{(j-1)!}\frac{d^{j-1}C}{d\eta^{j-1}}(0)\,,\\
\phi_{j-1,t} &= \cosh u\,\frac{1}{j!}\frac{d^jC}{d\eta^j}(0)\,.
\end{aligned}\tag{45}$$

From the second equation of (44) and (45) we obtain

$$\begin{aligned}
\cos\phi_0 &= -\tanh u\,,\\
\phi_j &= -\frac{1}{\sin\phi_0}\left[\frac{1}{\cosh u}\phi_{j-1,t} + \frac{1}{j}\sum_{i=1}^{j-1}\frac{j-i}{i!}\frac{d^iS}{d\eta^i}(0)\phi_{j-i}\right],\quad j\geqslant 1\,.
\end{aligned}\tag{46}$$

It is not difficult to show that Eqs. (46) satisfy the first equations of (44) and (45). The first functions are given by

$$\begin{aligned}
\cos\phi_0 &= -\tanh u\,,\\
\phi_1 &= -\frac{1}{\cosh u}u_t\,,\\
\phi_2 &= -\frac{1}{\sin\phi_0\cosh^2 u}\left(\frac{1}{2}\tanh u\, u_t^2 - u_{tt}\right),\\
&\vdots\,.
\end{aligned}\tag{47}$$

From Theorem 2.1 (ii), it follows that

$$\omega = (u_x\cos\phi - \eta\sin\phi)dx - (1/\eta)\cosh u\sin\phi\, dt$$

is a closed form whenever u and ϕ are solutions of (41) and (43), respectively. Since ϕ is analytic with respect to η, using (11) we reduce the above one-form to

$$\omega = -\frac{1}{\eta}\cosh u\sin\phi_0\, dt + \sum_{j=0}^{\infty}\Omega_j\eta^j\,, \tag{48}$$

where

$$\Omega_0 = \cos\phi_0(u_x\, dx - \phi_1 \cosh u\, dt)\,,$$

and, for $j \geqslant 1$,

$$\Omega_j = \left[\frac{u_x}{j!}\frac{d^jC}{d\eta^j}(0) - \frac{1}{(j-1)!}\frac{d^{j-1}S}{d\eta^{j-1}}(0)\right]dx - \frac{\cosh u}{(j+1)!}\frac{d^{j+1}S}{d\eta^{j+1}}(0)dt\,.$$

Now, we observe that since ϕ_0 satisfies (44), it follows that $\cosh u \sin\phi_0\, dt$ is a closed form. Therefore we conclude from (48) that

$$\Omega = \sum_{j=0}^{\infty} \Omega_j \eta^j$$

is also a closed form. Hence we obtain a sequence of conserved densities given by

$$u_x \cos\phi_0\,,$$

$$\frac{u_x}{j!}\frac{d^jC}{d\eta^j}(0) - \frac{1}{(j-1)!}\frac{d^{j-1}S}{d\eta^{j-1}}(0), \quad j \geqslant 1\,.$$

Using the expressions in Eq. (8) and the functions ϕ_j given by (47), we obtain the first conserved densities:

$$u_x \tanh u\,,$$

$$\sin\phi_0\left(1 - \frac{u_x u_t}{\cosh u}\right),$$

$$\frac{u_x}{\cosh^2 u}(-2u_t^2 \tanh u + 2u_{tt} + u_t \sinh u)\,,$$

$$\vdots\,.$$

[1]S. S. Chern and K. Tenenblat, "Pseudospherical surfaces and evolution equations," Stud. Appl. Math. **74**, 1 (1986).

[2]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, "Nonlinear evolution equations of physical significance," Phys. Rev. Lett. **31**, 125 (1973); "The inverse scattering transform—Fourier analysis for nonlinear problems," Stud. Appl. Math. **53**, 249 (1974).

[3]R. Sasaki, "Soliton equation and pseudospherical surfaces," Nucl. Phys. B **154**, 343 (1979).

[4]E. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955).

Nonlocal symmetries and the linearization of the massless Thirring and the Federbush models

Paul H. M. Kersten
Faculty of Applied Mathematics, University of Twente, P. O. Box 217, 7500 AE Enschede, The Netherlands

(Received 23 September 1987; accepted for publication 18 November 1987)

Multilation of nonlocal symmetries of the massless Thirring and the Federbush models leads to transformations linearizing these equations.

I. INTRODUCTION AND GENERAL

In recent papers[1–5] we studied local and nonlocal Lie–Bäcklund transformations of the massive Thirring and the Federbush models. Now we shall study nonlocal symmetries of the massless Thirring and the Federbush models. From these symmetries we obtain transformations that linearize these systems.

The nonlocal symmetries have to satisfy the symmetry condition[6]

$$\mathcal{L}_V I \subset I, \tag{1.1}$$

where V is a vector field, $\mathcal{L}_V$ denotes the Lie derivative with respect to the vector field V, and I denotes a closed ideal of differential forms generated by contact one-forms[7] and potential forms.[2]

In Sec. II we study the massless Thirring model, where we obtain the well-known transformation linearizing the system by mutilation of a nonlocal symmetry. In Sec. III the Federbush model is treated in an analogous way.

The computations have been carried through on a DEC System 20 computer using the symbolic language REDUCE 3.0.[8] We used software developed by Gragert *et al.*[9,10] to do differential geometric computations and software developed to solve overdetermined systems of partial differential equations.[11]

II. NONLOCAL SYMMETRIES AND THE LINEARIZATION OF THE MASSLESS THIRRING MODEL

The massless Thirring model is described by the following system of first-order partial differential equations:

$$\begin{aligned} &-U_{1x}+U_{1t}=-\lambda R_2V_1, \quad V_{1x}-V_{1t}=-\lambda R_2U_1,\\ &U_{2x}+U_{2t}=-\lambda R_1V_2, \quad -V_{2x}-V_{2t}=-\lambda R_1U_2, \end{aligned} \tag{2.1}$$

where $R_i=U_i^2+V_i^2$ $(i=1,2)$. In Ref. 1 we constructed ordinary and generalized infinitesimal symmetries and associated conserved vectors for the massive Thirring model. Associated with the infinitesimal symmetry

$$V_1\,\partial_{U_1}-U_1\,\partial_{V_1}+V_2\,\partial_{U_2}-U_2\,\partial_{V_2}, \tag{2.2}$$

there is a conserved current

$$(R_1+R_2)dx+(R_1-R_2)dt, \tag{2.3}$$

which is a conserved current for the massless Thirring model as well. We now construct an exterior differential system I for the massless Thirring model, including the potential p, defined on

$$\mathbb{R}^{11}=\{(x,t,U_1,\ldots,V_2,p,U_{1x},\ldots,V_{2x})\},$$

and generated by the differential forms

$$\begin{aligned} \alpha_1&=dU_1-U_{1x}\,dx-U_{1t}\,dt,\\ \alpha_2&=dV_1-V_{1x}\,dx-V_{1t}\,dt,\\ \alpha_3&=dU_2-U_{2x}\,dx-U_{2t}\,dt,\\ \alpha_4&=dV_2-V_{2x}\,dx-V_{2t}\,dt,\\ \alpha_5&=dp-(R_1+R_2)dx+(R_1-R_2)dt, \end{aligned} \tag{2.4}$$

and the exterior derivatives $d\alpha_1,\ldots,d\alpha_4$. The exterior derivative $d\alpha_5$ is in I by (2.3). In (2.4) $U_{1t},\ldots,V_{2t}$ are defined by (2.1).

The infinitesimal symmetry condition

$$\mathcal{L}_V I \subset I \tag{2.5}$$

leads to an overdetermined system of partial differential equations for the components of the vector field V. Solving this system by the assumption that the ∂_x and ∂_t components of the vector field V are independent of $U_1,\ldots,V_2$, we obtain the following result: The Lie algebra of the infinitesimal symmetries of (2.4) is generated by

$$\begin{aligned} V_1&=\partial_p,\\ V_2&=\tfrac12(U_1+\lambda pV_1)\partial_{U_1}+\tfrac12(V_1-\lambda pU_1)\partial_{V_1}\\ &\quad+\tfrac12(U_2-\lambda pV_2)\partial_{U_2}+\tfrac12(V_2+\lambda pU_2)\partial_{V_2}+p\,\partial_p,\\ V_3&=-H\,\partial_{U_1}+H\,\partial_{V_1},\\ V_4&=-K\,\partial_{U_2}+K\,\partial_{V_2},\\ V_5&=-\widetilde H\,\partial_x-\widetilde H\,\partial_t+\widetilde H\,\partial_{U_1}+\widetilde H\,\partial_{V_1},\\ V_6&=\widetilde K\,\partial_x-\widetilde K\,\partial_t+\widetilde K\,\partial_{U_2}+\widetilde K\,\partial_{V_2}, \end{aligned} \tag{2.6}$$

where H, K, $\widetilde H$, and $\widetilde K$ are functions of x and t satisfying the equations

$$\begin{aligned} &-H_x+H_t=0, \quad K_x+K_t=0,\\ &-\widetilde H_x+\widetilde H_t=0, \quad \widetilde K_x+\widetilde K_t=0. \end{aligned} \tag{2.7}$$

We now compute the local one-parameter group of symmetry transformations associated to the vector field V_2. The resulting system of differential equations is

$$\begin{aligned} &\frac{d\overline U_1}{ds}=\frac12(\overline U_1+\lambda\bar p\overline V_1), \quad \frac{d\overline U_2}{ds}=\frac12(\overline U_2-\lambda\bar p\overline V_2),\\ &\frac{d\overline V_1}{ds}=\frac12(\overline V_1-\lambda\bar p\overline U_1), \quad \frac{d\overline V_1}{ds}=\frac12(\overline V_2+\lambda\bar p\overline U_2),\\ &\frac{d\bar p}{ds}=\bar p, \end{aligned} \tag{2.8a}$$

with initial conditions

 0022-2488/88/041050-04$02.50

$$\overline{U}_1(0) = U_1,\ldots,\overline{p}(0) = p. \tag{2.8b}$$

The solution of (2.8a) and (2.8b) is

$$\begin{aligned}
\overline{U}_1 &= e^{s/2}\left\{U_1 \cos\frac{\lambda}{2}p(e^s-1) + V_1 \sin\frac{\lambda}{2}p(e^s-1)\right\},\\
\overline{V}_1 &= e^{s/2}\left\{V_1 \cos\frac{\lambda}{2}p(e^s-1) - U_1 \sin\frac{\lambda}{2}p(e^s-1)\right\},\\
\overline{U}_2 &= e^{s/2}\left\{U_2 \cos\frac{\lambda}{2}p(e^s-1) - V_2 \sin\frac{\lambda}{2}p(e^s-1)\right\},\\
\overline{V}_2 &= e^{s/2}\left\{V_2 \cos\frac{\lambda}{2}p(e^s-1) + U_2 \sin\frac{\lambda}{2}p(e^s-1)\right\},\\
\overline{p}_2 &= e^s p.
\end{aligned} \tag{2.9}$$

Now, setting $e^s = r$, we obtain

$$\begin{aligned}
\overline{U}_1 &= \sqrt{r}\left\{U_1 \cos\frac{\lambda}{2}p(r-1) + V_1 \sin\frac{\lambda}{2}p(r-1)\right\},\\
\overline{V}_1 &= \sqrt{r}\left\{V_1 \cos\frac{\lambda}{2}p(r-1) - U_1 \sin\frac{\lambda}{2}p(r-1)\right\},\\
\overline{U}_2 &= \sqrt{r}\left\{U_2 \cos\frac{\lambda}{2}p(r-1) - V_2 \sin\frac{\lambda}{2}p(r-1)\right\},\\
\overline{V}_2 &= \sqrt{r}\left\{V_2 \cos\frac{\lambda}{2}p(r-1) + U_2 \sin\frac{\lambda}{2}p(r-1)\right\},\\
\overline{p} &= rp.
\end{aligned} \tag{2.10}$$

From (2.4) we formally have

$$p = \int_{-\infty}^{x} (U_1^2 + V_1^2 + U_2^2 + V_2^2)dx. \tag{2.11}$$

Since (2.10) is a symmetry transformation, (2.1) is transformed into (2.1) with $U_1,\ldots,V_2$ replaced by $\overline{U}_1,\ldots,\overline{V}_2$.

We now "mutilate" the symmetry transformation (2.10), i.e.,

$$\begin{aligned}
\overline{U}_1 &= U_1 \cos\frac{\lambda}{2}p(r-1) + V_1 \sin\frac{\lambda}{2}p(r-1),\\
\overline{V}_1 &= V_1 \cos\frac{\lambda}{2}p(r-1) - U_1 \sin\frac{\lambda}{2}p(r-1),\\
\overline{U}_2 &= U_2 \cos\frac{\lambda}{2}p(r-1) - V_2 \sin\frac{\lambda}{2}p(r-1),\\
\overline{V}_2 &= V_2 \cos\frac{\lambda}{2}p(r-1) + U_2 \sin\frac{\lambda}{2}p(r-1),\\
\overline{p} &= p,
\end{aligned} \tag{2.12}$$

or

$$\begin{aligned}
U_1 &= \overline{U}_1 \cos\frac{\lambda}{2}\overline{p}(r-1) - \overline{V}_1 \sin\frac{\lambda}{2}\overline{p}(r-1),\\
V_1 &= \overline{V}_1 \cos\frac{\lambda}{2}\overline{p}(r-1) + \overline{U}_1 \sin\frac{\lambda}{2}\overline{p}(r-1),\\
U_2 &= \overline{U}_2 \cos\frac{\lambda}{2}\overline{p}(r-1) + \overline{V}_2 \sin\frac{\lambda}{2}\overline{p}(r-1),\\
V_2 &= \overline{V}_2 \cos\frac{\lambda}{2}\overline{p}(r-1) - \overline{U}_2 \sin\frac{\lambda}{2}\overline{p}(r-1),\\
p &= \overline{p}.
\end{aligned} \tag{2.13}$$

Substitution of (2.13) into (2.1) yields the following nonlinear first-order system:

$$\begin{aligned}
&-\overline{U}_{1x} + \overline{U}_{1t} = -r\lambda\overline{R}_2\overline{V}_1, \quad \overline{V}_{1x} - \overline{V}_{1t} = -r\lambda\overline{R}_2\overline{U}_1,\\
&\overline{U}_{2x} + \overline{U}_{2t} = -r\lambda\overline{R}_1\overline{V}_2, \quad -\overline{V}_{2x} - \overline{V}_{2t} = -r\lambda\overline{R}_1\overline{U}_2.
\end{aligned} \tag{2.14}$$

Now, for $r = 0$, the right-hand side of (2.14) vanishes, so the transformation

$$\begin{aligned}
\overline{U}_1 &= U_1 \cos\frac{\lambda}{2}p + V_1 \sin\frac{\lambda}{2}p,\\
\overline{V}_1 &= V_1 \cos\frac{\lambda}{2}p - U_1 \sin\frac{\lambda}{2}p,\\
\overline{U}_2 &= U_2 \cos\frac{\lambda}{2}p - V_2 \sin\frac{\lambda}{2}p,\\
\overline{V}_2 &= V_2 \cos\frac{\lambda}{2}p + U_2 \sin\frac{\lambda}{2}p,
\end{aligned} \tag{2.15}$$

linearizes the massless Thirring model into

$$\begin{aligned}
&-\overline{U}_{1x} + \overline{U}_{1t} = 0, \quad \overline{U}_{2x} + \overline{U}_{2t} = 0,\\
&\overline{V}_{1x} - \overline{V}_{1t} = 0, \quad -\overline{V}_{2x} - \overline{V}_{2t} = 0.
\end{aligned} \tag{2.16}$$

III. NONLOCAL SYMMETRIES AND THE LINEARIZATION OF THE FEDERBUSH MODEL

We construct nonlocal symmetries of the Federbush model. By mutilation of the resulting one-parameter group of symmetry transformations, we obtain transformations of the Federbush model into another nonlinear model that, for a certain value of the parameter, results in the well-known linearizing transformation of the Federbush model.

The Federbush model is described by

$$\begin{pmatrix} i(\partial_t + \partial_x) & -m(k) \\ -m(k) & i(\partial_t - \partial_x) \end{pmatrix}\begin{pmatrix} \psi_{k,1} \\ \psi_{k,2} \end{pmatrix} = k\lambda\begin{pmatrix} +|\psi_{-k,2}|^2\psi_{k,1} \\ -|\psi_{-k,1}|^2\psi_{k,2} \end{pmatrix} \quad (k = \pm 1) \tag{3.1}$$

where $\psi_k(x,t)$ $(k = \pm 1)$ are two-component functions defined on $\mathbb{C}$. Introducing U_i, V_i $(i,\ldots,4)$ by

$$\begin{aligned}
\psi_{1,1} &= U_1 + iV_1, \quad \psi_{-1,1} = U_3 + iV_3,\\
\psi_{1,2} &= U_2 + iV_2, \quad \psi_{-1,2} = U_4 + iV_4,
\end{aligned} \tag{3.2}$$

system (3.1) is rewritten as a system of eight nonlinear partial differential equations:

$$\begin{aligned}
U_{1t} + U_{1x} - m_1V_2 &= \lambda R_4V_1,\\
V_{1t} + V_{1x} + m_1U_2 &= -\lambda R_4U_1,\\
U_{2t} - U_{2x} - m_1V_1 &= -\lambda R_3V_2,\\
V_{2t} - V_{2x} + m_1U_1 &= \lambda R_3U_2,\\
U_{3t} + U_{3x} - m_2V_4 &= -\lambda R_2V_3,\\
V_{3t} + V_{3x} + m_2U_4 &= \lambda R_2U_3,\\
U_{4t} - U_{4x} - m_2V_3 &= \lambda R_1V_4,\\
V_{4t} - V_{4x} + m_2U_3 &= -\lambda R_1U_4,
\end{aligned} \tag{3.3}$$

where

$$\begin{aligned}
&m(1) = m_1, \quad m(-1) = m_2,\\
&R_i = U_i^2 + V_i^2 \quad (i = 1,\ldots,4).
\end{aligned} \tag{3.4}$$

In Ref. 3 we obtained two infinitesimal symmetries,

$$V_1 = -V_1\,\partial_{U_1} + U_1\,\partial_{V_1} - V_2\,\partial_{U_2} + U_2\,\partial_{V_2},$$
$$V_2 = -V_3\,\partial_{U_3} + U_3\,\partial_{V_3} - V_4\,\partial_{U_4} + U_4\,\partial_{V_4}, \tag{3.5}$$

giving rise to conserved currents[7]

$$(R_1+R_2)dx + (-R_1+R_2)dt,$$
$$(R_3+R_4)dx + (-R_3+R_4)dt. \tag{3.6}$$

We now formally introduce the nonlocal variables p_1 and p_2 by

$$p_1 = \int_{-\infty}^{x} (R_1+R_2)dx,$$
$$p_2 = \int_{-\infty}^{x} (R_3+R_4)dx, \tag{3.7}$$

and construct the exterior differential system I, defined on

$$\mathbb{R}^{20} = \{(x,t,U_1,\ldots,V_4,p_1,p_2,U_{1x},\ldots,V_{4x})\},$$

and generated by the differential one-forms

$$\begin{aligned}
\alpha_1 &= dU_1 - U_{1x}\,dx - U_{1t}\,dt,\\
\alpha_2 &= dV_1 - V_{1x}\,dx - V_{1t}\,dt,\\
\alpha_3 &= dU_2 - U_{2x}\,dx - U_{2t}\,dt,\\
\alpha_4 &= dV_2 - V_{2x}\,dx - V_{2t}\,dt,\\
\alpha_5 &= dU_3 - U_{3x}\,dx - U_{3t}\,dt,\\
\alpha_6 &= dV_3 - V_{3x}\,dx - V_{3t}\,dt,\\
\alpha_7 &= dU_4 - U_{4x}\,dx - U_{4t}\,dt,\\
\alpha_8 &= dV_4 - V_{4x}\,dx - V_{4t}\,dt,\\
\alpha_9 &= dp_1 - (R_1+R_2)dx - (-R_1+R_2)dt,\\
\alpha_{10} &= dp_2 - (R_3+R_4)dx - (-R_3+R_4)dt,
\end{aligned} \tag{3.8}$$

and the exterior derivatives $d\alpha_1,\ldots,d\alpha_8$. The exterior derivatives $d\alpha_9$, $d\alpha_{10}$ are in I by (3.6).

The symmetry condition

$$\mathscr{L}_V I \subset I \tag{3.9}$$

leads to an overdetermined system of partial differential equations for the components of the vector field V. The Lie algebra of infinitesimal symmetries is generated by the vector fields

$$\begin{aligned}
V_1 &= \partial_x, \quad V_2 = \partial_t,\\
V_3 &= V_1\,\partial_{U_1} - U_1\,\partial_{V_1} + V_2\,\partial_{U_2} - U_2\,\partial_{V_2},\\
V_4 &= V_3\,\partial_{U_3} - U_3\,\partial_{V_3} + V_4\,\partial_{U_4} - U_4\,\partial_{V_4},\\
V_5 &= t\,\partial_x + x\,\partial_t\\
&\quad + \tfrac{1}{2}(U_1\,\partial_{U_1} + V_1\,\partial_{V_1} - U_2\,\partial_{U_2} - V_2\,\partial_{V_2})\\
&\quad + \tfrac{1}{2}(U_3\,\partial_{U_3} + V_3\,\partial_{V_3} - U_4\,\partial_{U_4} - V_4\,\partial_{V_4}),\\
V_6 &= \partial_{p_1}, \quad V_7 = \partial_{p_2},\\
V_8 &= U_1\,\partial_{U_1} + V_1\,\partial_{V_1} + U_2\,\partial_{U_2} + V_2\,\partial_{V_2}\\
&\quad + \lambda p_1(-V_3\,\partial_{U_3} + U_3\,\partial_{V_3} - V_4\,\partial_{V_4} + U_4\,\partial_{U_4})\\
&\quad + 2p_1\,\partial_{p_1},\\
V_9 &= \lambda p_2(V_1\,\partial_{U_1} - U_1\,\partial_{V_1} + V_2\,\partial_{U_2} - U_2\,\partial_{V_2})\\
&\quad + U_3\,\partial_{U_3} + V_3\,\partial_{V_3} + U_4\,\partial_{U_4} + V_4\,\partial_{V_4} + 2p_2\,\partial_{p_2}.
\end{aligned} \tag{3.10}$$

The symmetries $V_1,\ldots V_5$ are the classical local ones [cf. Ref. 3, (2.6)]. We now compute the local one-parameter group of symmetries associated to the vector field $V_8 + V_9$. This leads to the following result:

$$\begin{aligned}
\tilde{U}_1(s) &= e^s(U_1\cos r_1 + V_1\sin r_1),\\
\tilde{V}_1(s) &= e^s(-U_1\sin r_1 + V_1\cos r_1),\\
\tilde{U}_2(s) &= e^s(U_2\cos r_1 + V_2\sin r_1),\\
\tilde{V}_2(s) &= e^s(-U_2\sin r_1 + V_2\cos r_1),\\
\tilde{U}_3(s) &= e^s(U_3\cos r_2 - V_3\sin r_2),\\
\tilde{V}_3(s) &= e^s(U_3\sin r_2 + V_3\cos r_2),\\
\tilde{U}_4(s) &= e^s(U_4\cos r_2 - V_4\sin r_2),\\
\tilde{V}_4(s) &= e^s(U_4\sin r_2 + V_4\cos r_2),
\end{aligned} \tag{3.11a}$$

or, from (3.2),

$$\begin{aligned}
\tilde{\psi}_{1,1} &= e^s\psi_{1,1}e^{ir_1}, \quad \tilde{\psi}_{1,2} = e^s\psi_{1,2}e^{ir_1},\\
\tilde{\psi}_{-1,1} &= e^s\psi_{-1,1}e^{-ir_2}, \quad \tilde{\psi}_{-1,2} = e^s\psi_{-1,2}e^{-ir_2},
\end{aligned} \tag{3.11b}$$

$$\tilde{p}_1(s) = e^{2s}p_1, \quad \tilde{p}_2(S) = e^{2s}p_2, \tag{3.11c}$$

where s is the group parameter and

$$r_1 = (\lambda/2)p_2(e^{2s}-1), \quad r_2 = (\lambda/2)p_1(e^{2s}-1), \tag{3.12}$$

$s = 0$ representing the identity map in (3.11). By definition, the transformation (3.11) maps solutions of the Federbush model into solutions of the same system.

We now mutilate the transformations defined by (3.11) in a way similar to that in Sec. II, i.e., by deleting e^s in (3.11a) and [as a consequence of (3.7)] e^{2s} in (3.11b), which results in the transformations

$$\begin{aligned}
\tilde{\psi}_{1,1} &= \psi_{1,1}e^{ir_1}, \quad \tilde{\psi}_{1,2} = \psi_{1,2}e^{ir_1},\\
\tilde{\psi}_{-1,1} &= \psi_{-1,1}e^{-ir_2}, \quad \tilde{\psi}_{-1,2} = \psi_{-1,2}e^{-ir_2},
\end{aligned} \tag{3.13}$$

and their inverse.

An easy calculation now shows that if $\psi_{i,j}$ ($i = \pm 1$, $j = 1,2$) is a solution of (3.1), then $\tilde{\psi}_{i,j}$ satisfies the following nonlinear system of partial differential equations:

$$\begin{pmatrix} i(\partial_t+\partial_t) & -m(k)\\ -m(k) & i(\partial_t+\partial_x)\end{pmatrix}\begin{pmatrix}\tilde{\psi}_{k,1}\\ \tilde{\psi}_{k,2}\end{pmatrix} = sk\lambda\begin{pmatrix} +|\tilde{\psi}_{-k,2}|^2\tilde{\psi}_{k,1}\\ -|\tilde{\psi}_{-k,1}|^2\tilde{\psi}_{k,2}\end{pmatrix}e^{2s}. \tag{3.14}$$

Now, for $s = 0$, the nonlinear term in the right-hand side of (2.14) cancels out, and, consequently, the transformations

$$\begin{aligned}
\tilde{\psi}_{1,1} &= \psi_{1,1}e^{(-i\lambda/2)p_2}, \quad \tilde{\psi}_{1,2} = \psi_{1,2}e^{(-i\lambda/2)p_2},\\
\tilde{\psi}_{-1,1} &= \psi_{-1,1}e^{(+i\lambda/2)p_1}, \quad \tilde{\psi}_{-1,2} = \psi_{-1,2}e^{(+i\lambda/2)p_1},
\end{aligned} \tag{3.15}$$

transform (3.1) into the *linear* system

$$\begin{pmatrix} i(\partial_t+\partial_x) & -m(k)\\ -m(k) & i(\partial_t+\partial_x)\end{pmatrix}\begin{pmatrix}\tilde{\psi}_{k,1}\\ \tilde{\psi}_{k,2}\end{pmatrix} = 0. \tag{3.16}$$

IV. CONCLUSION

By mutilation of nonlocal symmetries we obtain transformations of the massless Thirring and the Federbush models linearizing these equations. The Lie algebra structure of these and higher-order nonlocal symmetries will be studied elsewhere.

[1]P. H. M. Kersten and R. Martini, J. Math. Phys. **26**, 822 (1985).
[2]P. H. M. Kersten and R. Martini, J. Math. Phys. **26**, 1775 (1985).
[3]P. H. M. Kersten, J. Math. Phys. **27**, 1139 (1986).
[4]P. H. M. Kersten and H. M. M. ten Eikelder, J. Math. Phys. **27**, 2140 (1986).
[5]P. H. M. Kersten and H. M. M. ten Eikelder, J. Math. Phys. **27**, 2791 (1986).
[6]B. K. Harrison and F. B. Estabrook, J. Math. Phys. **12**, 653 (1971).
[7]F. A. E. Pirani, D. C. Robinson, and W. F. Shadwick, "Local jetbundle formulation of Bäcklund transformations," in *Mathematical Physica Studies* (Reidel, Boston, 1979), Vol. I.
[8]A. C. Hearn, *REDUCE User's Manual* (Version 3.0) (The Rand Corporation, Santa Monica, 1983).
[9]P. K. H. Gragert, Ph.D. thesis, Twente University of Technology, Enschede, The Netherlands, 1981.
[10]P. K. H. Gragert, P. H. M. Kersten, and R. Martini, Acta Appl. Math. **1**, 43 (1983).
[11]P. H. M. Kersten, "Infinitesimal symmetries: A computational approach," CWI Tract 34, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands.

All possible local charges in a local quantum field theory: Massive case

J. M. Amigó
Institut für Theoretische Physik, Bunsenstrasse 9, 3400 Göttingen, Federal Republic of Germany

(Received 5 August 1987; accepted for publication 9 December 1987)

A proof for local charges of the classical result first put forward by Coleman and Mandula [Phys. Rev. **159**, 1251 (1967)] is given. Local charges are operators defined as integrals of the time component of conserved Hermitian density currents; in interacting theories they happen to be generators of symmetries of the S matrix.

I. INTRODUCTION

Twenty years ago Coleman and Mandula[1] excluded the possibility of a nontrivial connection between internal and space-time symmetries of the S matrix in an interacting, massive, local quantum field theory. Their fundamental "no go" theorem not only established an important feature of the quantum field theories of that time (i.e., before the flourishing of the gauge field theories), but also influenced decisively the further research on physical symmetries, making possible the advent of the supersymmetric algebras.[2] What seems dubious to many people, however, is the way this theorem was proved, since the authors' reasoning is not rigorous but rather heuristic. As the authors themselves confess in the introduction to Ref. 1, "Although at times this [the proof] attains mathematical levels of obscurity, we make no claim for corresponding standards of rigor."

Considering the importance of such a result, several attempts were undertaken after the publication of Ref. 1 in order to provide a satisfactory proof.[3] The most careful approach known to me is that of Garber and Reeh.[4] However, their result is weaker than that of Coleman and Mandula in two aspects: (i) it does not involve arbitrary generators of Lie groups of symmetries of the S matrix, but only "local charges," i.e., bosonic symmetric generators defined as integrals over the Minkowski space of the time component of conserved current densities; and (ii) it covers only the case of scalar field theory. Later this second restriction was slightly removed by Reeh,[5] who extended the result of Ref. 4 to the case of scalar and spinor quantum field theory, and also by Łopuszański and Amigó (unpublished), who found different methods of coping with special theories with vectorial and scalar fields.

In this paper I present a proof concerning the structure of local charges in local quantum field theories (without spontaneous symmetry breaking) describing interacting massive particles with arbitrary spins. This completes the program initiated by Reeh in Ref. 5 in order to encompass massive theories with higher and higher spins. The present state of this question in the case of massless theories is briefly referred to in Sec. IV (c).

II. STATEMENTS AND RESULTS

Consider a Wightman field theory. Its fields $\Psi_j(x)$, $j\in I$, are meant to interpolate the asymptotic fields $\Psi_j^{\rm ex}(x)$ ("ex" stands for "in" or "out") of a relativistic scattering theory describing particles with mass $m=m(j)>0$ and spin $\sigma=\sigma(j)$. This amounts to a splitting of the Hilbert space $\mathscr{H}$ of the physical states in minimally invariant subspaces,

$$\mathscr{H}=\Omega\oplus\mathscr{H}^{(1)}\oplus\mathscr{H}^{(n\geqslant 2)},\tag{1}$$

where Ω is the vacuum, $\mathscr{H}^{(1)}$ is the one-particle subspace, and $\mathscr{H}^{(n\geqslant 2)}$ is the many-particle subspace. The one-particle states are created from the vacuum by the fields $\Psi_j^{\rm ex}$ and belong to the discrete irreducible representation $[m(j),\sigma(j)]$ $[m(j)>0,\ \forall j\in I]$ of $\tilde{\mathscr{P}}$, the covering group of the Poincaré group.[6] Correspondingly, each field $\Psi_j^{\rm ex}(x)$ has definite transformation properties under SL(2,$\mathbb{C}$), the covering group of the Lorentz group $\mathscr{L}^{\uparrow}_{+}$: It describes a boson or a fermion according to whether its (possibly reducible) representation is purely tensorial or purely spinorial.

In this framework let Q be an operator in $\mathscr{H}$ defined as the integral over the Minkowski space of the time component of a conserved (not necessarily covariant) Hermitian current density. Such Hermitian operators are called local charges; they induce an infinitesimal transformation of the fields belonging to the theory such that the transformed fields are local again, without changing their statistics.

Furthermore, let the theory be supplemented by the rather modest assumptions (A1) existence of a mass gap above the vacuum Ω, (A2) finite multiplicity of the one-particle hyperboloids, which are also supposed to lie below the continuum, and (A3) invariance of the vacuum, i.e., $Q\Omega=0$.

Within this extended framework one can show[7] that Q acts upon the (in general multicomponent) asymptotic fields

$$\Psi_j^{\rm ex}(x)=\{\Psi_{j\alpha}^{\rm ex}(x)\}_{\alpha=1}^{n(j)}\tag{2}$$

in the following way:

$$[Q,\Psi_{j\alpha}^{\rm ex}(x)]=\sum_k\sum_{\beta=1}^{n(k)}P_{j\alpha,k\beta}\left(x,\frac{1}{i}\partial\right)\Psi_{k\beta}^{\rm ex}(x).\tag{3}$$

Here the c-number coefficients $P_{j\alpha,k\beta}(x,(1/i)\partial)$ $[j,k\in I,\ 1\leqslant\alpha\leqslant n(j),\ 1\leqslant\beta\leqslant n(k)]$ are polynomials in the generators of the Poincaré group

$$P^\rho:=\frac{1}{i}\frac{\partial}{\partial x_\rho}=:\frac{1}{i}\partial^\rho$$

and

$$M^{\mu\nu}:=x^\mu\frac{1}{i}\partial^\nu-x^\nu\frac{1}{i}\partial^\mu+S^{\mu\nu}$$

$[S^{\mu\nu}$ are generators of SO(3,1) of the correct dimensionali-

 0022-2488/88/041054-07$02.50

ty], and the first sum runs over all particles k with the same statistics as the particle j.

Some basic properties of the local charge Q follow readily from (3).

(P1) Q commutes with the S matrix since $P_{j\alpha,k\beta}(x,(1/i)\partial)$ does not depend on the label "ex." Therefore local charges are always candidates for symmetry generators of the S matrix (their self-adjointness cannot be taken for granted).

(P2) Locality implies[8] that $P_{j\alpha,k\beta}(x,(1/i)\partial)$ vanishes for $m(j)\neq m(k)$. Thus Q connects only particles with the same mass.

(P3) Q acts "additively" on the asymptotic many-particle states, i.e., it transforms many-particle states as if they were tensor products.

The additivity of the local charges is a consequence of the [in (A3) assumed] invariance of the vacuum; it allows reconstruction of the action of Q on the whole Hilbert space $\mathscr{H}$ solely from its restriction to $\mathscr{H}^{(1)}$. Furthermore, we may confine our attention hereafter to the submatrix acting on a single mass multiplet of particles in agreement with (P2); all such multiplets are finite according to (A2).

The purpose of the present paper is to investigate further the structure of Q in the only interesting case—where the S matrix is nontrivial. Since, then, the optical theorem (short-range forces!) assures the occurrence of elastic forward scattering in $\mathscr{H}^{(2)}$, I state the precise assumption to be used later as to what "interaction" means in the following terms.

Assumption 2.1 (interaction assumption): There exists an open neighborhood $\mathscr{U}$ on the scattering manifold $\mathscr{M}$,

$$\mathscr{M}:=\{(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4)\in\mathbb{R}^{12}:$$
$$p_1+p_2=p_3+p_4,\ p^0=(\mathbf{p}^2+m^2)^{1/2}\},$$

with a nonempty intersection with the forward scattering, so that (i) in $\mathscr{U}$ all the particles of the considered mass multiplet interact with each other, and (ii) the elastic reduced two-particle scattering amplitude is continuous in $\mathscr{U}$ for all pairs of particles.

The interaction assumption is strong in the sense that it must hold for every basis of one-particle states. With this proviso it is possible to prove the following theorem.

Theorem 2.2: Q is at most a linear combination of the generators of the Poincaré group and of internal symmetries.

III. THE PROOF

Since only asymptotic fields appear in the proof and the theory is asymptotically complete, I shall exclusively use, say, "in" fields, omitting the label "in" henceforth. I split the proof into two steps.

A. The translationally invariant case

Consider first the case $[Q,P^\rho]=0$, $0\leqslant\rho\leqslant 3$, so that the polynomials $P_{j\alpha,k\beta}(x,(1/i)\partial)$ in (3) cannot depend on x, i.e.,

$$[Q,\Psi_{j\alpha}(x)]=P_{j\alpha,k\beta}((1/i)\partial)\Psi_{\beta k}(x) \tag{4a}$$

(summation convention!).

In order to keep the proof as simple as possible, I may assume that all the free fields $\Psi_j(x)$ belong to the irreducible representation $(\sigma(j),0)$ of $\mathrm{SL}(2,\mathbb{C})$. The reason is as follows: All the irreducible pieces of $\Psi_j(x)$ not transforming as $(\sigma(j),0)$ can be obtained from this representation by applying suitable polynomials in ∂ [$\sim(\frac{1}{2},\frac{1}{2})$], and, moreover, this procedure can also be inverted with such polynomials on account of the equations of motion of the fields $\Psi_j(x)$.[9,10] This being the case, one can absorb the differential operators which change the representation of $\Psi_j(x)$ and $\Psi_k(x)$ in (4a) to the $(\sigma(j),0)$ and $(\sigma(k),0)$, respectively, in the coefficients $P_{j\alpha,k\beta}((1/i)\partial)$ without changing their polynomical character at all. In such representations the asymptotic fields have as many different components, namely $2\sigma(j)+1$,

$$\{\Psi_{jr}(x)\}_{r=\sigma(j)}^{-\sigma(j)}$$
$$=\{\Psi_{j(1\cdots 11)}(x),\Psi_{j(1\cdots 10)}(x),\dots,\Psi_{j(0\cdots 00)}(x)\},$$

as physical degrees of freedom (polarizations) that the particles they describe have.

Hence, without loss of generality, one can take

$$[Q,\Psi_{jr}(x)]=P_{jr,ks}((1/i)\partial)\Psi_{ks}(x)\quad [\Psi_j(x)\sim(\sigma(j),0)] \tag{4b}$$

as the starting point. Here the sum runs both over all particles k in the multiplet with the same statistics as the particle j (i.e., both bosons or both fermions) and over all polarizations $s\in\{-\sigma(k),\dots,\sigma(k)\}$ (which, correspondingly, are integers or half-integers).

Next, decompose the fields into positive- and negative-frequency parts,[9,11]

$$\Psi_{jr}(x)=(2\pi)^{-3/2}\int\frac{d^3p}{2\omega_p}\times\sum_{r'=\sigma(j)}^{-\sigma(j)}(e^{-\theta(|\mathbf{p}|)\hat{\mathbf{p}}\cdot\mathbf{J}^{\sigma(j)}})_{r,r'}\Big\{a^*_{jr'}(\mathbf{p})e^{ipx}+\sum_{r''=\sigma(j)}^{-\sigma(j)}(C^{\sigma(j)})^{-1}_{r',r''}b_{jr''}(\mathbf{p})e^{-ipx}\Big\}, \tag{5}$$

in the distributive sense. The notation is as follows.

(i) The destruction and creation operators $a_{jr}(\mathbf{p})$ and $a^*_{jr}(\mathbf{p})$, respectively, of a particle j with momentum $\mathbf{p}$ and polarization r are canonically quantized, i.e.,

$$[a_{jr}(\mathbf{p}),a^*_{ks}(\mathbf{q})]_{(-)^{2\sigma(j)+1}}=2(\mathbf{p}^2+m^2)^{1/2}\delta_{jk}\delta_{rs}\delta^3(\mathbf{p}-\mathbf{q}), \tag{6}$$

and zero otherwise. The same applies to the corresponding operators for the antiparticles $b_{jr}(\mathbf{p})$ and $b^*_{jr}(\mathbf{p})$. If the particle j is self-conjugated, then one must set $a_{jr}(\mathbf{p})=b_{jr}(\mathbf{p})$.

(ii) In the exponents,

$$p^0=(\mathbf{p}^2+m^2)^{1/2}=:\omega_p, \tag{7}$$
$$\theta(|\mathbf{p}|):=\sinh^{-1}(|\mathbf{p}|/m)=:\theta, \tag{8}$$
$$\hat{\mathbf{p}}:=\mathbf{p}/|\mathbf{p}|. \tag{9}$$

The matrices $\mathbf{J}^{\sigma(j)}=(J_1^{\sigma(j)},J_2^{\sigma(j)},J_3^{\sigma(j)})$ are the (Hermitian) generators of the $(2\sigma(j)+1)$-dimensional irreducible representation of SU(2). The choice

$$(J_1^{\sigma(j)}\pm iJ_2^{\sigma(j)})_{r,r'}=((\sigma(j)\pm r)(\sigma(j)\mp r+1))^{1/2}\delta_{r,r'\pm 1}, \tag{10a}$$
$$(J_3^{\sigma(j)})_{r,r'}=r\delta_{r,r'}, \tag{10b}$$

$\forall r,r' \in \{-\sigma(j),...,\sigma(j)\}$, fixes all these representations; physically, this amounts to measuring the polarization of all the particles involved in the three-direction. Furthermore, one can show[9] that

$$(e^{\pm 2\theta \hat{\mathbf{p}}\cdot \mathbf{J}^{\sigma(j)}})_{r,r'} = \text{polynomial in } \omega_p, \mathbf{p}, \tag{11}$$

$\forall r,r'$ [so that (5) makes sense as an operator-valued tempered distribution].

(iii) The expression

$$(C^{\sigma(j)})_{r',r''} := (-1)^{\sigma(j)+r'} \delta_{r',-r''} \tag{12}$$

is the so-called charge conjugation matrix. The relation

$$C^{\sigma(j)} \overline{J_i^{\sigma(j)}} C^{\sigma(j)^{-1}} = -J_i^{\sigma(j)} \quad (i = 1,2,3) \tag{13}$$

will be used shortly.

Inserting (5) into (4b) one obtains, for the positive-frequency part (summation convention!),

$$[Q, a^*_{jr}(\mathbf{p})] = (e^{\theta \hat{\mathbf{p}}\cdot \mathbf{J}^{\sigma(j)}})_{r,r'} P_{jr',ks'}(\omega_p, \mathbf{p}) \times (e^{-\theta \hat{\mathbf{p}}\cdot \mathbf{J}^{\sigma(k)}})_{s',s} a^*_{ks}(\mathbf{p}), \tag{14a}$$

and for the negative-frequency one,

$$[Q, b^*_{jr}(\mathbf{p})] = -(e^{-\theta \hat{\mathbf{p}}\cdot \mathbf{J}^{\sigma(j)}})_{r,r'} (C^{\sigma(j)})_{r',r''} \times \overline{P_{jr'',ks''}(-\omega_p, -\mathbf{p})} (C^{\sigma(k)})^{-1}_{s'',s'} \times (e^{\theta \hat{\mathbf{p}}\cdot \mathbf{J}^{\sigma(k)}})_{s',s} b^*_{ks}(\mathbf{p}), \tag{14b}$$

where (13) has been used.

From (14a) and (14b) it does not follow that Q does not mix particles with antiparticles, since in general both $\Psi_k(x)$ and $\Psi^*_k(x)$ appear among the fields in (4a). In order to achieve a unified treatment, $a^*_{jr}(\mathbf{p})$, $a^*_{ks}(\mathbf{p})$, etc., stand hereafter for the creation operators of both particles and antiparticles (enlarge the index set I!).

Defining

$$H_{jr,ks}(\mathbf{p}) := (e^{\theta \hat{\mathbf{p}}\cdot \mathbf{J}^{\sigma(j)}})_{r,r'} P_{jr',ks'}(\omega_p, \mathbf{p}) (e^{-\theta \hat{\mathbf{p}}\cdot \mathbf{J}^{\sigma(k)}})_{s',s}, \tag{15}$$

Eq. (14a) reads

$$[Q, a^*_{jr}(\mathbf{p})] = H_{jr,ks}(\mathbf{p}) a^*_{ks}(\mathbf{p}). \tag{16a}$$

Observe that, by virtue of (11), Eq. (14b) has actually this form.

At this point, a notational simplification is convenient: Capital letters from the beginning of the alphabet, $A,B,...$ (also primed) will be used to enumerate the possible values of the multi-indices (j,r), (k,s),... ordered, say, lexicographically. Let $1 \leqslant A,B,... \leqslant N$, $N = \Sigma_j (2\sigma(j)+1)$, be their range.

With this convention (16a) becomes

$$[Q, a^*_A(\mathbf{p})] = H_{A,B}(\mathbf{p}) a^*_B(\mathbf{p}). \tag{16b}$$

Lemma 3.1: The $N \times N$ matrix

$$H(\mathbf{p}) := (H_{A,B}(\mathbf{p}))_{1 \leqslant A,B \leqslant N} \tag{17}$$

is Hermitian ($\forall \mathbf{p} \in \mathbb{R}^3$).

Proof: The Hermiticity of Q implies

$$(Q a^*_A(\mathbf{p})\Omega | a^*_B(\mathbf{q})\Omega) = (a^*_A(\mathbf{p})\Omega | Q a^*_B(\mathbf{q})\Omega),$$

and this, together with the invariance of the vacuum Ω and (16b), leads to

$$\overline{H_{A,A'}(\mathbf{p})} (\Omega | [a_{A'}(\mathbf{p}), a^*_B(\mathbf{q})] \Omega) = H_{B,B'}(\mathbf{q}) (\Omega | [a_A(\mathbf{p}), a^*_{B'}(\mathbf{q})] \Omega).$$

The canonical commutation relations (6) do the rest. □

In order to introduce the interaction at this stage, one possibility is to extend the technique employed in proving Lemma 3.1 to $\mathscr{H}^{(2)}$, for according to assumption (2.1), $S|\mathscr{H}^{(2)} \neq \mathbf{1}$ holds. In fact, from

$$(Q a^*_{A_1}(\mathbf{p}_1) a^*_{A_2}(\mathbf{p}_2)\Omega | S a^*_{A_3}(\mathbf{p}_3) a^*_{A_4}(\mathbf{p}_4)\Omega) = (a^*_{A_1}(\mathbf{p}_1) a^*_{A_2}(\mathbf{p}_2)\Omega | S Q a^*_{A_3}(\mathbf{p}_3) a^*_{A_4}(\mathbf{p}_4)\Omega) \tag{18}$$

(where $[Q,S] = 0$ has been used) and

$$Q a^*_{A_i}(\mathbf{p}_i) a^*_{A_{i+1}}(\mathbf{p}_{i+1})\Omega = [Q, a^*_{A_i}(\mathbf{p}_i)] a^*_{A_{i+1}}(\mathbf{p}_{i+1})\Omega + a^*_{A_i}(\mathbf{p}_i) [Q, a^*_{A_{i+1}}(\mathbf{p}_{i+1})]\Omega \quad (i = 1,3), \tag{19}$$

one obtains by means of (16b)

$$H_{A_1A_2A_3A_4,B_1B_2B_3B_4}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4) S_{B_1B_2B_3B_4}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4) = 0, \tag{20}$$

$\forall \mathbf{p}_1,...,\mathbf{p}_4 \in \mathbb{R}^3$, where $H_{A_1 \cdots B_4}(\mathbf{p}_1,...,\mathbf{p}_4)$ are the matrix elements of the $N^4 \times N^4$ matrix [cf. (17)]

$$H(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4) := \overline{H(\mathbf{p}_1)} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} + \mathbf{1} \otimes \overline{H(\mathbf{p}_2)} \otimes \mathbf{1} \otimes \mathbf{1} - \mathbf{1} \otimes \mathbf{1} \otimes H(\mathbf{p}_3) \otimes \mathbf{1} - \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes H(\mathbf{p}_4) \tag{21}$$

($\mathbf{1}$ is the unit $N \times N$ matrix and $\otimes$ stands for the usual tensorial product of matrices), and

$$S_{B_1B_2B_3B_4}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4) := (a^*_{B_1}(\mathbf{p}_1) a^*_{B_2}(\mathbf{p}_2)\Omega | S a^*_{B_3}(\mathbf{p}_3) a^*_{B_4}(\mathbf{p}_4)\Omega) \tag{22}$$

is the elastic two-particle scattering amplitude; it is a tempered distribution with support in the scattering manifold $\mathscr{M}$. Therefore, Eq. (20) is only nontrivial for $(\mathbf{p}_1,...,\mathbf{p}_4) \in \mathscr{M}$. Furthermore, since we are interested in nontrivial scattering, we may restrict our attention to the corresponding reduced scattering amplitude $\mathscr{T}_{B_1B_2B_3B_4}(\mathbf{p}_1,...,\mathbf{p}_4)$ defined in (22) after separation of the trivial scattering $S = \mathbf{1} + iT$ through

$$(a^*_{B_1}(\mathbf{p}_1) a^*_{B_2}(\mathbf{p}_2)\Omega | T a^*_{B_3}(\mathbf{p}_3) a^*_{B_4}(\mathbf{p}_4)\Omega) =: \delta^4(p_1 + p_2 - p_3 - p_4) \mathscr{T}_{B_1B_2B_3B_4}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4). \tag{23}$$

Hence the assumed occurrence of interaction in $\mathscr{H}^{(2)}$ leads from (20) to

$$H_{A_1A_2A_3A_4,B_1B_2B_3B_4}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4) \mathscr{T}_{B_1B_2B_3B_4}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4) = 0, \quad \forall A_1,...,A_4 \in \{1,...,N\}, \quad \forall (\mathbf{p}_1,...,\mathbf{p}_4) \in \mathscr{M}. \tag{24}$$

Let now

$$U(\mathbf{p}) := (U_{A,B}(\mathbf{p}))_{1 \leqslant A,B \leqslant N} \quad (p \in \mathbb{R}^3) \tag{25}$$

be any unitary matrix that diagonalizes the Hermitian matrix $H(\mathbf{p})$ (see Lemma 3.1), i.e.,

$$U(\mathbf{p}) H(\mathbf{p}) U(\mathbf{p})^{-1} = \operatorname{diag}\{\lambda_1(\mathbf{p}),...,\lambda_N(\mathbf{p})\}. \tag{26}$$

Set now

$$U(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4) := \overline{U(\mathbf{p}_1)} \otimes \overline{U(\mathbf{p}_2)} \otimes U(\mathbf{p}_3) \otimes U(\mathbf{p}_4). \tag{27}$$

Then it follows from (26) and (21) that the unitary matrix $U(\mathbf{p}_1,...,\mathbf{p}_4)$ diagonalizes the Hermitian matrix $H(\mathbf{p}_1,...,\mathbf{p}_4)$:

$$\begin{aligned}(U(\mathbf{p}_1,...,\mathbf{p}_4)H(\mathbf{p}_1,...,\mathbf{p}_4)U(\mathbf{p}_1,...,\mathbf{p}_4)^{-1})_{A_1A_2A_3A_4,B_1B_2B_3B_4} \\ = \delta_{A_1B_1}\delta_{A_2B_2}\delta_{A_3B_3}\delta_{A_4B_4} \\ \times(\lambda_{B_1}(\mathbf{p}_1) + \lambda_{B_2}(\mathbf{p}_2) - \lambda_{B_3}(\mathbf{p}_3) - \lambda_{B_4}(\mathbf{p}_4))\end{aligned} \tag{28}$$

(no summation over the B's!).

By means of the diagonalization (28), Eq. (25) (where the reduced scattering amplitudes are viewed as the components of a column vector) transforms to

$$\begin{aligned}(\lambda_{A_1}(\mathbf{p}_1) + \lambda_{A_2}(\mathbf{p}_2) - \lambda_{A_3}(\mathbf{p}_3) - \lambda_{A_4}(\mathbf{p}_4)) \\ \times\mathscr{T}^U_{A_1A_2A_3A_4}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4) = 0, \\ \forall A_1,...,A_4\in\{1,...,N\}, \quad \forall(\mathbf{p}_1,...,\mathbf{p}_4)\in\mathscr{M},\end{aligned} \tag{29}$$

with

$$\begin{aligned}\mathscr{T}^U_{A_1A_2A_3A_4}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4) \\ := \overline{U_{A_1,B_1}(\mathbf{p}_1)}\;\overline{U_{A_2,B_2}(\mathbf{p}_2)}\;U_{A_3,B_3}(\mathbf{p}_3)U_{A_4,B_4}(\mathbf{p}_4) \\ \times\mathscr{T}_{B_1B_2B_3B_4}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4).\end{aligned} \tag{30}$$

From (30) and (23) one obtains

$$\begin{aligned}\delta^4(p_1+p_2-p_3-p_4)\mathscr{T}^U_{A_1A_2A_3A_4}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4) \\ = (U_{A_1,B_1}(\mathbf{p}_1)a^*_{B_1}(\mathbf{p}_1)U_{A_2,B_2}(\mathbf{p}_2)a^*_{B_2}(\mathbf{p}_2)\Omega| \\ \times TU_{A_3,B_3}(\mathbf{p}_3)a^*_{B_3}(\mathbf{p}_3)U_{A_4,B_4}(\mathbf{p}_4)a^*_{B_4}(\mathbf{p}_4)\Omega),\end{aligned} \tag{31}$$

so that, in conformity with assumption 2.1 (i),

$$\mathscr{T}^U_{A_1A_2A_3A_4}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4)$$

can be interpreted as the reduced two-particle scattering amplitude for a new (nontrivial) reaction for all $(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4)\in\mathscr{U}$. This being the case, the optical theorem implies

$$\mathscr{T}^U_{A_1A_2A_1A_2}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_1,\mathbf{p}_2)\neq0, \quad \forall A_1,A_2, \quad \forall(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_1,\mathbf{p}_2)\in\mathscr{U}. \tag{32}$$

The continuity of

$$\mathscr{T}_{B_1B_2B_3B_4}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4)$$

in $\mathscr{U}$ for $1\leqslant B_1,...,B_4\leqslant N$, assumed in 2.1 (ii), leads further from (32) to the following corollary.

Corollary 3.2: There exists an open neighborhood $\mathscr{V}\subset\mathscr{M}$, $\mathscr{V}\cap\mathscr{U}\neq\varnothing$, such that

$$\mathscr{T}^U_{A_1A_2A_1A_2}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4)\neq0 \tag{33}$$

holds $\forall(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4)\in\mathscr{V}$ and $\forall A_1,A_2\in\{1,...,N\}$.

In turn, from (29) and (33) one concludes

$$\lambda_{A_1}(\mathbf{p}_1) + \lambda_{A_2}(\mathbf{p}_2) = \lambda_{A_1}(\mathbf{p}_3) + \lambda_{A_2}(\mathbf{p}_4), \tag{34}$$

$\forall(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3,\mathbf{p}_4)\in\mathscr{V}\subset\mathscr{M}$ (scattering manifold) and $1\leqslant A_1,A_2\leqslant N$. The most general solution of this conditioned functional equation [even for $\lambda_A(\mathbf{p})$ only locally integrable] is[12]

$$\lambda_A(\mathbf{p}) = \alpha_\mu p^\mu + \gamma_A \quad (p^0=\omega_p), \tag{35}$$

with $\alpha_\mu,\gamma_A\in\mathbb{R}$, and $\mathbf{p}\in\mathbb{R}^3$ arbitrary. Plugging (35) back into (26) one obtains

$$U(\mathbf{p})(H(\mathbf{p}) - \alpha_\mu p^\mu\mathbf{1})U(\mathbf{p})^{-1} = \operatorname{diag}\{\gamma_1,...,\gamma_N\}. \tag{36}$$

Lemma 3.3: The relation

$$H(\mathbf{p}) = \alpha_\mu p^\mu\mathbf{1} + \gamma \quad (p^0=\omega_p,\ \alpha_\mu\in\mathbb{R}) \tag{37}$$

holds, where γ is a Hermitian $N\times N$ matrix not depending on $\mathbf{p}$.

Proof: According to the definition (15), the matrix elements of $H(\mathbf{p})$ are linear combinations of products of polynomials in $\omega_p = (\mathbf{p}^2+m^2)^{1/2}$ and $\mathbf{p}$ [remember that $P_{jr,ks}(\omega_p,\mathbf{p})$ are polynomials] times square roots of such polynomials [see (11)]; hence the same holds for the matrix elements of $H(\mathbf{p}) - \alpha_\mu p^\mu\mathbf{1}$.

On the other hand, Eq. (36) spells out that the Hermitian matrix $H(\mathbf{p}) - \alpha_\mu p^\mu\mathbf{1}$ is unitary equivalent to a constant matrix, so that its norm has to be $\mathbf{p}$ independent. Reconciliation of these two facts can only be achieved if the matrix elements of $H(\mathbf{p}) - \alpha_\mu p^\mu\mathbf{1}$ are actually constants. □

Inserting (37) into (16b), one derives the following corollary.

Corollary 3.4: In the interaction case the relation

$$[Q,a^*_A(\mathbf{p})] = (\alpha_\mu p^\mu\delta_{AB} + \gamma_{A,B})a^*_B(\mathbf{p}) \tag{38}$$

holds, with $p^0=\omega_p$, $\alpha_\mu\in\mathbb{R}$, and $\gamma_{A,B} = \overline{\gamma_{B,A}}\in\mathbb{C}$.

Defining now the operator Q_{int} (first) on $\mathscr{H}^{(1)}$ through

$$Q_{\text{int}}a^*_A(\mathbf{p})\Omega := \gamma_{A,B}a^*_B(\mathbf{p})\Omega, \tag{39a}$$

Eq. (38) reads

$$Q = \alpha_\mu P^\mu + Q_{\text{int}}, \tag{39b}$$

where, of course, P^μ is the energy-momentum operator (in $\mathscr{H}^{(1)}$). By virtue of the additivity of Q and P^μ in the whole of the Hilbert space $\mathscr{H}$, Eq. (39b) extends to $\mathscr{H}$ and defines, in this way, Q_{int} on $\mathscr{H}$. As a result, Q_{int} is a translationally invariant local charge that annihilates the vacuum. It will turn out that Q_{int} is the generator of an internal symmetry of the S matrix.

In order to show this, one has to take advantage of the Lorentz covariance of the theory. Let $\Lambda\in\mathscr{L}^\uparrow_+$ be a (proper) Lorentz transformation in Minkowski space, and let $\mathscr{D}(\Lambda)$ denote a unitary representation of SL(2,$\mathbb{C}$) in $\mathscr{H}^{(1)}$ (positive metric!). [Since no confusion will arise, no care has been taken to distinguish Λ from any of its two representants in SL(2,$\mathbb{C}$).] Then

$$\mathscr{D}(\Lambda)a_{jr}(\mathbf{p})\mathscr{D}(\Lambda)^{-1} = \sum_{r'=\sigma(j)}^{-\sigma(j)} D^{\sigma(j)}_{r,r'}(\Lambda_{\mathbf{p}})a_{jr'}(\Lambda\mathbf{p}) \tag{40}$$

holds[11] ($\Lambda\mathbf{p}$ stands for the spatial part of Λp), where $\Lambda_{\mathbf{p}}\in$SU(2) is the so-called Wigner rotation, and the matrices $D^{\sigma(j)}$ span the unitary $(2\sigma(j)+1)$-dimensional irreducible representation of SU(2), the little group of SL(2,$\mathbb{C}$) in the case $P_\mu P^\mu>0$. Its generators $\mathbf{J}^{\sigma(j)}$ were given in (10a) and (10b).

For the purposes of proof it suffices to know the explicit form of (40) for Λ being an infinitesimal boost. In terms of the column vector

$$a^*_j(\mathbf{p}) := (a^*_{j\sigma(j)}(\mathbf{p}),...,a^*_{j-\sigma(j)}(\mathbf{p}))^{\text{transposed}},$$

one has[13]

$$\begin{aligned}\mathscr{D}(\mathbf{1}+\boldsymbol{\vartheta}\cdot\mathbf{B})a^*_j(\mathbf{p})\mathscr{D}(\mathbf{1}-\boldsymbol{\vartheta}\cdot\mathbf{B}) \\ = (\mathbf{1}+i(\boldsymbol{\vartheta}\times\mathbf{p})/(\omega_p+m)\cdot\mathbf{J}^{\sigma(j)})a^*_j(\mathbf{p}-\omega_p\boldsymbol{\vartheta}),\end{aligned} \tag{41}$$

where, of course, the **B** are the corresponding generators of SL(2,$\mathbb{C}$), and the $\boldsymbol{\vartheta}\in\mathbb{R}^3$ are infinitesimal velocities.

In order to stick to the compact notation $A = (j,r)$, $B = (k,s)$,... used so far, define

$$D_{A,B}(\Lambda_{\mathbf{p}}) := \delta_{jk} D^{\sigma(j)}_{r,s}(\Lambda_{\mathbf{p}}). \tag{42}$$

Thus (40) can be rewritten (take its Hermitian conjugate) as

$$\mathscr{D}(\Lambda) a^*_A(\mathbf{p}) \mathscr{D}(\Lambda)^{-1} = D_{A,A'}(\Lambda_{\mathbf{p}}^{-1}) a^*_{A'}(\Lambda\mathbf{p}). \tag{43}$$

Now define Q^Λ to be the local charge obtained from Q by performing the Lorentz transformation Λ upon it, i.e.,

$$Q^\Lambda := \mathscr{D}(\Lambda) Q \mathscr{D}(\Lambda)^{-1}. \tag{44}$$

An easy computation [using (43), (16b), and the identity $(\Lambda^{-1})_{\Lambda\mathbf{p}} = \Lambda_{\mathbf{p}}^{-1}$] shows that Q^Λ acts in $\mathscr{H}^{(1)}$ as

$$[Q^\Lambda, a^*_A(\Lambda\mathbf{p})] = D_{A,A'}(\Lambda_{\mathbf{p}}) H_{A',B'}(\mathbf{p}) D(\Lambda_{\mathbf{p}})^{-1}_{B',B} a^*_B(\Lambda\mathbf{p}). \tag{45}$$

On the other hand, since Q^Λ is a translationally invariant local charge, then, according to (4b),

$$[Q^\Lambda, \Psi_A(x)] = P^\Lambda_{A,B}((1/i)\partial)\Psi_B(x) \tag{46a}$$

holds, where $P^\Lambda_{A,B}((1/i)\partial)$, $1\leqslant A,B \leqslant N$, are polynomials. Inserting now the decomposition (5) into (46a), one gets in turn [cf. (16b) and (15)]

$$[Q^\Lambda, a^*_A(\mathbf{p})] = H^\Lambda_{A,B}(\mathbf{p}) a^*_B(\mathbf{p}), \tag{46b}$$

where the continuous functions

$$H^\Lambda_{jr,ks}(\mathbf{p}) := (e^{\theta\mathbf{p}\cdot\mathbf{J}^{\sigma(j)}})_{r,r'} P^\Lambda_{jr',ks'}(\omega_p,\mathbf{p})(e^{-\theta\mathbf{p}\cdot\mathbf{J}^{\sigma(k)}})_{s',s} \tag{47a}$$

build, analogously to Lemma 3.1, a Hermitian matrix

$$H^\Lambda(\mathbf{p}) := (H^\Lambda_{A,B}(\mathbf{p}))_{1\leqslant A,B\leqslant N}. \tag{47b}$$

The same reasoning that led to Corollary 3.4 obviously applies also to Q^Λ.

Lemma 3.5: In the interaction case the relation

$$H^\Lambda(\mathbf{p}) = \alpha^\Lambda_\mu p^\mu \mathbf{1} + \gamma^\Lambda \quad (p^0 = \omega_p,\ \alpha^\Lambda_\mu \in \mathbb{R}) \tag{48}$$

holds, $\forall\Lambda\in\mathscr{L}^\uparrow_+$, where γ^Λ is a Hermitian $N\times N$ matrix not depending on $\mathbf{p}$.

With $\Lambda = \mathbf{1}$ one recovers Lemma 3.3.

Substitution of (48) in (46b) yields

$$Q^\Lambda = \alpha^\Lambda_\mu P^\mu + Q^\Lambda_{\text{int}} \tag{49}$$

in $\mathscr{H}$, where $Q^\Lambda_{\text{int}}\Omega = 0$,

$$[Q^\Lambda_{\text{int}}, a^*_A(\mathbf{p})] := \gamma^\Lambda_{A,B} a^*_B(\mathbf{p}) \tag{50}$$

in $\mathscr{H}^{(1)}$, and it is additively defined on the rest of $\mathscr{H}$.

Comparing (45) and (46b) one arrives now at

$$H^\Lambda(\Lambda\mathbf{p}) = D(\Lambda_{\mathbf{p}}) H(\mathbf{p}) D(\Lambda_{\mathbf{p}})^{-1}. \tag{51}$$

Substitution of (48) into (51) renders

$$(\alpha^\Lambda_\mu (\Lambda p)^\mu - \alpha_\mu p^\mu)\mathbf{1} = D(\Lambda_{\mathbf{p}})\gamma D(\Lambda_{\mathbf{p}})^{-1} - \gamma^\Lambda. \tag{52}$$

Next, take the trace of (52) and observe that its rhs does not depend on $\mathbf{p}$. Hence

$$\alpha^\Lambda_\mu (\Lambda p)^\mu = \alpha_\mu p^\mu, \quad \text{i.e., } \alpha^\Lambda_\mu = (\Lambda^{-1})^\nu_{\ \mu}\alpha_\nu, \tag{53}$$

and this together with (52) implies

$$\gamma^\Lambda = D(\Lambda_{\mathbf{p}})\gamma D(\Lambda_{\mathbf{p}})^{-1}, \quad \forall\mathbf{p}\in\mathbb{R}^3, \quad \forall\Lambda\in\mathscr{L}^\uparrow_+. \tag{54}$$

Lemma 3.6: From (54) it follows that

$$\gamma_{jr,ks} = \delta_{\sigma(j)\sigma(k)}\delta_{rs}\lambda_{jk}, \quad \text{with } \lambda_{jk} = \overline{\lambda_{kj}}\in\mathbb{C}. \tag{55}$$

Proof: In accordance with Lemma 3.6, γ^Λ does not depend on $\mathbf{p}$, so that from (54) one obtains [see (42)]

$$D^{\sigma(j)}_{r,r'}(\Lambda_{\mathbf{p}})\gamma_{jr',ks'} D^{\sigma(k)}_{s',s}(\Lambda_{\mathbf{p}}^{-1}) = D^{\sigma(j)}_{r,r'}(\Lambda_{\mathbf{0}})\gamma_{jr',ks'} D^{\sigma(k)}_{s',s}(\Lambda_{\mathbf{0}}^{-1}). \tag{56}$$

For any j,k, define γ^{jk} to be the rectangular $(2\sigma(j)+1)\times(2\sigma(k)+1)$ matrix

$$\gamma^{jk} := (\gamma_{jr,ks})_{-\sigma(j)\leqslant r\leqslant\sigma(j),\ -\sigma(k)\leqslant s\leqslant\sigma(k)}, \tag{57}$$

and insert the representation (41) for an infinitesimal boost in (56) in order to derive

$$((\boldsymbol{\vartheta}\times\mathbf{p})\cdot\mathbf{J}^{\sigma(j)})\gamma^{jk} = \gamma^{jk}((\boldsymbol{\vartheta}\times\mathbf{p})\cdot\mathbf{J}^{\sigma(k)}), \tag{58}$$

for any $\mathbf{p}\in\mathbb{R}^3$ and any infinitesimal velocity $\boldsymbol{\vartheta}$. Thus

$$J^{\sigma(j)}_i\gamma^{jk} = \gamma^{jk}J^{\sigma(k)}_i, \quad \forall(j,k), \quad 1\leqslant i\leqslant 3. \tag{59}$$

Since $J^{\sigma(j)}_i$, $J^{\sigma(k)}_i$, $1\leqslant i\leqslant 3$, are generators of SU(2), Eq. (59) amounts to the matrix γ^{jk} intertwining both irreducible representations $D^{\sigma(j)}$ and $D^{\sigma(k)}$ of SU(2). This being the case, Schur's lemma implies (remember that the representations of the same dimension are equal)

$$(\gamma^{jk})_{r,s} = \delta_{\sigma(j)\sigma(k)}\lambda_{jk}\delta_{rs}. \tag{60}$$

□

As a matter of fact, one easily verifies $\gamma^\Lambda_{jr,ks} = \gamma_{jr,ks}$ by substituting (55) in (54), so that [see (50)]

$$[Q^\Lambda_{\text{int}}, a^*_{jr}(\mathbf{p})] = \delta_{\sigma(j)\sigma(k)}\delta_{rs}\lambda_{jk}a^*_{ks}(\mathbf{p}) = [Q_{\text{int}}, a^*_{jr}(\mathbf{p})]. \tag{61}$$

This shows that the local charge Q_{int} is Poincaré invariant.

Putting (55) in (37) and inverting (15), one ends up with

$$P_{jr,ks}(\omega_p,\mathbf{p}) = \delta_{rs}(\delta_{jk}\alpha_\mu p^\mu + \delta_{\sigma(j)\sigma(k)}\lambda_{jk}). \tag{62}$$

All that remains now is to return to the Minkowski space. This proves the following theorem [see (4b)].

Theorem 3.7: Let Q be a translationally invariant local charge, and suppose that the interaction assumption 2.1 is fulfilled. Then

$$[Q,\Psi_{jr}(x)] = \delta_{rs}(\delta_{jk}\alpha_\mu(1/i)\partial^\mu + \delta_{\sigma(j)\sigma(k)}\lambda_{jk})\Psi_{ks}(x), \tag{63}$$

i.e.,

$$Q = \alpha_\mu P^\mu + Q_{\text{int}}, \tag{64}$$

with

$$[Q_{\text{int}},\Psi_{jr}(x)] := \delta_{rs}\delta_{\sigma(j)\sigma(k)}\lambda_{jk}\Psi_{ks}(x), \tag{65}$$

and $\alpha_\mu\in\mathbb{R}$, $\lambda_{jk} = \overline{\lambda_{kj}}\in\mathbb{C}$.

Remark: Translationally invariant local charges are self-adjoint operators.[14] (By the way, this is in general not true for nontranslationally invariant local charges.[7]) Hence it turns out that Q_{int} is actually the generator of a symmetry of the S matrix called an "internal symmetry" because its action on $\mathscr{H}$ is independent of the space-time coordinates.

B. The general case

I continue to use the representations $(\sigma(j),0)$. The starting point is Eq. (3) again. The proof that, in an interacting theory (in the sense of assumption 2.1), the polynomials $P_{jr,ks}(x,(1/x)\partial)$ appearing in (3) are bound to be at most of first degree will be carried out by elimination according to their degree G in x. Of course, $G=0$ corresponds to the translationally invariant case we disposed of in Sec. III A.

(A) $G=1$. Then (3) can be written as

$$[Q,\Psi_{jr}(x)]=\Big\{x_\mu\frac{1}{i}\partial_\nu P^{[\mu,\nu]}_{jr,ks}\Big(\frac{1}{i}\partial\Big)+P_{jr,ks}\Big(\frac{1}{i}\partial\Big)\Big\}\Psi_{ks}(x), \quad (66)$$

where

$$P^{[\mu,\nu]}_{jr,ks}((1/i)\partial)=-P^{[\nu,\mu]}_{jr,ks}((1/i)\partial),$$

and $P_{jr,ks}((1/i)\partial)$ are polynomials in ∂. [The generators of the spin angular momentum $S^{\mu\nu}$ are obviously contained in the constant term of $P_{jr,ks}((1/i)\partial)$.]

In order to reduce the present case $G=1$ to the already solved case $G=0$, define the local charge

$$Q^\rho:=i[Q,P^\rho], \quad (67)$$

which happens to be translationally invariant. In fact,

$$[Q^\rho,\Psi_{jr}(x)]=(1/i)\partial_\nu P^{[\rho,\nu]}_{jr,ks}((1/i)\partial)\Psi_{ks}(x). \quad (68)$$

In a theory with interaction, Theorem 3.7 implies

$$p_\nu P^{[\rho,\nu]}_{jr,ks}(p)=\delta_{rs}(\delta_{jk}\alpha^{\rho\mu}p_\mu+\delta_{\sigma(j)\sigma(k)}\lambda^\rho_{jk}), \quad (69)$$

with $p=(\omega_p,\mathbf{p})$, $\alpha^{\rho\mu}\in\mathbb{R}$, and $\lambda^\rho_{jk}=\overline{\lambda^\rho_{kj}}\in\mathbb{C}$.

Before proceeding further, notice that all the polynomials $P(p)$ in ω_p and $\mathbf{p}$ we are dealing with can be brought by means of the on-mass-shell condition (7) to the form

$$P(p)=\omega_p P^{(1)}(\mathbf{p})+P^{(2)}(\mathbf{p}), \quad (70)$$

with $P^{(1)}(\mathbf{p})$ and $P^{(2)}(\mathbf{p})$ being now polynomials in $\mathbf{p}\in\mathbb{R}^3$ alone. Then one can easily prove the technical lemma that follows.

Lemma 3.8: For all polynomials $P(p)$ of the form (70), $P(p)\equiv 0$ if and only if all its coefficients vanish.

Corollary 3.9: From (69) it follows that

$$\alpha^{\rho\mu}=-\alpha^{\mu\rho}=:\alpha^{[\rho,\mu]} \quad\text{and}\quad \lambda^\rho_{jk}=0. \quad (71)$$

Proof: Multiply (69) times p_ρ, sum over ρ, and apply Lemma 3.8. □

Consequently, (69) gets simplified to

$$p_\nu P^{[\rho,\nu]}_{jr,ks}(p)=\delta_{jk}\delta_{rs}\alpha^{[\rho,\mu]}p_\mu. \quad (72)$$

Inserting (72) in (66), one gets

$$[Q,\Psi_{jr}(x)]=\Big\{\delta_{jk}\delta_{rs}\beta_{\mu\nu}\Big(x^\mu\frac{1}{i}\partial^\nu-x^\nu\frac{1}{i}\partial^\mu\Big)+P_{jr,ks}\Big(\frac{1}{i}\partial\Big)\Big\}\Psi_{ks}(x), \quad (73)$$

with $\beta^{\mu\nu}:=\frac{1}{2}\alpha^{[\mu,\nu]}$.

Since the expression between the curly brackets in (73) has to be a polynomial in the generators of the total angular momentum[7]

$$M^{\mu\nu}=x^\mu(1/i)\partial^\nu-x^\nu(1/i)\partial^\mu+S^{\mu\nu}$$

(as well as in those of the four-momentum), then

$$P_{jr,ks}((1/i)\partial)=\delta_{js}\beta_{\mu\nu}(S^{\mu\nu})_{r,s}+\hat{P}_{jr,ks}((1/i)\partial) \quad (74)$$

must hold.

Hence (73) becomes

$$[Q-\beta_{\mu\nu}M^{\mu\nu},\Psi_{jr}(x)]=\hat{P}_{jr,ks}((1/i)\partial)\Psi_{ks}(x), \quad (75)$$

and one ends up with the local charge $Q-\beta_{\mu\nu}M^{\mu\nu}$ being translationally invariant. Thus, in the case of interaction,

$$Q=\alpha_\rho P^\rho+\beta_{\mu\nu}M^{\mu\nu}+Q_{\rm int} \quad (76)$$

by means of Theorem 3.7.

One sees that the interaction suppresses the appearance of (nontranslationally invariant) local charges, which are not self-adjoint. This is a feature not present in free theories.[7]

(B) It remains to exclude the possibility $G\geqslant 2$. To begin with, one can write

$$P_{jr,ks}\Big(x,\frac{1}{i}\partial\Big)=\frac{1}{G!}(x_{\mu_1}\cdots x_{\mu_G})\Big(\frac{1}{i}\partial_{\nu_1}\cdots\frac{1}{i}\partial_{\nu_G}\Big)\times P^{[\mu_1,\nu_1]\cdots[\mu_G,\nu_G]}_{jr,ks}\Big(\frac{1}{i}\partial\Big)+o(G), \quad (77)$$

where the polynomials

$$P^{[\mu_1,\nu_1]\cdots[\mu_G,\nu_G]}_{jr,ks}((1/i)\partial)$$

are symmetric under the transpositions $\mu_g\leftrightarrow\mu_{g'}$, $\nu_g\leftrightarrow\nu_{g'}$, and

$$[\mu_g,\nu_g]\leftrightarrow[\mu_{g'},\nu_{g'}] \quad (1\leqslant g,g'\leqslant G),$$

and antisymmetric under $\mu_g\leftrightarrow\nu_g$; $o(G)$ denotes all the terms of lower degree in x. With $G=1$ one recovers (66).

Generalizing the technique used in (A), define the (in $\rho_2,\ldots,\rho_G$ symmetric) local charge

$$Q^{\rho_2\rho_3\cdots\rho_G}:=[\cdots[[Q,iP^{\rho_2}],iP^{\rho_3}]\cdots,iP^{\rho_G}]. \quad (78)$$

Taking into account (77), one easily computes

$$[Q^{\rho_2\cdots\rho_G},\Psi_{jr}(x)]=\Big\{x_\mu\Big(\frac{1}{i}\partial_{\nu_1}\cdots\frac{1}{i}\partial_{\nu_G}\Big)P^{[\mu,\nu_1][\rho_2,\nu_2]\cdots[\rho_G,\nu_G]}\Big(\frac{1}{i}\partial\Big)+o(1)\Big\}\Psi_{ks}(x), \quad (79)$$

so that $Q^{\rho_2\cdots\rho_G}$ is a local charge with $G=1$.

Comparison of (79) with (66) shows that the substitution

$$p_{\nu_1}P^{[\mu,\nu_1]}_{jr,ks}(p)\leftarrow(p_{\nu_1}p_{\nu_2}\cdots p_{\nu_G})P^{[\mu,\nu_1][\rho_2,\nu_2]\cdots[\rho_G,\nu_G]}_{jr,ks}(p) \quad (80)$$

($p_0=\omega_p$) is all one needs to transcribe the results of the precedent case $G=1$ for $Q^{\rho_2\cdots\rho_G}$. In particular, the result (72) (valid in the interaction case) becomes, under (80),

$$(p_{\nu_1}\cdots p_{\nu_G})P^{[\rho_1,\nu_1]\cdots[\rho_G,\nu_G]}_{jr,ks}(p)=\delta_{jk}\delta_{rs}\alpha^{\rho_1\cdots\rho_G\mu}p_\mu, \quad (81)$$

where the real constants $\alpha^{\rho_1\cdots\rho_G\mu}$ fulfill the generalization of (71),

$$\alpha^{\rho_1\cdots\rho_i\cdots\rho_G\mu}=-\alpha^{\rho_1\cdots\mu\cdots\rho_G\rho_i} \quad (1\leqslant i\leqslant G), \quad (82)$$

for any $\rho_1,\ldots,\rho_G\in\{0,\ldots,3\}$. Without loss of generality, the polynomials

$$P^{[\rho_1,\nu_1]\cdots[\rho_G,\nu_G]}_{jr,ks}(p)$$

are supposed to be expressed in the "reduced" form (70).

The next lemma ends the proof of Theorem 2.2.

Lemma 3.10: It follows from (81) that

$$P^{[\rho_1,\nu_1]\cdots[\rho_G,\nu_G]}_{jr,ks}(p)\equiv 0 \quad (G\geqslant 2). \tag{83}$$

Proof: In order to apply Lemma 3.8 to (81), one has first to bring its lhs into the form (70). Two cases appear in the summation.

(i) $\#\{\nu_g=0,\ 1\leqslant g\leqslant G\}=2L,\ 0\leqslant L\leqslant \lfloor G/2\rfloor$. Then the lhs of (81) becomes [notation like in (70), $i_g\in\{1,2,3\}$]

$$\begin{aligned}\omega_p^{2L}(p_{i_{2L+1}}\cdots p_{i_G})\{\omega_p P^{(1)\cdots}_{jr,ks}(\mathbf{p})+P^{(2)\cdots}_{jr,ks}(\mathbf{p})\}\\ =\omega_p\{(\mathbf{p}^2+m^2)^L(p_{i_{2L+1}}\cdots p_{i_G})P^{(1)\cdots}_{jr,ks}(\mathbf{p})\}\\ +(\mathbf{p}^2+m^2)^L(p_{i_{2L+1}}\cdots p_{i_G})P^{(2)\cdots}_{jr,ks}(\mathbf{p}).\end{aligned} \tag{84a}$$

(ii) $\#\{\nu_g=0,\ 1\leqslant g\leqslant G\}=2L+1,\ 0\leqslant L\leqslant \lfloor (G-1)/2\rfloor$. Analogously to (i), one obtains for the lhs of (81)

$$\begin{aligned}\omega_p\{(\mathbf{p}^2+m^2)^L(p_{i_{2L+2}}\cdots p_{i_G})P^{(2)\cdots}_{jr,ks}(\mathbf{p})\}\\ +(\mathbf{p}^2+m^2)^{L+1}(p_{i_{2L+2}}\cdots p_{i_G})P^{(1)\cdots}_{jr,ks}(\mathbf{p}).\end{aligned} \tag{84b}$$

Since we are assuming $G\geqslant 2$, the only possibility for both (84a) and (84b) to fulfill (81) without each $P^{(1)\cdots}_{jr,ks}(\mathbf{p})$ and $P^{(2)\cdots}_{jr,ks}(\mathbf{p})$ having to vanish identically (invoke Lemma 3.8) is that some summands on the lhs of (81) cancel among them. But this can easily be shown to be impossible due to the symmetry properties of $p_{\nu_1}\cdots p_{\nu_G}$ and

$$P^{[\rho_1,\nu_1]\cdots[\rho_G,\nu_G]}_{jr,ks}(\mathbf{p})$$

under transpositions of indices. □

IV. CONCLUDING REMARKS

(a) The Poincaré invariance of the generators of internal symmetries Q_{int} [see (61)] can be expressed as

$$[Q_{\text{int}},P^\rho]=0,\quad [Q_{\text{int}},M^{\mu\nu}]=0\quad (0\leqslant\rho,\mu,\nu\leqslant 3). \tag{85}$$

In a group-theoretical setting, the commutation relations (85), together with the statement of Theorem 2.2, Eq. (76), amount to the decomposition of the Lie algebra $\mathfrak{s}$ of those symmetries of the S matrix that are generated by local charges into two commutative subalgebras, namely, the Lie algebra of the Poincaré group $\mathfrak{p}$ and the Lie algebra $\mathfrak{s}_{\text{int}}$, which is generated by all the ($\mathbb{R}$-) linearly independent Q_{int} (supposing that their number is finite), i.e.,

$$\mathfrak{s}=\mathfrak{p}\oplus\mathfrak{s}_{\text{int}}. \tag{86}$$

It follows the local isomorphy of the corresponding Lie groups. By the way, it was in this form that Coleman and Mandula stated their result (with the difference that they allow for infinite-dimensional groups of symmetries of the S matrix).

(b) Another consequence of the Poincaré invariance of Q_{int} is that it prevents Q_{int} from containing any linear combination of the generators of the spin angular momentum. This shows that, in case of interaction, there is in general no separately conserved spin angular momentum (in a local massive relativistic quantum field theory).

(c) The most important assumptions underlying the proof of Theorem 1.2 are (i) interaction (in the form of assumption 2.1), (ii) Lorentz invariance, (iii) finite particle multiplets, (iv) massive particles, and (v) invariance of the vacuum.

Theorem 2.2 is known to be false whenever any of the three first assumptions is not satisfied.[1] As for massless theories, their investigation has not been completed yet, the main difficulty being the proof of an expression analogous to (3). (In the massless case one expects the generators of the conformal group to replace those of the Poincaré group in (3), but this does not at all entail that Theorem 2.2 cannot hold any more![15]) With regard to theories with spontaneous symmetry breaking, no result about the structure of the symmetries of the S matrix in such theories is known to me. Yet in non-Abelian gauge theories with magnetic monopoles, internal and space-time symmetries do couple ("spin from isospin"[16]).

Finally, it should be pointed out that assumptions (iv) and (v) are quite restrictive as far as the phenomenological application of Theorem 2.2 (and of the Coleman–Mandula theorem, for that case) is concerned. As a matter of fact this observation carries over, in particular, to the supersymmetries of the S matrix, for their structure rests completely on the Coleman–Mandula result.[2] In spite of this, these superalgebras are often applied as though they were universally valid. One should keep this shortcoming in mind whenever supersymmetric models are proposed for describing real physics.

ACKNOWLEDGMENTS

I wish to express my gratitude to Professor H. Reeh for his advice and guidance throughout this research, as well as for his constructive criticism on reading the original manuscript. During the first stage of this work I benefited also from discussions with Professor J. Łopuszański. Finally, I acknowledge the kind hospitality dispensed to me during the academic year 1985–1986 at the Physics Department of UCLA, where part of this work was done.

[1] S. Coleman and J. Mandula, Phys. Rev. **159**, 1251 (1967).
[2] R. Haag, J. Łopuszański, and M. Sohnius, Nucl. Phys. B **88**, 257 (1975).
[3] C. Orzalesi, J. Sucher, and C. H. Woo, Phys. Rev. Lett. **21**, 1550 (1968); J. Łopuszański, J. Math. Phys. **12**, 2401 (1971).
[4] W. D. Garber and H. Reeh, Commun. Math. Phys. **70**, 169 (1979).
[5] H. Reeh, in *Proceedings of the XIII International Conference on Differential Geometric Methods in Theoretical Physics*, Shumen, Bulgaria, 1984, edited by H. D. Doebner and T. D. Palev (World Scientific, Singapore, 1986), p. 143.
[6] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, Reading, MA, 1980).
[7] W. D. Garber and H. Reeh, Commun. Math. Phys. **67**, 179 (1979).
[8] W. D. Garber and H. Reeh, J. Math. Phys. **19**, 59 (1978).
[9] S. Weinberg, in *Lectures on Particles and Field Theory*, Vol. 2, Brandeis Summer Institute in Theoretical Physics, 1964, edited by S. Deser and K. W. Ford (Prentice–Hall, Englewood Cliffs, NJ, 1965), p. 405.
[10] J. Schwartz, J. Math. Phys. **2**, 271 (1961).
[11] J. Łopuszański, Fortschr. Phys. **26**, 261 (1978).
[12] J. M. Amigó and H. Reeh, J. Math. Phys. **24**, 1594 (1983).
[13] C. Fronsdal, Phys. Rev. **113**, 1367 (1959).
[14] K. Kraus and L. J. Landau, Commun. Math. Phys. **24**, 243 (1972).
[15] D. Strube, J. Phys. A: Math. Gen. **18**, 2603 (1985).
[16] P. Hasenfratz and G. 't Hooft, Phys. Rev. Lett. **36**, 1119 (1976); R. Jackiw and C. Rebbi, Phys. Rev. Lett. **37**, 172 (1976).