Spaces of analytic functions on a complex cone as carriers for the symmetric tensor representations of SO(n)

V. Bargmann*

Departments of Mathematics and Physics, Princeton University, Princeton, New Jersey 08540

I. T. Todorov†

The Institute for Advanced Study, Princeton, New Jersey 08540

(Received 28 July 1976)

We study the space P of all polynomial functions on the complex cone

$$\mathbb{K}_n = \{ z = (z_1 \cdots z_n) \in \mathbb{C}^n, z^2 = z_1^2 + \cdots + z_n^2 = 0 \} \quad (n = 3,4,\cdots).$$

Its subspaces K^l ($= K_n^l$) of homogeneous polynomials of degree l ($= 0,1,2,\cdots$) provide a convenient realization of the carrier spaces for the symmetric tensor representations of the real orthogonal group SO(n). The multiplication operator z_μ ($\mu = 1,\ldots,n$) maps K^l into K^{l+1}. We define its adjoint as an interior differential operator on $\mathbb{K}_n$ which maps K^{l+1} into K^l and transforms as an n-vector. We show that the lowest order differential operator with this property is proportional to $D_\mu = (n/2 - 1 + z\,\partial)\partial_\mu - (1/2)z_\mu\Delta$. We define a scalar product in P with respect to which the operators z_μ and D_μ are Hermitian adjoint to each other and consider the Hilbert space completion K_n of P with respect to this scalar product. The spaces K_n are imbedded for all $n(= 3,4,\cdots)$ in the Fock type spaces B_n, studied earlier by Bargmann. The space K_n possesses a reproducing kernel that allows us to define a (unique) harmonic extension of every analytic function in K_n. It is shown that the spaces K_3 and K_4 can be imbedded isometrically in the Hilbert spaces B_2 and B_4 associated with the representations of SU(2) and SU(2)×SU(2) [$\subset$SU(4)].

1. INTRODUCTION AND SUMMARY

It was recognized since the early days of representation theory that symmetric tensor representations are conveniently described in terms of homogeneous polynomials of a single (complex) vector variable. For example, the SU(n)-symmetric (rank r) tensors $\psi^{\alpha_1\cdots\alpha_r}$ ($\alpha_i = 1,\ldots,n$) are in one-to-one correspondence with the homogeneous polynomials (of degree r)

$$\psi(z) = (r!)^{-1/2}\sum_\alpha \psi^{\alpha_1\cdots\alpha_r} z_{\alpha_1}\cdots z_{\alpha_r}, \tag{1.1}$$

where $\sum_\alpha$ stands for an r-fold sum over the indices $\alpha_1,\ldots,\alpha_r$ ($1 \leq \alpha_i \leq n$). The set of all such polynomials is denoted by B^r ($= B_n^r$).

The rank r symmetric tensor representation S^r of SU(n) acts as a coordinate transformation on the functions $\psi \in B^r$:

$$[S^r(u)\psi](z) = \psi(u^{-1}z). \tag{1.2}$$

This realization has the following attractive algebraic property: The representation space $B^{r_1+r_2}$ for the symmetric tensor product representation $S^{r_1}\otimes S^{r_2} = S^{r_1+r_2}$ is spanned by the ordinary products $\psi_1(z)\psi_2(z)$ of functions $\psi_i(z) \in S^{r_i}$ ($i = 1,2$)—no symmetrization is needed.

The usual SU(n)-covariant scalar product of two tensors can be expressed in terms of the polynomials (1.1), if we postulate that the monomials

$$\frac{z_1^{k_1}\cdots z_n^{k_n}}{\sqrt{k_1!\cdots k_n!}} \quad (k_1 + \cdots + k_n = r) \tag{1.3}$$

form an orthonormal basis in the space B^r.

It was previously shown [1,2] that these rules are satisfied if we define the scalar product

$$\langle \phi, \psi \rangle = \int \overline{\phi(z)}\, \psi(z) d\mu_n(z),$$
$$d\mu_n(z) = \frac{1}{\pi^n}\exp(-\bar{z}z) d^2z_1\cdots d^2z_n \tag{1.4}$$
$$(\text{for } z = x + iy,\ d^2z = dx\,dy)$$

in the space $B(= B_n)$ of all entire analytic functions $\psi(z)$ satisfying

$$\langle \psi, \psi \rangle < \infty. \tag{1.5}$$

The Hilbert space B arises naturally in the Fock realization of the canonical commutation relations. The point is that the operators z_k and $\partial/\partial z_k$ are Hermitian conjugate to each other with respect to the scalar product (1.4):

$$\langle \phi, z_k\psi \rangle = \left\langle \frac{\partial}{\partial z_k}\phi, \psi \right\rangle \quad \left(\text{i.e., } z_k^* = \frac{\partial}{\partial z_k}\right). \tag{1.6}$$

A useful property of the space B is the existence of a *reproducing kernel* (see Ref. 3), the function $e^{\bar{w}z}$, which satisfies

$$\psi(z) = \int e^{\bar{w}z}\psi(w) d\mu_n(w) \quad \text{for every } \psi \in B. \tag{1.7}$$

The space B is the (topological) direct sum of the spaces B^r of irreducible symmetric (rank r) tensor representations of SU(n).

The objective of the present note is to work out a similar homogeneous and analytic function technique for the symmetric tensor representations of the orthogonal group SO(n) ($n \geq 3$). The technique is applied elsewhere[4] to study the irreducible representations (and intertwining operators) of the pseudo-orthogonal group SO($n+1,1$).

In contrast with the case of SU(n), an SO(n) symmetric tensor $f^{\mu_1\cdots\mu_l}$ is only irreducible if it is *traceless*. There are two related ways to account for this tracelessness in the language of homogeneous polynomials

$$f(\xi)=\sum_{\mu} f^{\mu_1\cdots\mu_l}\,\xi_{\mu_1}\cdots\xi_{\mu_l}. \tag{1.8}$$

Note that symmetric traceless tensors correspond to homogeneous *harmonic* polynomials $f(\xi)$, satisfying

$$\Delta f(\xi)=\Delta_{\xi} f(\xi)\equiv\sum_{j=1}^{n}\frac{\partial^2}{\partial\xi_j^2}f(\xi)=0. \tag{1.9}$$

Alternatively, we could represent any symmetric traceless tensor as the restriction of $f(\xi)$ to the complex "light cone"

$$\mathbb{K}_n=\{Z\in\mathbb{C}^n,\ Z^2\equiv Z_1^2+\cdots+Z_n^2=0\}. \tag{1.10}$$

The second possibility leads to simpler formulas, as illustrated in the discussion of Sec. 2 below. [Compare, for example, the expressions in the right-hand sides of Eqs. (2.17) and (2.18), which contain the same information about the underlying symmetric traceless tensor.] The algebraic basis for this simplification is again displayed in the realization of the semigroup of symmetric tensor representations. The point is that the product of two homogeneous polynomial functions of degrees l_1 and l_2 on the cone (1.10) is again a homogeneous polynomial on the cone (of degree l_1+l_2), while the product of two harmonic polynomials is not necessarily a harmonic polynomial.

Let K^l $(=K_n^l)$ be the space of all homogeneous polynomial functions of degree l on the cone (1.10). We shall consider (in Sec. 3) the *graded algebra*

$$P=P(\mathbb{K}_n)=\bigoplus_{l=0}^{\infty}K^l$$

of all polynomial functions on the cone $\mathbb{K}_n$.

We define the reducible representation

$$U(\Lambda)f(Z)=f(\Lambda^{-1}Z),\quad \Lambda\in\mathrm{SO}(n),\ f\in P, \tag{1.11}$$

of SO(n) which generates all symmetric tensor representations U^l, given by the restriction of U to K^l:

$$U=\bigoplus_{l=0}^{\infty}U^l.$$

The operator of multiplication by (any one of) the coordinates Z_μ is a mapping from K^l into K^{l+1}. We define Z_μ^* as the suitably normalized lowest order interior differential operator D_μ on the cone $\mathbb{K}$ which maps K^{l+1} into K^l and behaves under rotation as the μth component of an n-vector. It is shown that these conditions determine D_μ uniquely up to an over-all normalization. Moreover, the operators Z_μ, Z_ν^* and their commutators generate the Lie algebra of a unitary representation of the conformal group SO($n,2$), and the normalization of Z^* is chosen to fit the canonical commutation relation

$$[Z_\mu^*,Z_\nu]=X\delta_{\mu\nu}+X_{\mu\nu},\quad \mu,\nu=1,\ldots,n, \tag{1.12}$$

where $X_{\mu\nu}$ are the mathematical (anti-Hermitian) generators of rotations, viz.,

$$X_{\mu\nu}=Z_\nu\partial_\mu-Z_\mu\partial_\nu \tag{1.13a}$$

while

$$X=h-1+Z\partial\quad (Z\partial=\textstyle\sum_\mu Z_\mu\partial_\mu,\ h=n/2) \tag{1.13b}$$

is the Hermitian generator of rotations in the plane $(n+1,n+2)$, which plays the role of the physical dilatation generator in the Z-space. (The quantum mechanical angular momentum operators would be $i\,X_{\mu\nu}$.)

We construct (in Sec. 4) a scalar product (f,g) in P with respect to which the operators Z_μ and $Z_\mu^*=D_\mu$ are (Hermitian) adjoint to each other. It is shown that in the subspace K^l this scalar product is proportional to the scalar product (1.4) of the harmonic extensions of f and g. The construction of Secs. 3 and 4 gives a natural solution of the quantization problem for the harmonic oscillator with the constraints $Z^2=0=Z^{*2}$. The scalar product (f,g) is defined in terms of a measure m_n with support in the cone $\mathbb{K}_n$. The space K_n possesses a reproducing kernel $\delta_n(\xi,Z)$ with the following property: If $f(Z)$ is any analytic function on $\mathbb{K}_n$ with finite norm, then the scalar product

$$f(\xi)=(\overline{\delta}_n(\xi,\cdot),f) \tag{1.14}$$

provides a harmonic extension of $f(Z)$ to the entire complex space $\mathbb{C}^n(\ni\xi)$. A covariant isometric mapping of K_3 into B_3 and of K_4 into B_4 is defined in Sec. 5.

2. HOMOGENEOUS POLYNOMIALS ON THE ISOTROPIC CONE AND THEIR HARMONIC EXTENSION

In this section we shall present a constructive proof and some applications of the following classical result.[5]

Let B^l be the space of all homogeneous polynomials of $\xi\in\mathbb{C}^n$ of degree l and let $\mathcal{H}_l$ be the subspace of homogeneous harmonic polynomials.

Theorem 2.1: Every homogeneous polynomial $f(\xi)\in B^l$ has a unique expansion of the form

$$f(\xi)=\sum_{0\le 2k\le l}(\xi^2)^k Y_{l-2k}(\xi) \tag{2.1}$$

where $Y_{l-2k}(\xi)\in\mathcal{H}_{l-2k}$.

Proof: A basic role in our derivation of (2.1) will be played by the Casimir operator

$$\Omega=\tfrac{1}{2}\sum_{\mu,\nu}X_{\mu\nu}X_{\nu\mu}\quad (X_{\mu\nu}=\xi_\nu\partial_\mu-\xi_\mu\partial_\nu) \tag{2.2}$$

of SO(n). It can be expressed in the following way in terms of the dilatation generator $X=\xi\partial+h-1$ [(1.13b)] and the Laplacian Δ [(1.9)]:

$$\Omega=X^2-(h-1)^2-\xi^2\Delta. \tag{2.3}$$

For $f\in B^m$, $Xf=(m+h-1)f$. On the other hand Ω commutes with ξ^2 because of rotation invariance. It follows that

$$(\xi^2)^k Y_{l-2k}(\xi),\quad k=0,1,\ldots,[l/2], \tag{2.4}$$

are eigenfunctions of Ω in B^l, corresponding to the eigenvalues ω_{l-2k}, where

$$\omega_m=(m+h-1)^2-(h-1)^2=m(m+n-2). \tag{2.5}$$

It will be shown later that all eigenfunctions of Ω have the form (2.4) [see Eqs. (2.8)—(2.10) below]. The projection operator on the subspace of functions of the form (2.4) (for fixed l and k) will be denoted by P^{lk}.

Lemma 2.2: For each positive integer $m \leq [\,l/2]+1$

$$(\zeta^2)^m \Delta^m f = \prod_{j=0}^{m-1} (\omega_{l-2j} - \Omega) f \quad (f \in B^l). \tag{2.6}$$

We shall prove the lemma by induction on m. For $m=1$, (2.6) follows immediately from (2.3) and (2.5). Assume that it is proven for some $m \leq [l/2]$ and apply to both sides the operator $\omega_{l-2m} - \Omega$. For the left-hand side we find

$$(\zeta^2)^m (\omega_{l-2m} - \Omega)\Delta^m f = (\zeta^2)^m (\zeta^2 \Delta)\Delta^m f$$

(since $\Delta^m f \in B^{l-2m}$). The lemma follows.

Apply (2.6) to the special case $m=[l/2]+1$. Then $2m > l$ and, therefore, $\Delta^m f = 0$. Hence

$$\prod_{j=0}^{[l/2]} (\omega_{l-2j} - \Omega) f = 0 \quad \text{for } f \in B^l. \tag{2.7}$$

Thus, the linear operator Ω on B^l satisfies an algebraic equation with simple roots. Therefore, every f has a unique decomposition

$$f = \sum_{k=0}^{[l/2]} f_k, \quad (\Omega - \omega_{l-2k}) f_k = 0, \quad f_k = P^{lk} f. \tag{2.8}$$

It follows from (2.8) that the projection operator P^{lk} equals

$$P^{lk} = \prod_{\substack{0 \leq j \leq [l/2] \\ j \neq k}} \frac{\Omega - \omega_{l-2j}}{\omega_{l-2k} - \omega_{l-2j}}. \tag{2.9}$$

Let $m \leq [l/2]$. Using (2.6), we find that

$$(\zeta^2)^m \Delta^m f_m = \kappa f_m, \quad \text{where } \kappa = \prod_{j=0}^{m-1} (\omega_{l-2j} - \omega_{l-2m}) \neq 0.$$

Thus

$$f_m = (\zeta^2)^m g, \quad g \in B^{l-2m}. \tag{2.10}$$

On the other hand, because of (2.8),

$$0 = (\omega_{l-2m} - \Omega) f_m = (\zeta^2)^m (\omega_{l-2m} - \Omega) g = (\zeta^2)^{m+1} \Delta g.$$

Hence $\Delta g = 0$, so that $g = Y_{l-2m} \in H_{l-2m}$. This completes the proof of Theorem 2.1.

It is convenient to express the projection operator P^{l0} (as well as P^{lk}) in a different form. To this end we employ the following elementary algebraic identity.

Lemma 2.3: Let $\alpha_j (j = 0, 1, \ldots, m)$ be distinct numbers ($\alpha_i \neq \alpha_j$ if $i \neq j$). Consider two sets of polynomials $\phi_j(x)$ and $\psi_j(x)$ of one variable defined by

$$\phi_0(x) = \Psi_0(x) = 1, \quad \phi_k(x) = \prod_{j=1}^{k} \frac{x - \alpha_j}{\alpha_0 - \alpha_j},$$

$$\Psi_k(x) = \prod_{j=1}^{k} \frac{x - \alpha_{j-1}}{\alpha_0 - \alpha_j}, \quad k = 1, \ldots, m. \tag{2.11}$$

Then

$$\phi_k(x) = \sum_{j=0}^{k} \psi_j(x). \tag{2.12}$$

Proof: The lemma is easily verified by induction, noticing that

$$\phi_k(x) + \Psi_{k+1}(x) = \phi_k(x) \left(1 + \frac{x - \alpha_0}{\alpha_0 - \alpha_{k+1}}\right) = \phi_{k+1}(x).$$

Proposition 2.4: The projection operator P^{l0} is given by

$$P^{l0} = \sum_{j=0}^{[l/2]} \gamma_{lj} (\zeta^2)^j \Delta^j, \tag{2.13}$$

where

$$\gamma_{lj} = (4^j j!\,(2 - l - h)_j)^{-1}, \quad (a)_j \equiv \Gamma(a+j)/\Gamma(a). \tag{2.14}$$

Proof: We apply Lemma 2.3 with $\alpha_j = \omega_{l-2j}$ and $m = [l/2]$. Then

$$\phi_k(\Omega) = \sum_{j=0}^{k} \psi_j(\Omega)$$

and in view of (2.5), (2.6), (2.11)

$$\psi_j(\Omega) = \prod_{i=1}^{j} \frac{\Omega - \omega_{l-2i+2}}{\omega_l - \omega_{l-2i}}$$

$$= (-1)^j \Big(\prod_{i=1}^{j} [(l+h-1)^2 - (l-2i+h-1)^2]\Big)^{-1} (\zeta^2)^j \Delta^j$$

$$= \gamma_{lj} (\zeta^2)^j \Delta^j.$$

The representation (2.13) now follows from (2.9) by setting $k = [l/2]$, since $P^{l0} = \phi_{[l/2]}(\Omega)$.

Remark: The projection operators P^{lk} can also be written in a similar fashion. It suffices to notice that Eq. (2.9) can be rewritten in the form

$$P^{lk} = P^{l-2k,0} A_k$$

where

$$A_k f = \prod_{j=0}^{k-1} \frac{\Omega - \omega_{l-2j}}{\omega_{l-2k} - \omega_{l-2j}} f = (-1)^k \gamma_{l+1-k,k} (\zeta^2)^k \Delta^k f,$$

$$P^{lk} f = (-1)^k \gamma_{l+1-k,k} (\zeta^2)^k (P^{l-2k,0} \Delta^k f)$$

$$(f \in B^l, \text{ hence } \Delta^k f \in B^{l-2k}).$$

Proposition 2.4 allows us to write down a harmonic extension for any (polynomial) function $f(Z) \in K^l$ defined on the cone $\mathbb{K}_n$ [(1.10)]. Indeed, if $f_E(\zeta)$ is any homogeneous polynomial extension of $f(Z)$ ($\zeta \in \mathbb{C}^n$), then

$$Y_f(\zeta) = P^{l0} f_E(\zeta) \tag{2.15}$$

would provide us with a harmonic extension [since $Y_f(Z) = f(Z)$ on $\mathbb{K}_n$ because $\gamma_{l0} = 1$ in Eq. (2.14)]. The following proposition shows that the result does not depend on the intermediate extension f_E.

Proposition 2.5: Each polynomial $f(\zeta) \in K^l$ has a unique harmonic extension $Y_f(\zeta)$.

Proof: Assume that Y' and Y'' are two different harmonic extensions of f. Then the homogeneous harmonic polynomial $Y = Y' - Y''$ vanishes on the cone $\mathbb{K}_n$. On the other hand, if a polynomial Y vanishes on $\mathbb{K}_n$, it has necessarily the form

$$Y(\zeta) = \zeta^2 g(\zeta), \quad \text{where } g(\zeta) \in B^{l-2}. \tag{2.16}$$

We apply to both sides of (2.16) the operator

$$\prod_{j=0}^{[l/2]-1} (\omega_{l-2-2j} - \Omega)$$

and use (2.7). Since by assumption Y is harmonic, we obtain

$$\prod_{j=0}^{[l/2]-1} (\omega_{l-2-2j} - \omega_l) Y = 0.$$

Hence $Y = 0$. This completes the uniqueness proof.

Proposition 2.5 establishes an isomorphism between the space K^l of polynomial functions on $\mathbb{K}_n$ and the

space $\mathcal{H}_l$ of their harmonic extensions. The following simple example plays an important part in the applications.

Example 2.6: Let

$$f(Z)=(aZ)^l \quad (aZ=\textstyle\sum_\mu a_\mu Z_\mu). \tag{2.17}$$

The harmonic extension of f is given by

$$Y_f(\xi)=H_l(a,\xi)=(a\xi)^l F\left(-\frac{l}{2},\frac{1-l}{2},\,2-h-l;\,\frac{a^2\xi^2}{(a\xi)^2}\right)$$

$$=\frac{l!}{(h-1)_l}\left(\tfrac{1}{4}a^2\xi^2\right)^{l/2}C_l^{h-1}\left(\frac{a\xi}{(a^2\xi^2)^{1/2}}\right), \tag{2.18}$$

where $C_l^\nu(x)$ is the Gegenbauer polynomial defined as the regular solution of the differential equation

$$\left((1-x^2)\frac{d^2}{dx^2}-(2\nu+1)x\frac{d}{dx}+l(l+2\nu)\right)C_l^\nu(x)=0,$$

satisfying the normalization condition

$$\frac{1}{2^l(\nu)_l}\frac{d^l}{dx^l}\,C_l^\nu(x)=1;$$

$F(\alpha,\beta,\gamma;x)$ is the hypergeometric function.

To prove (2.18), we apply (2.13) and use the relations

$$(\xi^2)^j\Delta^j(a\xi)^l=\frac{l!}{(l-2j)!}\,(a^2\xi^2)^j(a\xi)^{l-2j},$$

$$\left(-\frac{l}{2}\right)_j\left(\frac{1-l}{2}\right)_j=\frac{l!}{4^j(l-2j)!}.$$

3. INTERIOR DIFFERENTIAL OPERATORS ON P. THE LIE ALGEBRA GENERATED BY RAISING AND LOWERING OPERATORS

We introduce the notion of an interior differential operator on the cone $\mathbb{K}_n$ (cf., e.g., Ref. 6). In order to keep the discussion on a purely algebraic level, we shall only consider (finite order) differential operators

$$Q=a(Z)+\textstyle\sum_\mu b_\mu(Z)\partial_\mu+\sum_{\mu,\nu}c_{\mu\nu}(Z)\partial_\mu\partial_\nu+\cdots,\ \partial_\mu\equiv\frac{\partial}{\partial Z_\mu} \tag{3.1}$$

with polynomial coefficients $a(Z)$, $b_\mu(Z)$, $c_{\mu\nu}(Z),\cdots$. We shall say that Q defines an *interior* differential operator on $\mathbb{K}_n$, if for every polynomial $f(Z)$,

$$(QZ^2f(Z))\,|_{Z^2=0}=0. \tag{3.2}$$

The interior operators on P form a (complex) algebra under addition and multiplication. To verify that if Q_1 and Q_2 are interior, then Q_1Q_2 is also interior, we notice that (3.2) is equivalent to

$$QZ^2f(Z)=Z^2f_Q(Z)\quad(f_Q\in P). \tag{3.2'}$$

For multiplication by a polynomial $a(Z)$ (i.e., for zeroth order differential operators) Eq. (3.2) imposes no restriction. The generators of rotations and dilatation (1.13) provide examples of first order interior differential operators which leave the spaces K^l invariant. The most general form of an operator with these properties is $Q=a+bX+\sum_{\mu\nu}\omega_{\mu\nu}X_{\mu\nu}$ where a, b, and $\omega_{\mu\nu}$ are complex numbers and $\omega_{\mu\nu}=-\omega_{\nu\mu}$. The multiplication by Z_μ is the simplest example of a *raising operator*, which maps K^l into K^{l+1}. We look for an adjoint *lowering operator* D_μ, satisfying the following conditions: (i) D_μ is an interior operator which maps K^{l+1} into K^l (and annihilates K^0); (ii) D_μ behaves as an n-vector under rotations; in particular,

$$[X_{\lambda\mu},D_\nu]=\delta_{\lambda\nu}D_\mu-\delta_{\mu\nu}D_\lambda; \tag{3.3}$$

(iii) D_μ is the lowest order nontrivial operator satisfying (i) and (ii).

Remark: If the variables Z_μ were independent, i.e., if we did not have the constraint $Z^2=0$, then the solution to our problem would have been proportional to $\partial_\mu(=\partial/\partial Z_\mu)$. In that case Z_μ and D_ν would satisfy the commutation relations of creation and annihilation operators. The solution of the above problem will provide a solution of the quantization problem for the oscillator Hamiltonian $H=\frac{1}{2}(p^2+q^2)$ with constraints $p^2-q^2=0=pq+qp$ $[p=(p_1\cdots p_n),\ q=(q_1\cdots q_n)]$. The connection between the two problems becomes obvious, if we set $Z_\mu=(1/\sqrt{2})(q_\mu-ip_\mu)$. Our approach avoids the use of nonpolynomial expressions encountered in the conventional procedure[7,8]; the price is that we introduce from the outset the operators ∂_μ, which do not have a simple interpretation in terms of the canonical variables, and then express $Z_\mu^*(=D_\mu)$ as functions of Z_μ and ∂_μ.

It is easily seen that there is no first order interior differential operator which maps K^{l+1} into K^l. We shall construct the most general second order operator satisfying (i) and (ii) [and hence, also (iii)].

The most general second order operator mapping K^{l+1} into K^l and satisfying the covariance condition (ii) is

$$D_\mu=a\partial_\mu+b(Z\partial)\partial_\mu+cZ_\mu\Delta,$$

where a,b, and c are complex numbers. The condition

$$[D_\mu,Z^2]\,|_{Z^2=0}$$

$$=2\{(a+b+nc)Z_\mu+(b+2c)Z_\mu(Z\partial)+bZ^2\partial_\mu\}\,|_{Z^2=0}=0$$

implies $a+b+nc=0=b+2c$; thus

$$D_\mu=b\{(h-1+Z\partial)\partial_\mu-\tfrac{1}{2}Z_\mu\Delta\}=b(X\partial_\mu-\tfrac{1}{2}Z_\mu\Delta). \tag{3.4}$$

This is the general solution to our problem. The operators D_μ commute with each other, and D^2 vanishes. We have

$$[Z_\mu,Z_\nu]=0,\quad [D_\mu,D_\nu]=0, \tag{3.5}$$

$$Z^2=0,\quad D^2=\textstyle\sum_\mu D_\mu^2=0. \tag{3.6}$$

We shall choose the normalization

$$b=1; \tag{3.7}$$

then Z_μ, D_ν, $X_{\mu\nu}$, and X satisfy the commutation relations (1.12)

$$[D_\mu,Z_\nu]=X\delta_{\mu\nu}+X_{\mu\nu} \tag{3.8}$$

[as well as (3.3)] of the conformal Lie algebra $so(n,2)$ (*cf.* Ref. 9). The algebra of all interior (polynomial) differential operators Q [(3.1)] appears as the universal enveloping algebra $\mathcal{A}$ of $so(n,2)$.

The operators D_μ allow us to write down an alternative form of the harmonic extension (2.15) of a homogeneous

polynomial $f(Z)\in K^l$:

$$f(\xi)=\frac{1}{l!\,(h-1)_l}(\xi D)^l f(Z). \tag{3.9}$$

Since we have established in Corollary 2.2 the existence and uniqueness of a harmonic extension, it is sufficient to verify (3.9) as an identity for harmonic functions on $\mathbb{C}^n$. For these it follows from the simple differentiation formula

$$\xi D\sum_\mu f^{\mu_1\cdots\mu_l}Z_{\mu_1}\cdots Z_{\mu_\nu}\xi_{\mu_{\nu+1}}\cdots\xi_{\mu_l}$$
$$=\nu(h+\nu-2)\sum_\mu f^{\mu_1\cdots\mu_l}Z_{\mu_1}\cdots Z_{\mu_{\nu-1}}\xi_{\mu_\nu}\cdots\xi_{\mu_l}. \tag{3.10}$$

4. THE HILBERT SPACE K_n. REPRODUCING KERNEL AND HARMONIC EXTENSION OF ANALYTIC FUNCTIONS ON THE CONE

Now we shall construct a scalar product in P such that the Hermitian conjugate to the multiplication operator Z_μ is given by

$$Z_\mu^*=D_\mu=X\partial_\mu-\tfrac{1}{2}Z_\mu\Delta. \tag{4.1}$$

[As a consequence this scalar product will be invariant under the conformal Lie algebra so$(n,2)$ defined above.]

Proposition 4.1: The SO(n) and dilatation invariant scalar product for which the relation (4.1) is valid and the function 1 has norm 1, is given by

$$(f,g)=\int\overline{f(Z)}g(Z)dm_n(Z), \tag{4.2}$$

$$dm_n(Z)=\rho(Z\bar Z)\delta(Z^2)d^{2n}Z, \tag{4.3a}$$

$$\rho(t)=(2/\pi)^{2h-1}[(16t)^{1-h/2}/\Gamma(h-1)]K_{h-2}(2\sqrt{t}),\quad h=\tfrac{1}{2}n, \tag{4.3b}$$

where K_ν is the MacDonald Bessel function. The completion $K\ (=K_n)$ of P with respect to the scalar product topology is a Hilbert space with a reproducing kernel

$$\delta_n(Z,\bar Z')=\Gamma(h-1)(Z\bar Z')^{1-h/2}I_{h-2}(2\sqrt{Z\bar Z'})$$
$$=\sum_{l=0}^{\infty}\delta_n^{(l)}(Z,\bar Z'), \tag{4.4}$$

where $\delta_n^{(l)}$ are the reproducing kernels of the SO(n) invariant subspaces K^l of homogeneous polynomials of degree l:

$$\delta_n^{(l)}(Z,\bar Z')=(Z\bar Z')^l/(h-1)_l l!, \tag{4.5}$$

I_ν is the modified Bessel function of the first kind [and $(\nu)_l$ is given in (2.14)].

Proof: We look for an invariant scalar product of the type (4.2), (4.3a) [with unknown $\rho(t)$, $t=Z\bar Z$] such that

$$(Z_\mu f,g)=(f,D_\mu g). \tag{4.6}$$

Integrating the right-hand side by parts, we obtain

$$\bar Z_\mu\rho(t)\delta(Z^2)=[\partial_\mu(h+1+Z\partial)-\tfrac{1}{2}\Delta Z_\mu]\rho(t)\delta(Z^2)$$
$$=[(\partial_\mu X-\tfrac{1}{2}\Delta Z_\mu)\rho(t)]\delta(Z^2)$$
$$=\bar Z_\mu[t\rho''(t)+(h-1)\rho'(t)]\delta(Z^2). \tag{4.7}$$

Here we have used the complex variable formalism for the δ function (see Gel'fand and Shilov[10]). Thus ρ satisfies the differential equation

$$t\rho''(t)+(h-1)\rho'(t)-\rho(t)=0\quad(t=Z\bar Z). \tag{4.8}$$

According to Ref. 11 every integrable solution is proportional to the expression (4.3b). This expression is positive and so normalized that $(1,1)=1$. In fact

$$\frac{8}{\pi^{n-1}\Gamma(h-1)}\int d^nx\int d^ny\,(x^2+y^2)^{1-h/2}$$
$$\times K_{h-2}(2(x^2+y^2)^{1/2})\delta(x^2-y^2)\delta(2xy)$$
$$=\frac{8}{\pi^{n-1}\Gamma(h-1)}\int d^nx\int d^{n-1}y\,\frac{(x^2+y^2)^{1-h/2}}{2(x^2)^{1/2}}$$
$$\times K_{h-2}(2(x^2+y^2)^{1/2})\delta(x^2-y^2)$$
$$=\frac{8s_{n-1}}{\pi^{n-1}\Gamma(h-1)}\int d^nx\,\frac{(x^2)^{h/2-1}}{2^{1+h/2}}K_{h-2}(2^{3/2}(x^2)^{1/2})$$
$$=\frac{2^{2-h/2}s_{n-1}s_n}{\pi^{n-1}\Gamma(h-1)}\int_0^\infty r^{3h-3}K_{h-2}(2^{3/2}r)dr=1. \tag{4.9}$$

[Here $s_\nu=2\pi^{\nu/2}/\Gamma(\nu/2)$ is the surface area of the unit sphere in $\mathbb{R}^\nu$, and the last integral equals

$$2^{-[1+(3/2)h]}\Gamma(h)\Gamma(2h-2)=\pi^{-1/2}2^{h/2-4}\Gamma(h)\Gamma(h-\tfrac{1}{2})\Gamma(h-1);$$

see, e.g., Eq. 6.561.16 of Ref. 12.]

To complete the proof of the proposition, it suffices to show that the function $\delta_n^{(l)}(Z,Z')$ [(4.5)] is the reproducing kernel in K^l. That will appear as a simple consequence of the following statement which has an interest of its own.

Proposition 4.2: In the subspace $\mathcal{H}^l$ $(\approx K^l)\subset B$ of harmonic homogeneous polynomials the scalar products (1.4) and (4.2) are related by

$$(f,g)=b_l\langle f,g\rangle, \tag{4.10}$$

where

$$b_l=(h-1)_l. \tag{4.11}$$

We shall first complete the proof of Proposition 4.1, assuming that Proposition 4.2 is true.

Let a be a complex lightlike vector (i.e., $a^2=0$) and let

$$s_a^l(\xi)=(\bar a\xi)^l/l!. \tag{4.12}$$

Then it follows from (4.10) and (1.7) that for $f(\xi)\in\mathcal{H}_l$

$$(s_a^l,f)=b_l\langle s_a^l,f\rangle=b_l f(a). \tag{4.13}$$

Thus

$$b_l^{-1}\,\overline{s_a^l}(Z)=[1/l!\,(h-1)_l](a\bar Z)^l=\delta_n^{(l)}(a,\bar Z) \tag{4.5'}$$

is indeed the reproducing kernel in K^l. It is, of course, also the reproducing kernel in $\mathcal{H}_l$ since the function

$$f(\xi)=\int\delta_n^{(l)}(\xi,\bar Z)f(\bar Z)dm_n(Z) \tag{4.14}$$

$[f(Z)\in K^l]$ is harmonic. [Equation (4.14) provides still another form of the harmonic extension formula for functions in K^l.] The corresponding property of δ_n(4.4) now follows from the power series expansion of the Bessel function and the orthogonality of homogeneous

polynomials of different degrees with respect to the scalar product (4.2). That completes the proof of Proposition 4.1.

Note 4.3: Equation (4.13) and Proposition 2.5 imply that the functions s_a^l with $a^2=0$ form a complete set in $\mathcal{H}_l$.

Proof of Proposition 4.2: Equation (4.10) is a consequence of Schur's lemma since the representation (1.11) of SO(n) is irreducible in K^l (and in $\mathcal{H}^l$) and both scalar products are rotation invariant. Indeed, let $f_1,\dots,f_d$ be an orthonormal basis in $\mathcal{H}^l\subset B^l$ with respect to the scalar product (1.4) so that

$$\langle f_j,f_k\rangle=\delta_{jk}. \tag{4.15}$$

Assume that

$$(f_j,f_k)=A_{jk}. \tag{4.16}$$

Let Λ be an arbitrary rotation and $D^l=D^l(\Lambda)$ its symmetric tensor representation, such that

$$f_j(\Lambda^{-1}\mathbb{Z})=\sum_{s=1}^{d} f_s(\mathbb{Z})D^l_{sj}. \tag{4.17}$$

Both scalar products are invariant under the transformation (4.17). The invariance of $\langle\ ,\ \rangle$ implies that the matrix D^l is unitary, and the invariance of $(\ ,\ )$ implies that

$$A=(A_{jk})=(\textstyle\sum_{st}\overline{D^l_{tj}}A_{ts}D^l_{sk})=D^{l*}AD^l$$

and hence that

$$D^lA=AD^l. \tag{4.18}$$

Since the representation $D^l(\Lambda)$ is irreducible (in K^l), it follows from Schur's lemma that A is a multiple of the unit matrix and that Eq. (4.10) holds.

To evaluate the constants b_l, we use again the harmonic functions $s_a^l(\xi)$ [(4.12)]. Since $(\bar a\,\xi)s_a^l(\xi)=(l+1)s_a^{l+1}(\xi)\in\mathcal{H}_{l+1}$, we obtain from Eqs. (4.1) and (1.6) (with z replaced by ξ) for any $f\in\mathcal{H}_{l+1}$

$$b_{l+1}\langle(\bar a\,\xi)s_a^l(\xi),f(\xi)\rangle$$

$$=((\bar a\,\mathbb{Z})s_a^l(\mathbb{Z}),f(\mathbb{Z}))=(s_a^l,(aD)f)=b_l\langle s_a^l(\xi),(aD)f(\xi)\rangle$$

$$=(h-1+l)b_l\langle s_a^l,(a\partial)f\rangle=(h-1+l)b_l\langle(\bar a\,\xi)s_a^l,f\rangle.$$

Hence, $b_{l+1}=(h-1+l)b_l$, and Eq. (4.11) follows because $b_0=1$ by (4.9).

Corollary 4.4: The scalar product (1.4) is majorized by the scalar product (4.2) and therefore K ($=K_n$) can be regarded as a proper subspace of B.

Indeed, it follows from (4.10), (4.11) that

$$\langle f,f\rangle\le\sup_l[1/(h-1)_l](f,f)$$

$$=2(f,f)\quad\text{for } n=3$$

$$=(f,f)\quad\text{for } n=4,5,\cdots. \tag{4.19}$$

Remark: The above formulas are obviously not applicable for $n=2$. This is not surprising, since the group SO(2) is Abelian and all its irreducible representations are one-dimensional.

5. ISOMETRIC HOMOMORPHISM OF K_3 INTO B_2 AND OF K_4 INTO B_4

It is well known that the universal covering groups of SO(3) and SO(4) are SU(2) and SU(2)×SU(2), respectively. This allows us to define a covariant mapping of K_3 into B_2 and of K_4 into B_4. We shall also show that this mapping can be normalized in such a way that the scalar product be preserved.

Consider the mapping from $\mathbb{C}^2$ onto $\mathbb{K}_3$ defined by

$$\mathbb{Z}_j=(1/2\sqrt{2})z\epsilon\sigma_j z,\quad j=1,2,3, \tag{5.1}$$

where σ_j are the Pauli matrices and ϵ is the 2×2 antisymmetric unit tensor:

$$\sigma_1=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \sigma_2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma_3=\begin{pmatrix}1&0\\0&-1\end{pmatrix},\quad \epsilon=\begin{pmatrix}0&1\\-1&0\end{pmatrix}. \tag{5.2}$$

[The condition $\mathbb{Z}^2=0$ for $\mathbb{Z}$ given by Eq. (5.1) is easily verified, —cf. Eq. (1.10) of Ref. 13.] This mapping allows us to regard the representation U [(1.11)] of SO(3) as a subrepresentation of the representation $S=\oplus S^r$ [see (1.2)] of SU(2); to this end one defines $\Lambda(u)\mathbb{Z}$ as the transformation induced by uz according to (5.1).

Proposition 5.1: The mapping $A: K_3\to B_2$ defined by

$$\phi(z)=f\left(\frac{z\epsilon\sigma_j z}{2\sqrt{2}}\right)\equiv(Af)(z) \tag{5.3}$$

preserves the scalar products (1.4), (4.2):

$$\langle\phi_1,\phi_2\rangle\equiv\int\bar\phi_1(z)\phi_2(z)d\mu_2(z)=\int f_1(\mathbb{Z})f_2(\mathbb{Z})dm_3(\mathbb{Z})=(f_1,f_2). \tag{5.4}$$

It is also an intertwining map for the representations U and S of SU(2):

$$AU(\Lambda)=S(u)A,\quad \Lambda=\Lambda(u), \tag{5.5}$$

and this exhibits the unitary equivalence of U with a subrepresentation of S. The (left) inverse mapping to (5.3) is the average of $\phi(uz)$ over the center of SU(2):

$$A^{-1}:\phi(z)\to f(\mathbb{Z})=\tfrac{1}{2}[\phi(z)+\phi(-z)]. \tag{5.6}$$

Proof: The only nontrivial point is to verify that the map A [(5.3)] is an isometry. It is clear that if f_1 and f_2 are polynomials of different degrees, they are orthogonal in K_3 [$(f_1,f_2)=0$] and so are their images ϕ_1 and ϕ_2 in B_2 [$\langle\phi_1,\phi_2\rangle=0$]. It follows (using in addition Note 4.3) that it suffices to verify (5.4) for

$$f_j(\mathbb{Z})=(\bar a_j\mathbb{Z})^l,\quad j=1,2, \tag{5.7}$$

where a_1 and a_2 are two arbitrary isotropic vectors: $a_1^2=0=a_2^2$. To do that, we note that each isotropic a_j can be written in the form

$$a_j=(1/2\sqrt{2})\,b_j\epsilon\sigma b_j,\quad j=1,2 \tag{5.8}$$

where each of the two-component spinors b_1 and b_2 is determined from (5.8) up to an over-all sign. Taking

into account the relation

$$\bar{a}_1 a_2 \ (=\sum_{\mu=1}^{3} \bar{a}_{1\mu} a_{2\mu}) = \tfrac{1}{4}(\bar{b}_1 b_2)^2 \quad (\bar{b}_1 b_2 = \sum_{\alpha=1}^{2} b_{1\alpha}\bar{b}_{2\alpha}) \tag{5.9}$$

and applying (4.13), (4.11), (4.12), we obtain

$$(f_1, f_2) = l!\,(1/2)_l (a_1\bar{a}_2)^l = (1/4^l) l!\,(1/2)_l (b_1\bar{b}_2)^{2l},$$

$$\langle \phi_1, \phi_2 \rangle = (1/4^{2l})(2l)!\,(b_1\bar{b}_2)^{2l} = (f_1, f_2).$$

This completes the proof of the isometry of A.

We notice that the reproducing kernel δ_3 in K_3 is related to the reproducing kernel $e^{z\bar{z}'}$ in B_2 by

$$\delta_3(\mathcal{Z}, \mathcal{Z}') = \tfrac{1}{2}(e^{z\bar{z}'} + e^{-z\bar{z}'})$$

$$= \cosh(2\sqrt{\mathcal{Z}\bar{\mathcal{Z}}'})$$

$$= \sqrt{\pi}\,(\mathcal{Z}\bar{\mathcal{Z}}')^{1/4} I_{-1/2}(2\sqrt{\mathcal{Z}\bar{\mathcal{Z}}'})$$

[in accord with (4.4) and (5.6)].

In order to display the homorphism of SU(2)×SU(2) onto SO(4), we introduce the quaternion units $q = (\mathbf{q}, q_4)$, satisfying the following (defining) relations:

$$q_4 = 1 \quad (\text{so that } q_4^2 = q_4, q_4\mathbf{q} = \mathbf{q}q_4 = \mathbf{q}) \tag{5.10a}$$

$$(\mathbf{aq})(\mathbf{bq}) = -(\mathbf{ab})q_4 + (\mathbf{a} \wedge \mathbf{b})\mathbf{q}, \tag{5.10b}$$

for any choice of the 3-vectors $\mathbf{a}$ and $\mathbf{b}$. We shall use the 2×2 matrix realization

$$\mathbf{q} = -i\boldsymbol{\sigma}, \quad q_4 = 1_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{5.11}$$

Let $w = (w_1, w_2)$ and $z = (z_1, z_2)$ be two pairs of complex numbers. We define a mapping from $\mathbb{C}^4[\ni(\bar{w}, z)]$ onto the complex light cone $\mathbb{K}_4$ [(1.10)] by

$$(\bar{w}, z) \to \mathcal{Z}_\mu = (1/\sqrt{2})\,\bar{w} q_\mu z. \tag{5.12a}$$

Any SU(2)×SU(2) transformation (u, v) of $(\bar{w}, z)$ generates an SO(4) transformation Λ of $\mathcal{Z}$ by

$$\Lambda\mathcal{Z} = (1/\sqrt{2})\,\bar{u}\,\bar{w} q v z = (1/\sqrt{2})\,\bar{w}\,u^{-1} q v z. \tag{5.12b}$$

The scalar product of two vectors $\mathcal{Z}[\leftarrow(\bar{w}, z)]$ and $\mathcal{Z}'$ $[\leftarrow(\bar{w}', z')]$ is given by

$$\mathcal{Z}\mathcal{Z}' = (z \epsilon z')(\bar{w} \epsilon\, \bar{w}') \tag{5.13}$$

[see again Eq. (1.10) of Ref. 13]. Equation (5.13) implies in particular that $\mathcal{Z}^2 = 0$ so that Eq. (5.12) indeed defines a mapping onto the "light cone" $\mathbb{K}_4$. Furthermore, one readily verifies the relation

$$\mathcal{Z}\bar{\mathcal{Z}}' = (w'\bar{w})(z\bar{z}'). \tag{5.14}$$

Consider the space $B_4 = B_2 \times B_2$ of entire analytic functions of four complex variables $(\bar{w}, z)$. According to (1.5) the scalar product in B_4 is given by

$$\langle \phi, \psi \rangle = \pi^{-4} \int \cdots \int \overline{\phi(\bar{w}, z)}\, \psi(\bar{w}, z) e^{-w\bar{w} - z\bar{z}} d^4 z\, d^4 w. \tag{5.15}$$

Let $T = T(u, v)$ be the (fully reducible) unitary representation of SU(2)×SU(2) defined by

$$[T(u, v)\psi](\bar{w}, z) = \psi(\bar{u}^{-1}\bar{w}, v^{-1} z). \tag{5.16}$$

Remark: The choice of the notation $(\bar{w}, z)$ for the coordinates in B_4 is, of course, a matter of convention; we could have used the pair (w, z) as well. With the present choice the SO(3) transformation law for the quaternion units q_j under the diagonal SU(2) subgroup [which consists of the pairs (u, u)] is the same as that for the matrices σ_j under the mapping $z \to uz$ [see Eq. (5.12b)].

Proposition 5.2: The mapping $A : K_4 \to B_4$ given by

$$\phi(\bar{w}, z) = (Af)(\bar{w}, z) \equiv f\left(\frac{1}{\sqrt{2}}\,\bar{w} q z\right) \tag{5.17}$$

is an isometry of K_4 into B_4 which establishes the equivalence of the representation U [(1.11)] of SU(2)×SU(2) [≈ (SO(4)] with a subrepresentation of T. The image of K_4 under the map A is the set

$$AK_4 = \{\phi(\bar{w}, z) \in B_4;\ \phi(\bar{w}, z) = M\phi(\bar{w}, z)\}, \tag{5.18}$$

where $M\phi$ is the mean value

$$M\phi(\bar{w}, z) \equiv \int_0^{2\pi} \phi(\bar{w} e^{i\alpha}, z e^{-i\alpha})\, \frac{d\alpha}{2\pi}. \tag{5.19}$$

In particular, the reproducing kernel $\delta_4(\mathcal{Z}, \bar{\mathcal{Z}}') = I_0(2(\mathcal{Z}\bar{\mathcal{Z}}')^{1/2})$ in K_4 is equal to the mean value of the reproducing kernel in B_4:

$$Me^{w\bar{w}' + z\bar{z}'} = \int_0^{2\pi} \exp(\bar{w}w'\, e^{-i\alpha} + z\bar{z}' e^{i\alpha})\, \frac{d\alpha}{2\pi}$$

$$= I_0(2[(\bar{w}w')(z\bar{z}')]^{1/2}) = A\delta_4(\mathcal{Z}, \mathcal{Z}'). \tag{5.20}$$

Proof: The argument follows the same pattern as the proof of Proposition 5.1. Let again a_1 and a_2 be two zero-length vectors represented by the pairs of spinors (u_1, v_1) and (u_2, v_2):

$$a_j = (1/\sqrt{2})(\bar{u}_j q v_j), \quad j = 1, 2, \tag{5.21}$$

and let f_j be given by (5.7). Using (4.10), (4.11), and (5.14), we obtain

$$(f_1, f_2) = (l!)^2 (a_1\bar{a}_2)^l = (l!)^2 (\bar{u}_1 u_2)^l (v_1\bar{v}_2)^l = \langle \phi_1, \phi_2 \rangle.$$

The remainder of Proposition 5.2 follows easily.

ACKNOWLEDGMENTS

It is a pleasure to thank G. Mack and T. Regge for helpful discussions. The hospitality extended to one of us (I.T.T.) at the Institute for Avanced Study is gratefully acknowledged.

*Supported in part by NSF Grant GP 39048.

†On leave from the Institute of Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia 1113, Bulgaria.

[1](a) For the finite-dimensional case see V. Bargmann, Comm. Pure Appl. Math. **14**, 187 (1961); 20, 1 (1967). In these papers the spaces B_n are denoted by F_n, and the spaces of homogeneous polynomials B^r (= B_n^r), introduced in Sec. 1, are denoted by $\mathfrak{P}_r$. (b) For the infinite-dimensional case see I. E. Segal, Illinois J. Math. 6, 500 (1962), and *Mathematical Problems of Relativistic Physics* (Amer. Math. Soc., Providence, R.I., 1963).

[2]V. Bargmann, Rev. Mod. Phys. 34, 829 (1962). This article is reprinted in the collection of papers, *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. Van Dam (Academic, New York, 1965), which also contains a related earlier paper by J. Schwinger.

[3]S. Bergman, *The Kernel Function and Conformal Mapping*, Mathematical Surveys No. V (American Mathematical Society, New York, 1950). N. Aronszajn, Trans. Amer. Math.

Soc. 68, 337 (1950). See also Refs. 1, 2, where the functions $e^{\overline{m}z}$ are called *principal vectors*.
[4]V. Dobrev, G. Mack, V. Petkova, S. Petrova, and I. Todorov, "Elementary representations and intertwining operators for the generalized Lorentz group," IAS preprint, Princeton (May, 1975); "On the Clebsch-Gordan expansion for the Lorentz group in n dimensions," Rep. Math. Phys. **9**, 219 (1976); V. Dobrev, V. Petkova, S. Petrova, and I. Iodorov, Phys. Rev. D **13**, 887 (1976). For other applications of the homogeneous polynomial techniques, see A.I. Oksak and I.T. Todorov, Rep. Math. Phys. **7**, 417 (1975).
[5]P. Appell and J. Kampé de Fériet, *Fonctions hypergéométriques et hypersphériques, Polynomes d'Hermite* (Gauthier-Villars, Paris, 1926); D.L. Van der Waerden, *Die Gruppentheoretische Methode in der Quantenmechanik* (Springer-Verlag, Berlin, 1932), p. 13; N.Ya. Vilenkin, *Special Functions and the Theory of Group Representations* (Nauka, Moscow, 1965, in Russian), Chap. IX, Secs. 2.5 and 2.6.
[6]I.M. Gel'fand, M.I. Graev, and N.Ya. Vilenkin, *Generalized Functions, Vol.* 5 (Academic, New York, 1966), Chap. II, Sec. 2.7.
[7]P.A.M. Dirac, Proc. Roy. Soc. Lond. A **246**, 326 (1958). See also *Lectures on Quantum Mechanics* (Belfer Grad. School of Sci., Yeshiva Univ., New York, 1964).
[8]A.J. Hanson, T. Regge, and C. Teitelboim, "Constraint Hamiltonian systems," Academia Nazionale dei Lincei, Rom., 1976.
[9]G. Mack and I. Todorov, J. Math. Phys. **10**, 2078 (1969). The Hermitian generators of the algebra are $L_{\mu\nu} = iX_{\mu\nu}$, $L_{\mu,n+2} = (1/\sqrt{2})(Z_\mu + D_\mu)$, $L_{n+1,\mu} = (i/\sqrt{2})(Z_\mu - D_\mu)$, $L_{n+1,n+2} = X$.
[10]I.M. Gel'fand and G.E. Shilov, *Generalized Functions, Vol. I* (Academic, New York, 1964), Sec. B2.1, in particular Eqs. (6)—(8) on p. 389.
[11]E. Kamke, *Differentialgleichungen Lösungsmethoden und Lösungen* (Akademische, Leipzig, 1943), see 2.162 (1a) (p. 440).
[12]I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965).
[13]I.T. Todorov and R.P. Zaikov, J. Math. Phys. **10**, 2014 (1969).

Conformal properties of a class of exactly solvable *N*-body problems in space dimension one

G. Barucchi

Istituto di Fisica dell' Università, Torino, Italy,
Istituto di Fisica Matematica dell'Università, Torino, Italy,
and Istituto Nazionale di Fisica Nucleare, Torino, Italy

T. Regge*

Istituto de Fisica dell' Università, Torino, Italy
and Institute for Advanced Study, Princeton, New Jersey
(Received 20 September 1976)

An algebra of collective variables in the generalized (Calogero–Sutherland) N-body classical one-dimensional model is introduced, and their transformation properties under the conformal group are discussed in detail.

1. INTRODUCTION

In recent years[1] a class of exactly solvable N-body problems in space dimension one has been investigated. The idea of the model is extremely simple. The $N\times N$ matrix $L=P+D$ is introduced

$$P_{ij}=\delta_{ij}p_i,\quad D_{ij}=\begin{cases}0, & i=j,\\ iA/(q_i-q_j), & i\neq j,\end{cases}$$

where p_i and q_i are the momentum and position of the ith particle. If one considers the Hamiltonian

$$H=\tfrac{1}{2}\mathrm{Tr}L^2=\tfrac{1}{2}\sum_i p_i^2+\tfrac{1}{2}\sum_{i\neq j}\frac{A^2}{(q_i-q_j)^2} \tag{1.1}$$

one finds the equations of motion

$$\begin{cases}\dfrac{dp_i}{dt}=\{p_i,H\}=2\sum\limits_{j\neq i}\dfrac{A^2}{(q_i-q_j)^3},\\[2ex] \dfrac{dq_i}{dt}=\{q_i,H\}=p_i.\end{cases} \tag{1.2}$$

One finds also as a direct consequence of (1.2)

$$\frac{dL_{ij}}{dt}=\sum_k(L_{ik}M_{kj}-M_{ik}L_{kj})=[L,M]_{ij},$$

or briefly

$$\frac{dL}{dt}=[L,M],$$

where

$$M_{ij}=\begin{cases}\sum\limits_{k\neq i}\dfrac{iA}{(q_i-q_k)^2}, & i=j\\[2ex] -\dfrac{iA}{(q_i-q_j)^2}, & i\neq j.\end{cases}$$

It follows that

$$\frac{dL^n}{dt}=[L^n,M]$$

and therefore

$$\frac{d}{dt}\mathrm{Tr}L^n=0.$$

The quantities

$$H_n^0=(1/n!)\,\mathrm{Tr}L^n$$

are constants of motion. In particular we have $H_2^0=H$. Obviously also the P.b. (Poisson brackets) $\{H_n^0,H_m^0\}$ are constants of motion, but since at $t\to\infty$

$$H_n^0\to\frac{1}{n!}\sum_i p_i^n,$$

we have $\{H_n^0,H_m^0\}\to 0$ and

$$\{H_n^0,H_m^0\}\equiv 0$$

for any t. The system has N independent constants of motion since all H_n^0 with $n>N$ can be expressed as polynomials in the H_n^0 for $n\leq N$.

In order to solve completely the equations of motion, one can introduce the auxiliary matrix Q:

$$Q_{ij}=\delta_{ij}q_i.$$

It can be easily checked that

$$\frac{dQ}{dt}=[Q,M]+L.$$

Therefore, if we consider the quantities

$$H_n^1=\frac{1}{n!}\mathrm{Tr}(QL^{n-1}),$$

we find

$$\frac{dH_n^1}{dt}=\frac{1}{n!}\mathrm{Tr}L^n=H_n^0=\text{const.}$$

It follows that H_n^1 are linear functions of the time. Given the initial data $q_i(0)$, $p_i(0)$, one can calculate $H_n^1(0)$, $H_n^0(0)$. At the time t one has

$$H_n^0(t)=H_n^0(0),$$
$$H_n^1(t)=H_n^1(0)+tH_n^0(0).$$

From the knowledge of H_n^0, H_n^1 we can retrieve the $q_i(t)$, $p_i(t)$, by solving a set of algebraic equations.

These results are essentially already contained in Calogero *et al.*,[4] who also consider very interesting generalizations to periodic geometries where the inverse square potential in (1.1) is replaced by Weierstrass $\wp$ function. We plan another generalization of these results.

2. GENERALIZED HAMILTONIANS

We consider the generalized Hamiltonian

$$H_f = \sum_{n=0}^{\infty} f_n H_n^0$$

together with the matrix $M^{(f)}$:

$$M_{ij}^{(f)} = \begin{cases} \sum_{k\neq i} \frac{1}{q_i - q_k} \left(\frac{df(L)}{dL}\right)_{ik}, & i=j, \\ -\frac{1}{q_i - q_j} \left(\frac{df(L)}{dL}\right)_{ij}, & i\neq j, \end{cases}$$

where

$$f(L) = \sum_{n=0}^{\infty} \frac{f_n}{n!} L^n.$$

By direct and easy computation we can check that

$$\frac{dQ}{dt} = [Q, M^{(f)}] + \frac{df(L)}{dL}. \tag{2.1}$$

A more elaborate procedure (see Appendix A), however, also shows that

$$\frac{dL}{dt} = [L,\ M^{(f)}]. \tag{2.2}$$

From (2.1) and (2.2) in analogy to what we did already in Sec. 1, we find that all H_n^0 and *a fortiori* all H_g (where $H_g = \sum_{n=0}^{\infty} g_n H_n^0$) are constants of motion, that is

$$\{H_g, H_f\} = 0.$$

Similary we have

$$\{\mathrm{Tr}[Qg(L)], H_f\} = \mathrm{Tr}\left(\frac{df(L)}{dL} g(L)\right) = \text{constant of motion.}$$

The solution of the equations of motion proceeds as before since all quantities $\mathrm{Tr}[Qg(L)]$ are now linear in time. (An alternate solution is derived in Appendix B).

The result can be further generalized by considering traces of the kind

$$\mathrm{Tr}[Qg_1(L)Qg_2(L)\cdots Qg_p(L)]$$

A trace of product of degreee p in Q is a polynomial of degree p in the time.

Of particular interest is the case

$$H = \mathrm{Tr} f(L), \quad f(L) = (L^2 + m^2)^{1/2},$$

which yields a relativistic theory. Indeed one can introduce the total momentum $\sum_i p_i = H_1^0$ and the boost B,

$$B = \mathrm{Tr}[Q(L^2 + m^2)^{1/2}],$$

and check the P.b.,

$$\{H_1^0, B\} = -H, \quad \{H, B\} = -H_1^0.$$

Cleary H, H_1^0, B generate the Poincaré group in space–time dimension $1+1$.

3. THE DISTINGUISHED VECTOR v

Consider the matrix commutator

$$[L, Q]_{ij} = iA(\delta_{ij} - v_i v_j) = iA(1 - T)_{ij},$$

where $v_i = 1$ $(i = 1, 2, \ldots, N)$. From the explicit formula for $M^{(f)}$ we find

$$\sum_j M_{ij}^{(f)} v_j = \sum_i v_i M_{ij}^{(f)} = \sum_i M_{ij}^{(f)} = \sum_j M_{ij}^{(f)} = 0. \tag{3.1}$$

(3.1) implies the equation

$$[M^{(f)}, [L, Q]] = 0,$$

which also follows directly from $(d/dt)\,[L, Q] = 0$ and (2.1), (2.2). The vector v can be used to give an alternate definition of H_f:

$$H_f = (v,\ f(L)v) = \sum_{ij} [f(L)]_{ij} = \mathrm{Tr}[T \cdot f(L)]. \tag{3.2}$$

The equivalence of the definitions in (3.2) can be easily proved by noticing that all lead to constants of the motion which become identical in the limit $t \to \infty$, $|q_i - q_j| \to \infty$. The vector v and the associated matrix T can be used in order to classify all collective variables of the kind:

$$\mathrm{Tr}(Q^{a_1} L^{b_1} Q^{a_2} L^{b_2} \cdots Q^{a_p} L^{b_p}). \tag{3.3}$$

Of particular interest are

$$H_n^p = \frac{1}{n!} \frac{\partial^p}{\partial \lambda^p} \mathrm{Tr}(L + \lambda Q)^n \Big|_{\lambda=0}.$$

All expressions defined by (3.3) can be reduced to polynomials in the H_n^p through repeated reordering of the factors in (3.3) and use of the formula

$$LQ - QL = iA(1 - T) \tag{3.4}$$

and of the equivalent definitions:

$$H_n^p = \frac{1}{n!} \frac{\partial^p}{\partial \lambda^p} \mathrm{Tr}[T(L + \lambda Q)^n] \Big|_{\lambda=0}$$

$$= \frac{1}{n!} \frac{\partial^p}{\partial \lambda^p} (v, (L + \lambda Q)^n v) \Big|_{\lambda=0}.$$

In deriving these identities we may use the equivalent definition:

$$H_n^p = \frac{1}{n!} \frac{\partial^p}{\partial \lambda^p} h_n \Big|_{\lambda=0}$$

where

$$h_n = \mathrm{Tr}(L + \lambda Q)^n = \mathrm{Tr}[T(L + \lambda Q)^n]. \tag{3.5}$$

(3.5) follows from

$$\mathrm{Tr}[(L + \lambda Q)^n (LQ - QL)] = \mathrm{Tr}\{(L + \lambda Q)^n [(L + \lambda Q)Q - Q(L + \lambda Q)]\} = 0,$$

using (3.4).

As an example we have the identity:

$$\mathrm{Tr}(L^2 Q^2) = 2H_4^2 - \tfrac{1}{3} A^2 (1 - N) N.$$

4. THE CONFORMAL GROUP

Of specific interest are the three variables H_2^p, $p = 0$, 1, 2. They have the explicit expressions

$$H_2^0 = \tfrac{1}{2} \sum_i p_i^2 + \tfrac{1}{2} \sum_{i\neq j} \frac{A^2}{(q_i - q_j)^2},$$

$$H_2^1 = \sum_i p_i q_i,$$

$$H_2^2 = \sum_i q_i^2$$

and the P.b.

$$\{H_2^0, H_2^1\} = -2H_2^0, \quad \{H_2^1, H_2^2\} = -2H_2^2,$$

$$\{H_2^2, H_2^0\} = 2H_2^1, \tag{4.1}$$

which show that the H_2^p ($p=0, 1, 2$) obey the same P.b. as the Lie algebra of the conformal group O(2,1) considered in Ref. 2 and of which they can be considered as an extension to many-body systems.

In addition to (4.1) we have also

$$\begin{aligned} \{H_2^0, H_n^p\} &= -p(n-p+1)H_n^{p-1}, \\ \{H_2^1, H_n^p\} &= (n-2p)H_n^p, \\ \{H_2^2, H_n^p\} &= 2H_n^{p+1}, \end{aligned} \tag{4.2}$$

which show that all H_n^p belonging to a given n form an $(n+1)$-multiplet under the action of the group. Equivalently we may observe that we can generate all H_n^p from H_n^0 by repeated P.b. with H_2^2. Indeed many of the properties of H_n^{p-1} can be inferred from (4.2) as definition.

We have also the doublet

$$H_1^0 = \sum_i p_i, \quad H_1^1 = \sum_i q_i$$

and the P.b.

$$\{H_1^0, H_n^p\} = -pH_{n-1}^{p-1}, \quad \{H_1^1, H_n^p\} = H_{n-1}^p.$$

However, the computation of the generic P.b.

$$\{H_n^p, H_m^q\} \tag{4.3}$$

can be reduced to the computation of expressions of the kind (3.3), which needs extensive reordering before it can be expressed as a polynomial in the H_k^h. In Appendix C it is shown that if we normalize the H_n^p by introducing

$$\begin{aligned} &K_a^\alpha = [(a+\alpha)!/(a-\alpha)!]^{1/2} H_{2a}^{a-\alpha}, \\ &a = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots, \quad \alpha = -a, -a+1, \ldots, a-1, a, \end{aligned}$$

we find that the contribution of linear terms in K_a^α to the P.b. (4.3) is given by

$$\begin{aligned} \{K_s^m, K_t^n\} = \sum_{p \text{ integer} \geq 0} & \omega(s+t-2p-1, s, t) \\ & \times (s, m, t, n | s, t, s+t-2p-1, m+n) \\ & \times K_{s+t-2p-1}^{m+n} + \cdots, \end{aligned} \tag{4.4}$$

where

$$\omega(s+t-1, s, t) = \left[(2s+2t)\binom{2s+2t-2}{2t-1}\right]^{1/2}.$$

Formula (4.4) takes into full account the conformal invariance of the theory and the Jacobi's identities involving two generic K_a^α, K_b^β with $K_{1/2}^{\pm 1/2}$ (that is, H_1^0, H_1^1). In general in the rhs other terms (multilinear in K_d^δ) involving higher powers in A will appear destroying the simplicity of the theory.

5. COMMENTS

We have shown that the collective variables in the Calogero *et al.*[1] model form an algebra which extends the set of commuting variables H_n^0, already known to Calogero, to a conformally invariant set H_n^p. This set appears as the natural generalization of the conformal structure investigated in Ref. 2. We plan to investigate the formal properties of this set and their relevance to the physical behavior of the system in a forth-coming paper.

APPENDIX A: PROOF OF (2, 2)

Clearly by linearity the proof for a generic $f(L)$ can be carried out separately on each $f=(1/n!)L^n$. In this case

$$\begin{aligned} M_{ii}^{(n)} &= \frac{1}{(n-1)!} \sum_{k \neq i} D_{ik} (L^{n-1})_{ik}, \\ M_{ij}^{(n)} &= -\frac{1}{(n-1)!} D_{ij} (L^{n-1})_{ij}, \quad i \neq j, \end{aligned} \tag{A1}$$

where

$$D_{ij} = \begin{cases} 0 & i=j, \\ 1/(q_i - q_j), & i \neq j. \end{cases}$$

The proof of (2.2) runs separately for the diagonal and the off-diagonal elements:

Case 1, $i=j$

From (A1) we derive

$$\begin{aligned} [L, M^{(n)}]_{ii} &= \sum_k (L_{ik} M_{ki}^{(n)} - M_{ik}^{(n)} L_{ki}) \\ &= \sum_{k \neq i} (L_{ik} M_{ki}^{(n)} - M_{ik}^{(n)} L_{ki}) \\ &= \frac{iA}{(n-1)!} \sum_k D_{ik}^2 [(L^{n-1})_{ki} - (L^{n-1})_{ik}]. \end{aligned}$$

Therefore, using

$$\{p_i, L_{hk}\} = iA\{p_i, D_{hk}\} = iA(\delta_{ih} - \delta_{ik}) D_{hk}^2,$$

we have

$$\begin{aligned} [L, M^{(n)}]_{ii} &= \frac{1}{(n-1)!} \sum_{h,k} \{p_i, L_{hk}\} (L^{n-1})_{kh} \\ &= \frac{1}{n!} \{p_i, \mathrm{Tr} L^n\}, \end{aligned}$$

where a factor $1/n$ appears on account of the cyclical properties of the trace. This ends Case 1.

Case 2, $i \neq j$

Using

$$\begin{aligned} \{q_i, \mathrm{Tr} L^n\} &= n \sum_{h,k} \{q_i, L_{hk}\} (L^{n-1})_{kh} \\ &= n \sum_{h,k} \delta_{hk} \delta_{ki} (L^{n-1})_{kh} = n (L^{n-1})_{ii}, \end{aligned}$$

we have

$$\begin{aligned} (n-1)! \frac{dL_{ij}}{dt} &= -\frac{iA}{(q_i - q_j)^2} \left(\frac{dq_i}{dt} - \frac{dq_j}{dt} \right) \\ &= -iA D_{ij}^2 [(L^{n-1})_{ii} - (L^{n-1})_{jj}]. \end{aligned}$$

As for $[L, M^{(n)}]_{ij}$ we derive from (A1)

$$\begin{aligned} (n-1)! [L, M^{(n)}]_{ij} &= (n-1)! \sum_k (L_{ik} M_{kj}^{(n)} - M_{ik}^{(n)} L_{kj}) \\ &= -\sum_k [L_{ik} D_{kj} (L^{n-1})_{kj} - L_{kj} D_{ik} (L^{n-1})_{ik}] \\ &\quad + iA D_{ij} \sum_k [D_{jk} (L^{n-1})_{jk} - D_{ik} (L^{n-1})_{ik}]. \end{aligned}$$

By repeated use of

$$L_{ij} = P_{ij} + iA D_{ij} \tag{A2}$$

we then obtain

$$(n-1)![L,M^{(n)}]_{ij}=(p_j-p_i)\mathcal{D}_{ij}(L^{n-1})_{ij} + iA\sum_k \mathcal{D}_{kj}\mathcal{D}_{ik}[(L^{n-1})_{ik}-(L^{n-1})_{kj}] + \mathcal{D}_{ij}[(L^n)_{ii}-(L^n)_{jj}-p_i(L^{n-1})_{ii} + p_j(L^{n-1})_{jj}]. \tag{A3}$$

Using

$$\mathcal{D}_{kj}\mathcal{D}_{ik}=\mathcal{D}_{ij}(\mathcal{D}_{kj}-\mathcal{D}_{ki}),\quad i\neq j,\ k\neq i,j,$$

and again (A2), we derive from (A3)

$$(n-1)![L,M^{(n)}]_{ij}=-iA\mathcal{D}_{ij}^2[(L^{n-1})_{ii}-(L^{n-1})_{jj}].$$

This ends Case 2.

APPENDIX B

In this appendix we derive another way of solving the equations of motion of a system with generalized Hamiltonian

$$H_f=\sum_{k=0}^{\infty} f_k H_k^0=\mathrm{Tr}f(L).$$

A generalization of (3.5) leads to the definition of $h_n^{(f)}$:

$$h_n^{(f)}=\mathrm{Tr}\left[\frac{df(L)}{dL}+\lambda Q\right]^n,\quad n=1,\ 2,\ \cdots.$$

$h_n^{(f)}$ satisfies the differential equation:

$$\lambda^2\frac{\partial h_n^{(f)}}{\partial\lambda}+\frac{dh_n^{(f)}}{dt}=n\lambda h_n^{(f)}. \tag{B1}$$

If we define $H_{f,n}^p$,

$$H_{f,n}^p=\frac{1}{n!}\frac{\partial^p}{\partial\lambda^p}h_n^{(f)}\Big|_{\lambda=0},$$

we obtain from (B1)

$$\frac{d}{dt}H_{f,n}^p=p(n-p+1)H_{f,n}^{p-1},$$

whence

$$\frac{d^k}{dt^k}H_{f,n}^p=\begin{cases}\dfrac{p!(n-p+k)!}{(p-k)!(n-p)!}H_{f,n}^{p-k}, & k\leq p,\\ 0, & k>p.\end{cases}$$

The polynomial expansion in t of $H_{f,n}^p$ then has the following expression:

$$H_{f,n}^p(t)=\sum_{k=0}\binom{p}{k}\frac{(n-p+k)!}{(n-p)!}H_{f,n}^{p-k}(0)\cdot t^k,$$

which considering that $H_{f,n}^n=\mathrm{Tr}Q^n=\sum_i q_i^n$, leads to

$$\sum_i q_i^n=\sum_{k=0}^{n}\frac{n!}{(n-k)!}H_{f,n}^{n-k}(0)\cdot t^k,\quad n=1,\ 2,\cdots. \tag{B2}$$

If $f(L)=\frac{1}{2}L^2$ (that is, $H_f=H_2^0=H$), (B2) become

$$\sum_i q_i^n=\sum_{k=0}\frac{n!}{(n-k)!}H_n^{n-k}(0)\cdot t^k,\quad n=1,\ 2,\cdots. \tag{B3}$$

Given initial data $q_i(0)$, $p_i(0)$, it is possible to calculate all $H_n^{n-k}(0)$ and therefore $\sum_i q_i^n$ as functions of t from (B2), (B3). This yields another way of solving the equations of motion. It implies, of source, the solution of an algebraic equation of degree N since we know all symmetric polynomials in the q_i.

APPENDIX C

We have the easily derived P.b.:

$$K_s^m=[(s+m)!/(s-m)!]^{1/2}\cdot H_{2s}^{s-m},$$

$$s=0,\tfrac{1}{2},1,\tfrac{3}{2},..,\quad m=-s,-s+1,\ldots,s-1,s.$$

This normalization is inspired by the one used in the representation theory of SU(2).
We have the easily derived P.b.:

$$\{K^-,K^+\}=N,$$

where

$$K^+=K_{1/2}^{1/2}=\sum_i p_i,\quad K^-=K_{1/2}^{-1/2}=\sum_i q_i,$$

$$\{Q_{ij},K^+\}=\delta_{ij},\quad \{P_{ij},K^+\}=0,\quad \{L_{ij},K^+\}=0,$$
$$\{Q_{ij},K^-\}=0\quad \{P_{ij},K^-\}=-\delta_{ij},\quad \{L_{ij},K^-\}=-\delta_{ij}, \tag{C1}$$
$$\{K_s^m,K^+\}=(s-m)^{1/2}K_{s-1/2}^{m+1/2},\quad \{K_s^m,K^-\}=-(s+m)^{1/2}K_{s-1/2}^{m-1/2}.$$

Similarly,

$$\{K_1^1,K^+\}=0,\quad \{K_1^0,K^+\}=K^+,\quad \{K_1^{-1},K^+\}=\sqrt{2}K^-,$$
$$\{K_1^1,K^-\}=-\sqrt{2}K^+,\quad \{K_1^0,K^-\}=-K^-,\quad \{K_1^{-1},K^-\}=0, \tag{C2}$$
$$\{K_1^1,K_1^0\}=-2K_1^1,\quad \{K_1^0,K_1^{-1}\}=-2K_1^{-1},\quad \{K_1^{-1},K_1^1\}=2K_1^0.$$

Of crucial interest is the recursion relation for the Clebsch–Gordan coefficients:

$$\left[\frac{(a+s+t+1)(a-\alpha)}{2a+1}\right]^{1/2}(s,\ m,\ t,\ n\,|\,s,\ t,\ a,\ \alpha) =\left[\frac{(a+s-t)(s-m)}{2a}\right]^{1/2}(s-\tfrac{1}{2},m+\tfrac{1}{2},t,n\,|\,s-\tfrac{1}{2},t,a-\tfrac{1}{2},a+\tfrac{1}{2}), \tag{C3}$$

$$+\left[\frac{(a+t-s)(t-n)}{2a}\right]^{1/2}(s,m,t-\tfrac{1}{2},n+\tfrac{1}{2}\,|\,s,t-\tfrac{1}{2},a-\tfrac{1}{2},\alpha+\tfrac{1}{2}),$$

which can be derived from formula (3.7.12) of Ref. 3 through the use of the complete symmetries of the 3-j symbol.[4] From the general discussion in the text we know that the P.b.

$$\{K_s^m,K_t^n\}$$

is the trace of a polynomial in Q and L, and therefore, if suitably ordered according to Sec. 3, should have a leading term proportional to K_{s+t-1}^{m+n}. Therefore, we postulate the ansatz:

$$\{K_s^m,K_t^n\}=\sum_a\omega(a,s,t)(s,m,t,n\,|\,s,t,a,m+n)K_a^{m+n} + \text{multilinear terms in } K_b^\beta. \tag{C4}$$

The linear part of (C4) takes fully into account the conformal invariance of the theory. Hence we do not obtain any new information by using (C4) and the Jacobi's identity together with (C2). Of interest instead is the use of P.b. with $K^\pm$. Be writing explicitly

$$\{\{K_s^m,K_t^n\},K^\pm\}+\{\{K_t^n,K^\pm\},K_s^m\} +\{\{K^\pm,K_s^m\},K_t^n\}=0$$

using (C4), (C1) repeatedly and introducing (C3), we arrive at the recursion relation for ω:

$$\left(\frac{a+s-t}{a+s+t+1}\ \frac{2a+1}{2a}\right)^{1/2}\omega(a,s,t)=\omega(a-\tfrac{1}{2},s-\tfrac{1}{2},t). \quad \text{(C5)}$$

Similarly, from our ansatz (C4) and the symmetries of the Clebsch–Gordan coefficients, we obtain:

$$\omega(a,t,s)=-(-1)^{s+t-a}\omega(a,s,t). \quad \text{(C6)}$$

From (C6) we see that

$$\omega(2s-2p,s,s)=0.$$

Using the recursion relation (C5), we see that more generally

$$\omega(s+t-2p,s,t)=0, \quad p \text{ integer} \geq 0.$$

The general formula for the leading term is

$$\omega(s+t-1,s,t)=\left[(2s+2t)\binom{2s+2t-2}{2t-1}\right]^{1/2}. \quad \text{(C7)}$$

(C1) and (C2) are simple examples of (C7).

Note added in manuscript: It was kindly pointed out by the referee that the results in Sec. 2 have been found independently at an earlier date by M. Adler and presented at the 1976 Symposium in Tucson and are as well present in his dissertation thesis at New York University. Also the bibliography on Olshanetsky and Perelomov omits a series of papers in Inv. Math. We regret these and other omissions.

*Research sponsored in part by the National Science Foundation Grant No. GP-40768X.

[1]F. Calogero, J. Math. Phys. **12**, 419 (1971); Lett. Nuovo Cimento **13**, 411 (1975); **16**, 22 (1976); F. Calogero, O. Ragnisco, and C. Marchioro, Lett. Nuovo Cimento **13**, 383 (1975); B. Sutherland, Phys. Rev. A **4**, 2019 (1971); A **5**, 1372 (1972); J. Moser, Adv. Math. **16**, 197 (1975); M.A. Olshanetsky and A.M. Perelomov, Lett. Nuovo Cimento **16**, 333 (1976). References to other papers on this subject can be found in the papers quoted above.

[2]V. de Alfaro, S. Fubini and C. Furlan, "Conformal Invariance in Quantum Mechanics" (CERN preprint, 1976).

[3]A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1957).

[4]T. Regge, Nuovo Cimento **10**, 544 (1958).

Conditional expectations in generalized probability theory

H. Cycon and K.-E. Hellwig

Institut für Theoretische Physik der Technischen Universität Berlin, Germany
(Received 29 November 1976)

Expectations have been considered as dual objects of instruments in several papers on generalized probability theory and quantum theory. Here, we relate a generalized conditional (GC) expectation to a given instrument and a given state by two requirements, which are analogous to the axioms by which the classical conditional expectation is related to a given sub-σ-algebra and a given probability measure. In examples we illustrate the close similarity of GC expectations with classical conditional expectations. Eventually, we introduce a rich class of quantum stochastic processes, which are Markovian.

1. INTRODUCTION

In classical probability theory conditional expectations are related to a probability measure μ on a Borel space $(\Omega,\mathcal{M})$ and a sub-σ-algebra $\mathcal{M}_c$ of $\mathcal{M}$. The conditional expectation $\mathcal{E}_c(f)$ of a given random variable f on $(\Omega,\mathcal{M})$ is determined up to a μ-null set by two axioms: $\mathcal{E}_c(f)$ has to be $\mathcal{M}_c$-measurable and to fulfill $\int_M \mathcal{E}_c(f)d\mu = \int_M f d\mu$ for any $M\in\mathcal{M}_c$. Both axioms, which in the following are referred to as the classical averaging requirement and the classical mean value requirement, respectively, are essential features which make conditional expectations meaningful in the theory of martingales and stochastic processes.

In generalized probability theory expectations have been considered which are related to instruments. The instrument replaces in a sense, which we will describe below, the supposition of a sub-σ-algebra in the classical theory. So far, averaging and mean value requirements have not been considered seriously in many papers about the generalized theory. We attempt to formulate such requirements and find them suitable to relate close analogs to classical conditional expectations to certain instruments. We call them generalized conditional expectations or, for short, GC expectations, with respect to the given instrument and a given state.

We now describe how classical conditional expectation can be related to an instrument. Denote by $\mathcal{V}$ the space of bounded Borel measures on $(\Omega,\mathcal{M})$. Then the mapping

$$\mathcal{J}:\mathcal{M}_c\times\mathcal{V}\to\mathcal{V},$$
$$(M,\nu)\mapsto\nu(M\cap(\cdot))$$

defines an instrument. It has to be interpreted as follows. Consider a statistical ensemble the state of which is given by the probability measure μ on $(\Omega,\mathcal{M})$.

Assume $\mu(M)\neq 0$, $M\in\mathcal{M}_c$. Then

$$\mu_M := [1/\mu(M)]\,\mathcal{J}(M,\mu),$$

known as conditional probability, describes the state of the subensemble selected by occurence of the event M. Adopting the classical philosophy that events can be observed without disturbing the state of the statistical system and that such observations are the only natural ones, this instrument is canonically coordinated $\mathcal{M}_c$. The mean value requirement

$$\int f d\mu_M = \int \mathcal{E}_c(f)d\mu_M$$

just expresses that the expectation values of f and $\mathcal{E}_c(f)$ have to coincide on subensembles selected by occurence of a fixed event M, $M\in\mathcal{M}_c$.

We shall adapt our classical consideration a little more to the situation we encounter in the generalized theory. Allowing for incompatible observables, the basic structure of the generalized theory is more involved, and it is impossible to consider events independently of observables. Assume, therefore, $\mathcal{M}_c$ to be given by the range of $g^{-1}\colon\mathcal{L}\to\mathcal{M}$, where g is some Borel function on $(\Omega,\mathcal{M})$ and $\mathcal{L}$ denotes the σ-algebra of Borel sets on the real line. Using the same symbol as above, we consider now the instrument

$$\mathcal{J}\colon\mathcal{L}\times\mathcal{V}\to\mathcal{V}$$
$$(E,\nu)\mapsto\nu(g^{-1}(E)\cap(\cdot)).$$

Denote by $\tilde{\mathcal{V}}'$ the space of bounded Borel functions on $(\Omega,\mathcal{M})$. Any Borel function f on $(\Omega,\mathcal{M})$ defines a $\mathcal{V}'$-valued measure χ^f on $\mathcal{L}$, where $\chi^f(E)$, $E\in\mathcal{L}$, is the characteristic function of $f^{-1}(E)$. Classical conditional expectation is closely adapted to the notion of expectation in the generalized theory if we consider the mapping

$$\mathbb{E}_{\mathcal{J},\mu}\colon\mathcal{L}\times\tilde{\mathcal{V}}'\to\tilde{\mathcal{V}}',$$
$$(E,h)\mapsto\chi^g(E)\mathcal{E}_c(h).$$

Denote by $\tilde{\mathcal{C}}_c$ the space of $\mathcal{M}_c$-measurable functions on Ω. Then the averaging requirement can be expressed by

$$\mathbb{E}_{\mathcal{J},\mu}(E,h)\in\tilde{\mathcal{C}}_c \tag{1.1}$$

for any $h\in\tilde{\mathcal{V}}'$ and any $E\in\mathcal{L}$. The classical mean value requirement can be expressed by

$$\int h d\mathcal{J}(E,\mu)=\int\mathbb{E}_{\mathcal{J},\mu}(E,h)d\mu \tag{1.2}$$

for any $h\in\tilde{\mathcal{V}}'$, any $E\in\mathcal{L}$, and the fixed probability measure μ.

Both formulas (1.1) and (1.2) maintain their meaning in generalized probability theory if they are reformulated in the appropriate terms, which will be done in Definitions 3.2 and 3.3. The interpretations just mentioned remain also the same if we forget the classical assumption of nondisturbance of state if subensembles were selected. This assumption cannot be reconciled with the existences of incompatible observables. Hence, state changes other than the classical ones have always to be taken into account. We do not restrict our consideration on ideal instruments in the Lüders and von Neumann sense. We accept that for any given observable

there are many different instruments describing different state changes and relate, therefore, GC expectations not to observables, but to instruments.

Let us add one further remark. Expectations are defined by the characteristic properties of dual objects of instruments and, in many papers, only related to those instruments to which they are dual. Our GC expectation is related to an instrument by averaging and mean value requirements and is, in general, not the dual object of that instrument. Nevertheless, in some cases one can find another instrument to which the GC expectation is related in both ways. We show such classical and quantum mechanical cases in Sec. 4. An interesting point of quantum theory is that formulating a quantum stochastic process with an instrument related to a GC expectation in both ways is Markovian.

We review shortly the main concepts of generalized probability theory which we will use in Sec. 2. Section 3 contains the definition of GC expectations and two theorems stating such expectations to exist. Examples will be given in Sec. 4. In the Appendix we develop a functional calculus which we need for our constructions.

2. SOME CONCEPTS OF GENERALIZED PROBABILITY THEORY

For the most part, definitions and main concepts introduced by Davies and Lewis[1] are maintained. For details we refer the reader to that paper and to Edwards.[2]

The basic setting is a state space (V,B), where V is a complete base norm space and B is the base of a closed generating cone K of positive elements in V. $K\setminus\{0\}$ is the set of possible states of a statistical ensemble, and B is the set of normed states.

The Banach dual space of (V,B) is an order unit space (V',e) with order unit e. The positive cone K' in V' is $\sigma(V',V)$-closed. Observables are certain V'-valued measures on some Borel space $(X,\mathcal{A})$. Such a measure α is an observable if the range of α is contained in the order interval $[0,e]$, $\alpha(X)=e$, and α is σ-additive in the $\sigma(V',V)$-topology. In this paper we take for $(X,\mathcal{A})$ the Borel space $(\mathbb{R},\mathcal{L})$ of the real line. Sometimes we will also use $(I,\mathcal{P}(I))$, where $I\subseteq\mathbb{N}$ and $\mathcal{P}(I)$ denotes the power set of I. The elements of $[0,e]$ are called simple observables.

Denote by $L(V,V)$ the ordered vector space of bounded linear operators on V. Instruments are certain positive σ-additive $L(V,V)$-valued measures on $(\mathbb{R},\mathcal{L})$ or, sometimes, on $(I,\mathcal{P}(I))$. Such a measure $\mathbb{I}$ is an instrument if $\langle e,\mathbb{I}(\mathbb{R},v)\rangle=\langle e,v\rangle$ and if $\mathbb{I}(\cdot,v)$ is σ-additive with respect to the strong operator topology for any $v\in V$. We have preferred to write $\mathbb{I}(E,v)$ instead of $\mathbb{I}(E)v$, $E\in\mathcal{L}$, $v\in V$.

Any instrument determines uniquely an observable $\alpha_{\mathbb{I}}$ by $\alpha_{\mathbb{I}}(E):=\mathbb{I}^*(E,e)$, $E\in\mathcal{L}$, where $\mathbb{I}^*(E,\cdot)$ is the transposed operator of $\mathbb{I}(E,\cdot)$ in $L(V',V')$. We often refer to $\mathbb{I}^*(\cdot,\cdot)$ as the dual object of the instrument $\mathbb{I}$. For any given observable β there are, conversely, instruments $\mathbb{I}$ which determine β by $\beta=\alpha_{\mathbb{I}}$.

As already mentioned in the Introduction, instruments are to be interpreted in the following way: For a given statistical ensemble with state $v_0\in B$ the subensemble, which has been selected by occurrence of the event E, $E\in\mathcal{L}$, on the scale of some observable β in a certain way described by an instrument $\mathbb{I}$, is in state $\|\mathbb{I}(E,v_0)\|^{-1}\mathbb{I}(E,v_0)\in B$.

A state $v_0\in B$ is called effective with respect to a given instrument $\mathbb{I}$ if $\langle\alpha_{\mathbb{I}}(E),v_0\rangle=0$ implies $\alpha_{\mathbb{I}}(E)=0$, $E\in\mathcal{L}$. For any given instrument effective states exist if B is separable (e.g., Ludwig,[3] Theorem 4.2).

A instrument is called nuclear whenever it can be represented by an (α)-integral as

$$\mathbb{I}(E,\cdot)=\int_E\varphi d\langle\alpha_{\mathbb{I}},\cdot\rangle$$

for any $E\in\mathcal{L}$. Here φ is a V-valued α-measurable function on $\mathbb{R}$ taking values in B $\alpha_{\mathbb{I}}$-almost everywhere. For details of the (α)-integral we refer to the Appendix.

3. GC EXPECTATIONS

Throughout this section we consider a fixed state space (V,B) and is dual order unit norm space (V',e). A detailed definition of expectations as considered by Davies[4] is the following.

Definition 3.1: An expectation is a mapping $\mathbb{E}$: $\mathcal{L}\times V'\to V'$ such that

(i) $\mathbb{E}(E,a)\geq 0$ for any $E\in\mathcal{L}$ whenever $a\geq 0$, $a\in V'$,

(ii) $\mathbb{E}(E,\cdot)\in\mathcal{L}(V',V')$ for any $E\in\mathcal{L}$

(iii) $\mathbb{E}(\cdot,a)$ is σ-additive with respect to the $\sigma(V',V)$-topology for any $a\in V'$,

(iv) for any monotone sequence $\{a_n\}_{n\in\mathbb{N}}\subseteq V'$ converging to $a\in V'$ in the $\sigma(V',V)$-topology and any fixed $E\in\mathcal{L}$ the monotone sequence $\{\mathbb{E}(E,a_n)\}_{n\in\mathbb{N}}$ converges to $\mathbb{E}(E,a)$ in that topology,

(v) $\mathbb{E}(\mathbb{R},e)=e$.

We enter into the discussion of the introduction and suppose an instrument $\mathbb{I}$ to be given. To reformulate the averaging requirement we denote by $C_{\mathbb{I}}$ the $\sigma(V',V)$-closure of the span of range of $\alpha_{\mathbb{I}}$, i.e.,

$$C_{\mathbb{I}}:=\overline{\text{span}}\,\{x\,|\,x\in V',\ x=\alpha_{\mathbb{I}}(E),\ E\in\mathcal{L},\ \text{suitable}\}.$$

Note that $\alpha_{\mathbb{I}}$ replaces the random variable g, or, more closely, χ^g, in the classical case.

Definition 3.2: An expectation $\mathbb{E}$ is said to satisfy the averaging requirement with respect to the instrument $\mathbb{I}$, if for any $a\in V'$ and any $E\in\mathcal{L}$

$$\mathbb{E}(E,a)\in C_{\mathbb{I}}$$

holds. Such an expectation will henceforth be denoted by $\mathbb{E}_{\mathbb{I}}$.

Recall that for a given state $v_0\in B$, B being the base of V, $\|\mathbb{I}(E,v_0)\|^{-1}\mathbb{I}(E,v_0)$ is interpreted as the state of the subensemble selected by occurrence of $E\in\mathcal{L}$ on the scale of $\alpha_{\mathbb{I}}$, and the interpretation of the classical mean value requirement given in the Introduction. This motivates the following definition.

Definition 3.3: An expectation $\mathbb{E}_{\mathbb{I}}$ is said to satisfy

the mean value requirement with respect to the instrument $\mathbb{I}$ and the state $v_0 \in B$, if for any simple observable a, i.e., $a \in [0,e]$, and any $E \in \mathcal{L}$

$$\langle a, \mathbb{I}(E, v_0)\rangle = \langle \mathbb{E}(E,a), v_0\rangle$$

holds. Such an expectation will henceforth be denoted by $\mathbb{E}_{\mathbb{I},v_0}$.

We shall drop the terms "with respect to ..." if $\mathbb{I}$ or v_0 are understood. We now introduce the concept of generalized conditional expectation, GC expectation for short.

Definition 3.4 For a given state $v_0 \in B$ and a given instrument $\mathbb{I}$ we call an expectation $\mathbb{E}_{\mathbb{I},v_0}$ a generalized conditional expectation with respect to $(\mathbb{I}, v_0)$ if both the averaging and the mean value requirements hold.

The following theorems show under which assumptions GC expectations exist.

Theorem 3.1: GC expectations with respect to $(\mathbb{I}, v_0)$ exist whenever v_0 is effective with respect to $\mathbb{I}$.

Proof: We divide the proof into three parts. In part (1) we construct a mapping $\mathbb{E}_{\mathbb{I},v_0}: \mathcal{L} \times V' \to V'$. In part (2) we show it to be an expectation, and in part (3) that it is indeed a GC expectation.

(1) For any given $a \in V'$ we denote by ν_a the real measure on $(\mathbb{R}, \mathcal{L})$ defined by $\nu_a(E) := \langle a, \mathbb{I}(E, v_0)\rangle$, $E \in \mathcal{L}$. From $-(\|a\|+\epsilon)e < a < (\|a\|+\epsilon)e$ for any $\epsilon > 0$, and from $\mathbb{I}(E, v_0) \geq 0$ for any $E \in \mathcal{L}$, we conclude that $\nu_e(E) = 0$ implies $\nu_a(E) = 0$. Thus, the Radon–Nikodym density ϵ_a exists such that $\nu_a(E) = \int_E \epsilon_a d\nu_e$. We show that $|\epsilon_a(s)| \leq 4\|a\|$, $s \in \mathbb{R}$, holds ν_a-almost everywhere. Let $tv(\nu_a(\cdot))$ denote the total variation of ν_a. Then we have (e.g., Dunford and Schwartz,[5] III. 1.5) $tv(\nu_a(E)) \leq 4 \sup_{F \subseteq E} |\nu_a(F)| \leq 4\|a\|\nu_e(E)$, which gives the desired inequality if we use $tv(\nu_a(E)) = \int_E |\epsilon_a| d\nu_e$. Since ν_0 is effective, we have $\nu_e(E) = 0$ if and only if $\alpha_{\mathbb{I}}(E) = 0$. Hence, ϵ_a is $\alpha_{\mathbb{I}}$-measurable because it is ν_e-measurable. Moreover, since ϵ_a turns out to be $\alpha_{\mathbb{I}}$-essentially bounded for the same reason, we conclude that ϵ_a is $\alpha_{\mathbb{I}}^*$-integrable [e.g., Theorem A1, (iii) in the Appendix]. Thus we can define the mapping $\mathbb{E}_{\mathbb{I},v_0}: \mathcal{L} \times V' \to V'$ by

$$\mathbb{E}_{\mathbb{I},v_0}(E,a) := \int_E \epsilon_a d\alpha_{\mathbb{I}}, \quad E \in \mathcal{L}, \ a \in V'.$$

(2) We now prove that $\mathbb{E}_{\mathbb{I},v_0}$ is an expectation. Since $a \mapsto \nu_a$, $\nu_a \mapsto \epsilon_a$, and $\epsilon_a \mapsto \int_E \epsilon_a d\alpha_{\mathbb{I}}$ are positive linear mappings, (i) and (ii) of Definition 3.1 are obvious. (v) is also immediate. (iii) is a direct consequence of Theorem A1, (ii) in the Appendix. We have only to prove (iv): Let $\{a_n\}_{n \in \mathbb{N}}$ be a monotone sequence in V' converging to a in the $\sigma(V', V)$-topology. Then, for any $E \in \mathcal{L}$, $\{\nu_{a_n}(E)\}_{n \in \mathbb{N}}$ is monotone and converges to $\nu_a(E)$. Straightforward calculations show that $\{\epsilon_{a_n}\}_{n \in \mathbb{N}}$ is monotone $\alpha_{\mathbb{I}}$-almost everywhere and converges to ϵ_a $\alpha_{\mathbb{I}}$-almost everywhere. From Lemma A2 in the Appendix, we conclude that $\{\int_E \epsilon_{a_n} d\alpha_{\mathbb{I}}\}_{n \in \mathbb{N}}$ is monotone and converges to $\int_E \epsilon_a d\alpha_{\mathbb{I}}$ in the $\sigma(v,v)$-topology. This completes the proof that $\mathbb{E}_{\mathbb{I},v_0}$ is an expectation.

(3) By definition of the $\alpha_{\mathbb{I}}^*$-integral in the Appendix, Defintion A2, $\mathbb{E}_{\mathbb{I},v_0}$ fulfills the averaging requirement. To prove the mean value requirement, recall that the density ϵ_a is defined in part (1) of the proof by

$$\langle a, \mathbb{I}(E, v_0)\rangle = \int_E \epsilon_a d\langle \alpha_{\mathbb{I}}, v_0\rangle.$$

and that, moveover, ϵ_a is $\alpha_{\mathbb{I}}$-measurable and $\alpha_{\mathbb{I}}$-essentially bounded. From Theorem A2 in the Appendix we have

$$\int_E \epsilon_a d\langle \alpha_{\mathbb{I}}, v_0\rangle = \langle \int_E \epsilon_a d\alpha_{\mathbb{I}}, v_0\rangle.$$

Thus, the mean value reqirement is also fulfilled. This shows that $\mathbb{E}_{\mathbb{I},v_0}$ is indeed a GC expectation, which completes the proof of Theorem 3.1.

We mentioned in the Introduction that dual objects of instruments have been considered as expectations. In general, such dual objects $\mathbb{I}^*$ do not fulfill the averaging requirement with respect to $\mathbb{I}$. An example is given by the following situation.

Take a_1, $a_2 \in [0,e]$, $a_1 \neq 0$, $a_2 \neq 0$, $a_1 \neq a_2$, and $a_1 + a_2 \in [0,e]$. Such a situation arises in quantum theory if a_i are taken to be two orthogonal projections on mutual orthogonal subspaces. Let $a := a_1 + a_2$. Take, furthermore, two states v_1, $v_2 \in B$, i.e., $\langle e, v_1\rangle = 1$ and $\langle e, v_2\rangle = 1$, such that $\langle a, v_1\rangle = \langle a, v_2\rangle$ but with $\langle a_1, v_1\rangle \neq \langle a_1, v_2\rangle$. An instrument $\mathbb{I}: \mathcal{L} \times V \to V$ is uniquely determined by

$$\mathbb{I}((-\infty, s], v) := \begin{cases} 0 & \text{for} \quad s < 0, \\ \langle a_1, v\rangle v_1 + \langle a_2, v\rangle v_2 & \text{for } 0 \leq s < 1, \\ v & \text{for } 1 \leq s. \end{cases}$$

From $a = a_1 + a_2$ one easily checks that the range of $\alpha_{\mathbb{I}}$ is given by $\{0, a, e\}$. Hence $\mathcal{C}_{\mathbb{I}} = \mathrm{span}\{a, e\}$.

Assume $\mathbb{I}^*$ to fulfill the averaging requirement with respect to $\mathbb{I}$. Then $\mathbb{I}^*(\{0\}, a_1) \in \mathcal{C}_{\mathbb{I}}$ would hold and so from $\langle a, v_1\rangle = \langle a, v_2\rangle$, we would have

$$\langle \mathbb{I}^*(\{0\}, a_1), v_1\rangle - \langle \mathbb{I}^*(\{0\}, a_1), v_2\rangle = 0.$$

But actually we have

$$\langle a_1, \mathbb{I}(\{0\}, v_1)\rangle = \langle a_1, v_1\rangle\langle a_1, v_1\rangle + \langle a_2, v_1\rangle\langle a_1, v_2\rangle,$$

$$\langle a_1, \mathbb{I}(\{0\}, v_2)\rangle = \langle a_1, v_2\rangle\langle a_1, v_1\rangle + \langle a_2, v_2\rangle\langle a_1, v_2\rangle.$$

Using $\langle a, v_1\rangle = \langle a, v_2\rangle$ and $a = a_1 + a_2$ to eliminate a_2 from these equations, simple computation leads to

$$\langle \mathbb{I}^*(\{0\}, a_1), v_1\rangle - \langle \mathbb{I}^*(\{0\}, a_1), v_2\rangle = (\langle a_1, v_1\rangle - \langle a_1, v_2\rangle)^2.$$

By construction the right-hand side is unequal to zero, which is a contradiction.

Nevertheless, there are instruments the dual objects of which fulfill the averaging requirement and are, therefore, GC expectations. A class of such instruments is given by the nuclear ones. Such instruments have been constructed in the proof of Theorem 1 in Davies and Lewis.[1] Let an observable $\beta: \mathcal{L} \to V'$ be given. A nuclear instrument $\mathbb{I}$, such that $\alpha_{\mathbb{I}} = \beta$ is constructed for any countable partition of $\mathbb{R}$ into disjoint Borel sets E_i, $i \in I$, $I \subseteq \mathbb{N}$, and for any choice $\{v_i\}_{i \in I} \subseteq B$ by

$$\mathbb{I}(E, v) := \sum_{i \in I} v_i \langle \beta(E \cap E_i), v\rangle, \quad E \in \mathcal{L}, \ v \in V.$$

Thus, given an observable, there is likewise a rich set of nuclear instruments describing measuring processes for this observable, and the dual objects are GC-

expectations as will be shown by the following theorem.

Theorem 3.2: The dual object $\mathbb{I}^*$ of any nuclear instrument $\mathbb{I}$ is a GC expectation with respect to $(\mathbb{I}, v)$ for any normed state $v \in B$.

Proof: Nuclear instruments are represented by $(\alpha_{\mathbb{I}})$-integrals as introduced in the Appendix. For any $(E, v) \in \mathcal{L} \times V$ we have

$$\mathbb{I}(E, v) = \int_E \varphi d\langle \alpha_{\mathbb{I}}, v\rangle,$$

where $\varphi: \mathbb{R} \to V$ is a fixed V-valued function which is $\alpha_{\mathbb{I}}$-measurable and takes $\alpha_{\mathbb{I}}$-almost everywhere values in B. For any given $a \in V'$ we denote by f_a the real function defined by $f_a(s) := \langle a, \varphi(s)\rangle$, $s \in \mathbb{R}$, which, by Theorem A1 in the Appendix, is $\alpha_{\mathbb{I}}^*$-integrable. From the Corollary of Theorem A3, formula (ii), we have for the dual object of $\mathbb{I}$

$$\mathbb{I}^*(E, a) = \int_E f_a d\alpha_{\mathbb{I}}$$

for any $(E, a) \in \mathcal{L} \times V'$. $\mathbb{I}^*$ is an expectation because it is the dual object of an instrument. The same argument implies that the mean value requirement is satisfied, where v is arbitrary in B. By definition of the $\alpha_{\mathbb{I}}^*$-integral the averaging requirement is obvious. This completes the proof.

4. SOME EXAMPLES

We first describe the classical case more closely as a special case of the generalized theory. Given a Borel space $(\Omega, \mathcal{M})$, the state space $(V, \mathcal{B})$ is the ordered linear space V of real bounded Borel measures on $\mathcal{M}$ and $\mathcal{B}$ is the subset of probability measures. The space $\tilde{V}'$ of bounded Borel functions on $(\Omega, \mathcal{M})$, which is an order unit space, is embedded into (V', e) by $\langle h, \nu\rangle = \int h d\nu$ where $h \in V'$, $\nu \in V$, The function χ_Ω is identified with the order unit e.

In the Introduction, we have already considered the classical ideal instrument $\mathcal{J}$ on $\mathcal{L} \times V$ defined by $\mathcal{J}(E, \nu) = \nu(g^{-1}(E \cap (\cdot)))$ and the mapping $\mathbb{E}_{\mathcal{J}, \mu}$ on $\mathcal{L} \times \tilde{V}'$ which has, for elements h of the embedded space V', all properties of a GC expectation with respect to $(\mathcal{J}, \mu)$.

For convenience we shall restrict ourselves here to the case when the range of g is I, $I \subseteq \mathbb{N}$. We remark that all arguments would also go through equally well if the more general case of an expectation kernel were considered. Let us assume $\mathcal{M}_c$, the sub-σ-algebra of $\mathcal{M}$, to be given by the range of g^{-1}: $\mathcal{P}(I) \to \mathcal{M}$. Then the instrument $\mathcal{J}$ can be written more simply as

$$\mathcal{J}: \mathcal{P}(I) \times V \to V,$$
$$(m, \nu) \mapsto \nu(g^{-1}(m) \cap (\cdot))$$

and, obviously, $\alpha_{\mathcal{J}} = \chi^g$, where, as in the Introduction, χ^g denotes the V'-valued measure on $\mathcal{L}$ defined by $\chi^g(E) := \chi_{g^{-1}(E)}$, $E \in \mathcal{L}$. Since $\Omega = \bigcup \{M_i \mid M_i := g^{-1}(\{i\}),\ i \in I\}$, we have for $h \in \tilde{V}'$ the classical conditional expectation with respect to $\mathcal{M}_c$ and $\mu \in \mathcal{B}$

$$\mathcal{E}_c(h) = \sum_{i \in I} \frac{1}{\mu(M_i)} \int_{M_i} h d\mu \ \chi_{M_i}.$$

Hence, on $\tilde{V}'$, we have

$$\mathbb{E}_{\mathcal{J}, \mu}: \mathcal{P}(I) \times \tilde{V}' \to \tilde{V}',$$
$$(m, h) \mapsto \sum_{i \in m} \left(\frac{1}{\mu(M_i)} \int_{M_i} h d\mu \right) \chi_{M_i}.$$

One checks easily that $\mathbb{E}_{\mathcal{J}, \mu}$ is the restriction to $\mathcal{P}(I) \times \tilde{V}'$ of the dual object of

$$\tilde{\mathcal{J}}: \mathcal{P}(I) \times V \to V,$$
$$(m, \nu) \mapsto \sum_{i \in m} \nu(M_i) \frac{1}{\mu(M_i)} \mu(M_i \cap (\cdot))$$

on $\mathcal{P}(I) \times V'$. The instrument $\tilde{\mathcal{J}}$ differs in general from $\mathcal{J}$. Now, $\alpha_{\tilde{\mathcal{J}}} = \alpha_{\tilde{\mathcal{J}}} = \chi^g$ and, defining

$$\mathcal{J}: I \to \mathcal{B},$$
$$i \mapsto \mu_{M_i} := \frac{1}{\mu(M_i)} \mu(M_i \cap (\cdot)),$$

$\tilde{\mathcal{J}}$ can be written in the form

$$\tilde{\mathcal{J}}(m, \nu) = \sum_{i \in m} \varphi(i) \langle \alpha_{\mathcal{J}}(\{i\}), \nu\rangle.$$

Hence, $\tilde{\mathcal{J}}$ is a nuclear instrument and from Theorem 3.2 we conclude that $\tilde{\mathcal{J}}^*$ is a GC expectation with respect to $\tilde{\mathcal{J}}$ and any state $\nu \in V$. Since $\alpha_{\tilde{\mathcal{J}}} = \alpha_{\tilde{\mathcal{J}}}$, $\tilde{\mathcal{J}}^*$ also fulfills the averaging requirement with respect to $\mathcal{J}$. Moreover, for any $m \in M$ we have $\tilde{\mathcal{J}}(m, \mu) = \mathcal{J}(m, \mu)$ for the fixed state μ. Therefore, $\mathcal{J}^*$ also satisfies the mean value requirement with respect to $(\mathcal{J}, \mu)$. Thus $\mathcal{J}^*$ extends $\mathbb{E}_{\mathcal{J}, \mu}$ as a GC expectation with respect to $(\mathcal{J}, \mu)$ from $\mathcal{P}(I) \times \tilde{V}'$ onto $\mathcal{P}(I) \times V'$.

By the latter part of the example we have shown that the classical conditional expectation, written as a GC expectation $\mathbb{E}_{\mathcal{J}, \mu}$, is not the dual object of the ideal instrument $\mathcal{J}$, but of another instrument $\tilde{\mathcal{J}}$. We find exactly this situation in the following quantum mechanical example.

Consider quantum mechanics in Hilbert space. The state space is given by $(V_H, \mathcal{B}_H)$, where V_H is the real space of Hermitian trace class operators on the separable Hilbert space H, and B_H is the subset of positive elements of V_H with unit trace. The Banach dual space (V', e_H) is norm-isomorphic to the real Banach space of bounded Hermitian operators on H. A bounded linear operator A is identified with the linear functional defined by $\langle A, W\rangle := \mathrm{tr}(AW)$, $W \in V_H$. The unit operator 1_H is identified with the order unit e_H.

For simplicity we consider a discrete partition of the unit operator into orthogonal projections $\{P_i\}_{i \in I}$, $I \subseteq \mathbb{N}$, $1_H = \sum_{i \in I} P_i$. Then

$$\mathcal{J}: \mathcal{P}(I) \times V_H \to V_H,$$
$$(m, W) \mapsto \sum_{i \in m} P_i W P_i,$$

the ideal instrument of Lüders and von Neumann,[6] is known to be the closest analog of the classical instrument $\mathcal{J}$ just considered. $\alpha_{\mathcal{J}}$ is given by

$$\alpha_{\mathcal{J}}: \mathcal{P}(I) \to V',$$
$$m \mapsto \sum_{i \in m} P_i.$$

We first show that the dual object

$$\mathcal{J}^*: \mathcal{P}(I)\times \mathcal{V}_H' \to \mathcal{V}_H',$$
$$(m,F) \mapsto \sum_{i\in m} P_i F P_i$$

is in general not a GC expectation with respect to $\mathcal{J}$ and any state.

In the general case at least one projection, say P_{i_0}, $i_0 \in I$, may be nonatomic (i.e., rank $P_{i_0} > 1$). Then there is a nonzero projection Q, $Q \neq P_{i_0}$, and $Q \leq P_{i_0}$. From $\mathcal{J}^*(\{i_0\}, Q) = Q$ we conclude $\mathcal{J}^*(\{i_0\}, Q) \notin \mathcal{C}_{\mathcal{J}}$, where $\mathcal{C}_{\mathcal{J}}$ is, as in Definition 3.2, the $\sigma(\mathcal{V}', \mathcal{V})$-closed span of the range of $\alpha_{\mathcal{J}}$, i.e., of $\{P_i \mid i \in I\}$. Thus $\mathcal{J}^*$ fails to be a GC expectation with respect to $\mathcal{J}$ and some state.

In the special case, where any P_i, $i \in I$, is atomic, however, $\mathcal{J}^*$ is a GC expectation with respect to $\mathcal{J}$ and any state. Defining

$$\varphi: I \to \mathcal{B}_H,$$
$$i \mapsto P_i,$$

and recalling that $\mathrm{tr}(P_i B)P_i = P_i B P_i$ for any bounded linear operator B whenever P_i is atomic, we find

$$\mathcal{J}(m, W) = \sum_{i\in m} \varphi(i)\, \langle \alpha_{\mathcal{J}}(\{i\}), W\rangle.$$

Hence, $\mathcal{J}$ is nuclear and by Theorem 3.2 our statement is proven.

We return to the general case. Let $W_0 \in \mathcal{B}_H$ be an effective state with respect to $\mathcal{J}$, i.e., $\mathrm{tr}(P_i W_0) \neq 0$ for any $i \in I$. Then

$$\mathbb{E}_{\mathcal{J}, W_0}: \mathcal{P}(I)\times \mathcal{V}_H' \to \mathcal{V}_H',$$
$$(m,F) \mapsto \sum_{i\in m} \left(\frac{1}{\mathrm{tr}(P_i W_0)} \mathrm{tr}(P_i W_0 P_i F)\right) P_i$$

is easily seen to be a GC expectation with respect to $(\mathcal{J}, W_0)$ and to coincide with $\mathcal{J}^*$ in the special atomic case. We remark that the construction scheme used in proof of Theorem 3.1 leads just to $\mathbb{E}_{\mathcal{J}, W_0}$. $\mathbb{E}_{\mathcal{J}, W_0}(I, \cdot)$ coincides with the conditional expectation $\mathbb{E}_{W_0}(\cdot, \mathbb{B})$ which has been introduced by Gudder and Marchand,[7] where $\mathbb{B}$ is just $\mathcal{C}_{\mathcal{J}}$.

$\mathbb{E}_{\mathcal{J}, W_0}$ is the dual object of

$$\tilde{\mathcal{J}}: \mathcal{P}(I)\times \mathcal{V}_H \to \mathcal{V}_H,$$
$$(m, W) \mapsto \sum_{i\in m} \mathrm{tr}(P_i W) \frac{P_i W_0 P_i}{\mathrm{tr}(P_i W_0)}.$$

Defining

$$\varphi_H: I \to \mathcal{V}_H,$$
$$i \mapsto \frac{P_i W_0 P_i}{\mathrm{tr}(P_i W_0)},$$

we have for any $m \in \mathcal{P}(I)$ and for any $W \in \mathcal{B}_H$

$$\tilde{\mathcal{J}}(m, W) = \sum_{i\in m} \varphi_H(i)\, \langle \alpha_{\mathcal{J}}(\{i\}), W\rangle,$$

which shows $\tilde{\mathcal{J}}$ to be nuclear. Since W_0 has been assumed to be effective with respect to $\mathcal{J}$, $\varphi_H(i) = P_i$ holds whenever P_i is atomic. Hence $\tilde{\mathcal{J}} = \mathcal{J}$ if this is the case for any $i \in I$. Application of Theorem 2 shows that $\tilde{\mathcal{J}}^*$, which is identical with $\mathbb{E}_{\mathcal{J}, W_0}$, is a GC expectation with respect to $\mathcal{J}$ and any state $W \in \mathcal{B}_H$.

We conclude this section with a remark on Markovicity of quantum stochastic processes. We enter into the discussion given in Sec. 5 of a paper by Srinivas.[8] A quantum stochastic process is a one-parameter family of instruments $\{\mathcal{J}_t\}_{t\in\mathbb{R}^+}$ on a quantum probability space $(\mathcal{V}_H, \mathcal{B}_H, W_0)$, $W_0 \in \mathcal{B}_H$. We continue to consider the discrete case and assume a partition of the unit operator in H which depends on the parameter t, I being fixed, $\{P_{i,t}\}_{(i,t)\in I\times\mathbb{R}^+}$.

Let us first consider the one-parameter family of ideal instruments

$$\mathcal{J}_t: \mathcal{P}(I)\times \mathcal{V}_H \to \mathcal{V}_H,$$
$$(m, W) \mapsto \sum_{i\in m} P_{i,t} W P_{i,t}.$$

$t \in \mathbb{R}^+$. As shown by Srinivas, the finite-dimensional distributions of the form

$$\mathbb{P}_r(i_1, t_1; i_2, t_2; \cdots; i_r, t_r)$$
$$= \langle 1_H, \mathcal{J}_{t_r}(i_r, \cdot) \circ \mathcal{J}_{t_{r-1}}(i_{r-1}, \cdot) \circ \cdots \circ \mathcal{J}_{t_1}(i_1, W_0)\rangle,$$
$$(i_1, t_1; i_2, t_2; \cdots; i_r, t_r) \in (I\times\mathbb{R}^+)^r,$$
$$t_1 \leq t_2 \leq \cdots \leq t_r,$$

have a Markov property which may be expressed by

$$\mathbb{P}_r(i_1, t_1; i_2, t_2; \cdots; i_r, t_r) = \mathbb{P}_1(i_1, t_1)\mathbb{W}(i_2, t_2 \mid i_1, t_1)$$
$$\times \mathbb{W}(i_3, t_3 \mid i_2, t_2) \cdots \mathbb{W}(i_r, t_r \mid i_{r-1}, t_{r-1}),$$

whenever any $P_{i,t}$, $(i,t) \in (I, \mathbb{R}^+)$, is atomic, $\mathbb{W}$ is the transition probability given by

$$\mathbb{W}(i_r, t_r \mid i_{r-1}, t_{r-1}; \cdots; i_1, t_1)$$
$$= \mathbb{P}_r(i_1, t_1; \cdots; i_r, t_r) / \mathbb{P}_{r-1}(i_1, t_1; \cdots; i_{r-1}, t_{r-1}).$$

In the nonatomic case the process is in general non-Markovian.

Replacing $\mathcal{J}_t$ by the dual objects of GC expectations with respect to $(\mathcal{J}_t, W_0)$,

$$\tilde{\mathcal{J}}_t: \mathcal{P}(I)\times \mathcal{V}_H \to \mathcal{V}_H,$$
$$(m, W) \mapsto \sum_{i\in m} \mathrm{tr}(P_{i,t} W) \frac{P_{i,t} W_0 P_{i,t}}{\mathrm{tr}(P_{i,t} W_0)},$$

$t \in \mathbb{R}^+$, we find distributions

$$\mathbb{P}_r(i_1, t_1; \cdots; i_r, t_r) = \langle 1_H, \tilde{\mathcal{J}}_{t_r}(i_r, \cdot) \circ \cdots \circ \tilde{\mathcal{J}}_{t_1}(i_1, W_0)\rangle$$
$$= \mathrm{tr}(P_{i_1,t_1} W_0) \frac{\mathrm{tr}(P_{i_2,t_2} P_{1,t_1} W_0 P_{1,t_1})}{\mathrm{tr}(P_{i_1,t_1} W_0)} \cdots$$
$$\times \frac{\mathrm{tr}(P_{i_r,t_r} P_{i_{r-1},t_{r-1}} W_0 P_{i_{r-1},t_{r-1}})}{\mathrm{tr}(P_{i_{r-1},t_{r-1}} W_0)}$$

also in case the $P_{i,t}$ are nonatomic. Hence, realizing that the transition probabilities $\mathbb{W}(i,t \mid j,s)$, $t \geq s$, are given by

$$\mathbb{W}(i,t \mid j,s) = \frac{\langle 1_H, \tilde{\mathcal{J}}_s(j, \cdot) \circ \tilde{\mathcal{J}}_t(i, W_0)\rangle}{\langle 1_H, \tilde{\mathcal{J}}_t(i, W_0)\rangle}$$
$$= \frac{\mathrm{tr}(P_{j,s} P_{i,t} W_0 P_{i,t})}{\mathrm{tr}(P_{i,t} W_0)},$$

we find the Markov property in the above form to be fulfilled.

APPENDIX: GENERALIZED FUNCTIONAL CALCULUS

In this appendix we develop an integration theory for vector valued functions and vector valued measures which is adapted to our purposes. Since we use mainly the technique of Dunford and Schwartz[5] and Bartle,[9] we keep proofs as short as possible and refer the reader to these references for analogous definitions and for the structure of calculations omitted here.

If ν is a finite real valued measure on the measure space $(\mathbb{R}, \mathcal{L})$, we denote by $tv(\nu(\cdot))$ the total variation of ν, which is a positive measure. Then the notions ν-almost everywhere, ν-simple, ν-measurable, ν-almost uniformly, ν-essentially bounded are to be understood with respect to $tv(\nu(\cdot))$. We remind the reader that by the theorem of Egoroff a sequence of ν-measurable functions converges ν-almost everywhere uniformly if and only if it converges ν-almost everywhere.

If (V, B) is a state space and α is an observable, we call a set $N \subseteq \mathbb{R}$ a α-null set if there exists an $E \in \mathcal{L}$ with $N \subseteq E$ and $\alpha(E) = 0$. α-simple functions, α-almost everywhere convergence, and α-essential boundedness are then defined in the obvious way. We call a function f on $\mathbb{R}$ α-measurable if there exists a sequence of α-simple functions converging to f α-almost everywhere. As a direct consequence of the theorem of Egoroff we have, for a sequence of α-measurable functions, that α-almost everywhere convergence to a function f implies $\langle \alpha, v \rangle$-almost uniform convergence to f for any $v \in V$.

If f is a function on $\mathbb{R}$ and $r \in \mathbb{R}$, $r > 0$, then $f^{(r)}$ shall denote the r-truncation of f, which is defined by

$$f^{(r)}(s) := \begin{cases} f(s) & \text{if } \|f(s)\| < r \\ f(s)/\|f(s)\| & \text{if } \|f(s)\| \geq r \end{cases}, \quad s \in \mathbb{R}.$$

Short calculation shows that if f is α-essentially bounded, $r = \alpha\text{-ess sup}(f)$, and if f is the limit of the α-almost everywhere convergent sequence $\{\varphi_n\}_{n \in \mathbb{N}}$ of α-simple functions, then f is as well the limit of the α-almost everywhere convergent sequence $\{\varphi_n^{(r)}\}_{n \in \mathbb{N}}$ of r-truncated α-simple functions. Hence, by the Egoroff theorem, any α-essentially bounded α-measurable function can be assumed to be for all $v \in V$ the $\langle \alpha, v \rangle$-almost uniform limit of a uniformly bounded sequence of α-simple functions. We will use this fact freely in the following.

For any α-simple real function with representation

$$\varphi := \sum_{i=1}^{n} x_i \chi_{E_i}, \quad x_i \in \mathbb{R}, \quad E_i \in \mathcal{L}, \quad i = 1, \ldots, n,$$

we define the α^*-integral of φ by

$$\int_E \varphi d\alpha := \sum_{i=1}^{n} x_i \alpha(E_i \cap E), \quad E \in \mathcal{L}.$$

We omit the standard proofs of the following statements:

(1) The α^*-integral is independent of the particular representation of the α-simple function φ.

(2) The α^*-integral is a linear mapping from the linear space of α-simple functions to V' for any $E \in \mathcal{L}$.

(3) The α^*-integral is a V'-valued σ-additive set function on $(\mathbb{R}, \mathcal{L})$ with respect to the $\sigma(V', V)$-topology for any α-simple function.

(4) For any α-simple function φ with $|\varphi(s)| \leq r$, $s \in E$, $r \in \mathbb{R}$, we have

$$\left| \left\langle \int_E \varphi d\alpha, v \right\rangle \right| \leq r\, tv(\langle \alpha(E), v \rangle) \leq 4r\|v\|, \quad E \in \mathcal{L}, \ v \in V.$$

Definition A1: Given an observable α of the state space (V, B) then a real function f on $\mathbb{R}$ is said to be α^*-integrable if there is a sequence $\{\varphi_n\}_{n \in \mathbb{N}}$ of α-simple real functions satisfying

(i) $\{\varphi_n\}_{n \in \mathbb{N}}$ converges to f α-almost everywhere and

(ii) the sequence $\{\int_E \varphi_n d\alpha\}_{n \in \mathbb{N}}$ converges in V' with respect to the $\sigma(V', V)$-topology for any $E \in \mathcal{L}$.

Lemma A1: If the real function f is α^*-integrable, then $\lim^*_{n \to \infty} \int_E \varphi_n d\alpha$ does not depend on the particular sequence $\{\varphi_n\}_{n \in \mathbb{N}}$ satsifying (i) and (ii) of Definition A1.

The proof of this lemma proceeds in almost the same way as that given for Theorem IV.10.8. (a) in Dunford and Schwartz.[5] For any two sequences $\{f_n\}_{n \in \mathbb{N}}$ and $\{g_n\}_{n \in \mathbb{N}}$ satisfying (i) and (ii) of Definition A1, define a sequence $\{h_n\}_{n \in \mathbb{N}}$ of α-simple real functions by $h_n(s) := 0$ for the exceptional points $s \in \mathbb{R}$ where $\{f_n\}_{n \in \mathbb{N}}$ or $\{g_n\}_{n \in \mathbb{N}}$ fail to converge to $f(s)$ and by $h_n(s) := f_n(s) - g_n(s)$, $s \in \mathbb{R}$, otherwise. For any $v \in V$ the Vitali–Hahn–Saks theorem applies to the sequence $\{\langle \int_{(\cdot)} h_n d\alpha, v \rangle\}_{n \in \mathbb{N}}$ of real valued $\langle \alpha, v \rangle$-continuous measures. The proof is then finished by the same arguments as given in Dunford and Schwartz.

Definition A2: If f is α^*-integrable, we write for any $E \in \mathcal{L}$

$$\int_E f d\alpha := \lim^*_{n \to \infty} \int_E \varphi_n d\alpha$$

and call it the α^*-integral of f over E.

Theorem A1: (i) The set of α^*-integrable functions is a linear space and for any $E \in \mathcal{L}$ the α^*-integral over E is a linear mapping from this space into V'.

(ii) For any α^*-integrable function f the α^*-integral of f is a σ-additive set function on $(\mathbb{R}, \mathcal{L})$ with respect to $\sigma(V', V)$.

(iii) If f is α-measurable and if $\alpha\text{-ess sup}(f) =: r < \infty$, then f is α^*-integrable and

$$\left| \int_E f d\alpha \right| \leq 4r, \quad E \in \mathcal{L}.$$

Proof: Using the Vitali–Hahn–Saks theorem, (i) and (ii) follow easily from the corresponding properties of the α-simple functions.

We prove (iii). Since f is α-measurable and α-essentially bounded it is the α-almost everywhere limit of a sequence $\{\varphi_n\}_{n \in \mathbb{N}}$ of α-simple functions uniformly bounded by r. Moreover, $\{\varphi_n\}_{n \in \mathbb{N}}$ converges to f $\langle \alpha, v \rangle$-almost uniformly for any $v \in V$. Hence only (ii) of Definition A1 remains to be shown. Take $v \in V$ and $E \in \mathcal{L}$. For any $\epsilon > 0$ there is an $A_\epsilon \in \mathcal{L}$ with $tv(\langle \alpha(A_\epsilon), v \rangle) < \epsilon$ and $\|\varphi_n(s) - \varphi_m(s)\| < \epsilon$ for n, m sufficiently large and $s \notin A_\epsilon$.

We have

$$|\langle\int_E(\varphi_n-\varphi_m)d\alpha,v\rangle| \le |\langle\int_{E\cap A_\epsilon}\varphi_n d\alpha,v\rangle| + |\langle\int_{E\cap A_\epsilon}\varphi_m d\alpha,v\rangle|$$
$$+ |\langle\int_{E\setminus A_\epsilon}(\varphi_n-\varphi_m d\alpha,v\rangle| \le 2r\cdot tv(\langle\alpha(E\cap A_\epsilon),v\rangle)$$
$$+\epsilon\cdot tv(\langle\alpha(E\setminus A_\epsilon),v\rangle) \le \epsilon(2r+4\|v\|),$$

which implies the $\sigma(V',V)$-convergence of $\{\int_E f_n d\alpha\}_{n\in\mathbb{N}}$ in V'.

Lemma A2: If $\{f_n\}_{n\in\mathbb{N}}$ is a nondecreasing sequence of α^*-integrable α-essentially bounded real functions converging α-almost everywhere to a α^*-integrable α-essentially bounded real function f, then for any $E\in\mathcal{L}$ the sequence $\{\int_E f_n d\alpha\}_{n\in\mathbb{N}}$ is nondecreasing and converges to $\int_E f d\alpha$ in the $\sigma(V',V)$-topology.

Proof: The proof will be given in two parts.

(1) We show that the α^*-integral of a nonnegative α-essentially bounded real function g is positive in V'. Let ess $\sup(g) =: r$. Then g is the α-almost everywhere limit of a sequence $\{g_n\}_{n\in\mathbb{N}}$ of α-simple functions which is uniformly bounded by r. Moreover, the g_n can be assumed to be nonnegative. Since $\{g_n\}_{n\in\mathbb{N}}$ converges to g $\langle\alpha,v\rangle$-almost uniformly for any $v\in V$, the same arguments used in the proof of Theorem A1 (iii) show that $\{\int_E g_n d\alpha\}_{n\in\mathbb{N}}$ converges to $\int_E g d\alpha$ in the $\sigma(V',V)$-topology for any $E\in\mathcal{L}$. Now $\int_E g_n d\alpha \ge 0$, $n\in\mathbb{N}$. Since the positive cone in V' is $\sigma(V',V)$-closed, $\int_E g d\alpha \ge 0$.

(2) To show the statement of the theorem, we use that the assumptions on $\{f_n\}_{n\in\mathbb{N}}$ and f imply the existence of a common finite α-essential bound, say $r<\infty$, for f and f_n, $n\in\mathbb{N}$. Since $\{f_n\}_{n\in\mathbb{N}}$ converges to f $\langle\alpha,v\rangle$-almost uniformly for any $v\in V$, the arguments used in the proof of Theorem A1 (iii) can be applied again to give for any $\epsilon>0$ and n sufficiently large

$$|\langle\int_E(f_n-f)d\alpha,v\rangle| \le \epsilon(2r+4\|v\|), \quad E\in\mathcal{L}.$$

which completes the proof.

Theorem A2: If α is an observable of the state space (V,B) and if f is a real α-measurable α-essentially bounded function, then we have for any $v\in V$.

$$\langle\int_E f d\alpha,v\rangle = \int_E f d\langle\alpha,v\rangle.$$

Proof: For any $v\in V$ the real function f is the $\langle\alpha,v\rangle$-almost everywhere limit of $\langle\alpha,v\rangle$-simple functions, and, because f is essentially bounded, we can assume it to be the $\langle\alpha,v\rangle$-almost uniform limit of a uniformly bounded sequence of simple functions. One checks easily that $\int_E |f_n-f_m|\,d\,tv(\langle\alpha,v\rangle)\to 0$ for $n,m\to\infty$. Hence, f is $\langle\alpha,v\rangle$-integrable (e.g., III.2.17 in Dunford and Schwartz[5]). The stated equation holds true if f is an α-simple function. Hence, if f is not α-simple, both sides of the equation can be defined by the limit of the same Cauchy sequence of real numbers. This concludes the proof.

We now consider V-valued functions. For a given observable α of the state space (V,B), an α-simple V-valued function on $\mathbb{R}$ is represented by

$$\varphi=\sum_{i=1}^n v_i\chi_{E_i}, \quad v_i\in V, \quad E_i\in\mathcal{L}.$$

We define a linear operator in V by

$$\int_E\varphi d\langle\alpha,\circ\rangle := \sum_{i=1}^n v_i\langle\alpha(E_i\cap E),\circ\rangle, \quad E\in\mathcal{L},$$

which we call (α)-integral of φ over E. We omit the simple proofs of the following statements.

(1) The (α)-integral is independent of the particular representation of the α-simple function φ.

(2) The (α)-integral is a linear mapping from the linear space of α-simple V-valued functions to $L(V,V)$ for any $E\in\mathcal{L}$, where $L(V,V)$ denotes the linear space of linear operators in V.

(3) The (α)-integral is a $L(V,V)$-valued σ-additive set function on $(\mathbb{R},\mathcal{L})$ with respect to the strong operator topology for any α-simple V-valued function.

(4) For any α-simple V-valued function φ with $\|\varphi(s)\| \le r$, $s\in E$, $r\in\mathbb{R}$, r suitable, we have

$$\|\int_E\varphi d\langle\alpha,v\rangle\| \le r\cdot tv(\langle\alpha(E),v\rangle) \le 4r\|v\|, \quad E\in\mathcal{L}, \quad v\in V.$$

Definition A3: Given an observable α of the state space (V,B), then a V-valued function f on $\mathbb{R}$ is said to be (α)-integrable if there is a sequence $\{\varphi_n\}_{n\in\mathbb{N}}$ of α-simple V-valued functions satisfying the following:

(i) $\{\varphi_n\}_{n\in\mathbb{N}}$ converges to f α-almost everywhere with respect to the norm in V.

(ii) For any $v\in V$ and for any choice of $\epsilon>0$ there is a $\delta>0$ such that, for any $A_\delta\in\mathcal{L}$ with $tv(\langle\alpha(A_\delta),v\rangle)<\delta$ and for any $n\in\mathbb{N}$, $\|\int_{A_\delta}\varphi_n d\langle\alpha,v\rangle\|<\epsilon$.

Lemma A3: If the V-valued function f is (α)-integrable, then the strong operator limit s-$\lim_{n\to\infty}\int_E\varphi_n d\langle\alpha,\circ\rangle$ exists and is independent of the particular sequence $\{\varphi_n\}_{n\in\mathbb{N}}$ satisfying (i) and (ii) of Definition A3.

Proof: Since $\{\varphi_n\}_{n\in\mathbb{N}}$ converges for any $v\in V$ to f $\langle\alpha,v\rangle$-almost everywhere and since $\langle\alpha,v\rangle$ is a finite measure on $\mathcal{L}$, $\{\varphi_n\}_n$ converges $\langle\alpha,v\rangle$-almost uniformly to f. Combining this with (ii) of Definition A3, we conclude that for any $\epsilon>0$ there are a $\delta>0$ and an $A_\delta\in\mathcal{L}$ with $tv(\langle\alpha(A_\delta),v\rangle)<\delta$ such that $\|\int_{A_\delta}\varphi_n d\langle\alpha,v\rangle\|<\epsilon$ and $\|\varphi_n(s)-\varphi_m(s)\|<\epsilon$ for $s\notin A_\delta$ and n,m sufficiently large. Hence, for any E

$$\|\int_E(\varphi_n-\varphi_m)d\langle\alpha,v\rangle\|$$
$$\le \|\int_{E\cap A_\delta}\varphi_n d\langle\alpha,v\rangle\| + \|\int_{E\cap A_\delta}\varphi_m d\langle\alpha,v\rangle\|$$
$$+\|\int_{E\setminus A_\delta}(\varphi_n-\varphi_m)d\langle\alpha,v\rangle\| \le 2\epsilon+4\epsilon\|v\|.$$

Thus the first assertion is proven. To prove the second, assume $\{f_n\}_{n\in\mathbb{N}}$ and $\{g_n\}_{n\in\mathbb{N}}$ satisfy (i) and (ii) of Definition A3. We define $\{h_n\}_{n\in\mathbb{N}}$ by $h_n(s):=0$ for the exceptional points $s\in\mathbb{R}$ where $\{f_n\}_{n\in\mathbb{N}}$ or $\{g_n\}_{n\in\mathbb{N}}$ fail to converge to f and by $h_n(s):=f_n(s)-g_n(s)$, $s\in\mathbb{R}$, otherwise. By the same arguments as above there are for any $\epsilon>0$ a $\delta>0$ and an $A_\delta\in\mathcal{L}$ such that for any $E\in\mathcal{L}$

$$\|\int_E(f_n-g_n)d\langle\alpha,v\rangle\| \le \|\int_{E\cap A_\delta}f_n d\langle\alpha,v\rangle\| + \|\int_{E\cap A_\delta}g_n d\langle\alpha,v\rangle\|$$
$$+\|\int_{E\,A_\delta}h_n d\langle\alpha,v\rangle\| \le 2\epsilon+4\epsilon\|v\|.$$

This completes the proof.

Definition A4: If f is (α)-integrable, we write for any

$E \in \mathcal{L}$

$$\int_E f d\langle \alpha, \cdot \rangle := \text{s-}\lim_{n\to\infty} \int_E \varphi_n d\langle \alpha, \cdot \rangle$$

and call it the (α)-integral of f over E.

Theorem A3: (i) If f is a V-valued α-measurable α-essentially bounded function, then f is (α)-integrable.

(ii) The (α)-integral is a linear mapping from the linear space of V-valued α-measurable α-essentially bounded functions into $L(V,V)$ for any $E \in \mathcal{L}$.

(iii) For any V-valued α-measurable α-essentially bounded function f the (α)-integral of f is a σ-additive $L(V,V)$-valued set function on $(\mathbb{R}, \mathcal{L})$ with respect to the strong operator topology.

Proof: Let ess sup$(f) =: r$ and $\{\varphi_n\}_{n\in\mathbb{N}}$ denote a sequence of α-simple functions uniformly bounded by r and converging to f α-almost everywhere. For $\epsilon > 0$ take $\delta := \epsilon/r$. Requirement (ii) of Definition A3 is then a direct consequence of property (4) of the (α)-integral for α-simple functions. Hence, f is (α)-integrable. The easy proof of statement (ii) of the theorem will be omitted. Statement (iii) is proven, if we show $\int_{(\cdot)} f d\langle \alpha, v \rangle$ to be σ-additive for any fixed $v \in V$ with respect to the norm in V. Since, for any $E \in \mathcal{L}$, we have $\int_E f d\langle \alpha, v \rangle = \lim_{n\to\infty} \int_E \varphi_n d\langle \alpha, v \rangle$ with respect to the norm in V, and, for any $n \in \mathbb{N}$, $\int_{(\cdot)} \varphi_n d\langle \alpha, v \rangle$ is a vector valued measure in the sense of Dunford and Schwartz, a generalization of the Nikodym theorem proven in Dunford and Schwartz,[5] IV.10.6, applies to show that $\int_{(\cdot)} f d\langle \alpha, v \rangle$ is a vector valued measure. This completes the proof.

Corollary: If f is a V-valued α-measurable α-essentially bounded function, we have for any $v \in V$, $a \in V'$, and $E \in \mathcal{L}$

(i) $\langle a, \int_E f d\langle \alpha, v \rangle \rangle = \int_E \langle a, f \rangle d\langle \alpha, v \rangle$,

(ii) $\langle a, \int_E f d\langle \alpha, v \rangle \rangle = \langle \int_E \langle a, f \rangle d\alpha, v \rangle$,

where $\langle a, f \rangle$ denotes the real function $s \mapsto \langle a, f(s) \rangle$, $s \in \mathbb{R}$.

Proof: Equation (i) holds true for α-simple functions. Considering a sequence of α-simple functions which is uniformly bounded and converges to f α-almost everywhere, it is easy to see that $\langle a, f \rangle$ is $\langle \alpha, v \rangle$-integrable. Hence both sides of (i) are defined and, obviously, equal.

From Theorem A2 we conclude that $\langle a, f \rangle$ is also α^*-integrable and $\langle \int_E \langle a, f \rangle d\alpha, v \rangle = \int_E \langle a, f \rangle d\langle \alpha, v \rangle$. Hence (ii) follows directly from (i).

[1] E. B. Davies and J. T. Lewis, Commun. Math. Phys. **17**, 239–60 (1970).
[2] C. M. Edwards, Commun. Math. Phys. **16**, 207–30 (1970).
[3] G. Ludwig, *Deutung des Begriffes "physikalische Theorie" und axiomatische Grundlegung der Hilbertraumstruktur...*, Lecture Notes in Physics 4 (Springer-Verlag, Berlin, 1970).
[4] E. B. Davies, J. Funct. Anal. **6**, 318–46 (1970).
[5] N. Dunford and J. T. Schwartz, *Linear Operators, Part I* (Interscience, New York, 1958).
[6] G. Lüders, Ann. Physik (6) **8**, 322–28 (1951).
[7] S. Gudder and J. P. Marchand, J. Math. Phys. **13**, 799–806 (1972).
[8] M. D. Srinivas, J. Math. Phys. **16**, 1672–85 (1975).
[9] R. G. Bartle, Studia Math. **15**, 337–52 (1956).

Highest weights of semisimple Lie algebras

W. Laskar*

Département de Mathématiques, Université de Montréal, Montréal, Québec, Canada
(Received 8 September 1976)

The nine well-known semisimple Lie algebras are partitioned in two classes: $W_{lpc=1}$ (all roots have the same length) and $W_{lzc\neq 1}$ (the roots have two different lengths of ratio equal to $c^{1/2}$). For each of these two classes a general expression is given for few elements of interest as the highest weight vector (h.w.v.) L and its power $\delta(L)$, the eigenvalues of the second order Casimir operator, the width of a weight diagram, the dimensions and the matrix elements of irreducible representations of the algebras. In the Appendix are given two examples of application of this paper.

INTRODUCTION

This paper is concerned with semisimple Lie algebras defined over an algebraically closed field of characteristic zero only (in brief, s.L.a.), i.e., with the type of algebras widely used by physicists. Calculations of highest weight vectors in particular cases[4,11-13] have of course been done already. However, here the use of a general procedure yields general formulas which give a very simple proof that no other s.L.a. than the well-known ones do exist.

To make the paper relatively self-contained and to define notations, we first recall the usual definitions of roots of an algebra, the Dynkin diagram and the highest weight vector (in brief, h.w.v.) of a given representation of that algebra.[1-14]

In the second part the calculation of the h.w.v. is performed firstly when all the roots have the same length and secondly when the roots have two different lengths of ratio equal to $\sqrt{c}$; these two cases correspond respectively to the two classes of s.L.a. $W_{lpc=1}$ and $W_{lzc\neq 1}(c=2 \text{ or } 3)$.

The third part is devoted to the interpretation of the results obtained in the second part; in a first step[15] it is very simply shown that no other semisimple Lie algebras (defined over an algebraically closed field of characteristic zero) than the ones already known do exist: the four series A_l, B_l, C_l, D_l and the five "exceptional" Lie algebras $\{E_6, E_7, E_8, F_4, G_2\}$ that we reclassify according to our scheme as

$$W_{lpc=1} = \{A_l, D_l, E_l \text{ with } l=6,7,8 \text{ only}\},$$

$$W_{lzc\neq 1} = \{B_l, C_l, F_4, G_2\}.$$

In a second step we calculate and tabulate the power $\delta(L)$ of the highest weight vector L and link it to $R = \frac{1}{2}\sum_{\mu>0}\mu$; hence the eigenvalues of the Casimir operator and the width of a weight diagram can be deduced.

In a third step the results so obtained are used to build up the matrices of representations for the two classes of algebras.

In the appendix two examples are briefly studied as applications of this paper.

I. ROOTS, DYNKIN DIAGRAM AND HIGHEST WEIGHT

The following fundamental facts are well known:

1. If $\Sigma = \{\alpha_1, \ldots, \alpha_i, \ldots, \alpha_j, \ldots, \alpha_l\}$ is an irreducible fundamental system of simple roots we have

(i) $\alpha_1, \ldots, \alpha_l$ are linearly independent;

(ii) $$2(\alpha_i, \alpha_j)/(\alpha_i, \alpha_i) = -m, \quad 2(\alpha_i, \alpha_j)/(\alpha_j, \alpha_j) = -c \quad (m, c \in Z > 0); \tag{1}$$

(iii) Σ is not decomposable into two mutually orthogonal subsets. Consequently,

$$[2(\alpha_i, \alpha_j)]^2/(\alpha_i, \alpha_i)(\alpha_j, \alpha_j) = 4\cos^2\theta = mc \leq 4 \tag{2}$$

and for $m=1$ one only gets:

$c=0$, $(\theta=90°)$; $c=1$, $(\theta=120°)$
0 line (i.e., no connection); 1 line

$c=2$, $(\theta=135°)$; $c=3$, $(\theta=150°)$;
2 lines; 3 lines

$$c=4, \begin{cases} \theta=0, & \alpha_j=\alpha_i \\ \theta=\pi, & \alpha_j=-\alpha_i. \end{cases} \tag{3}$$

Also

$$\frac{2(\alpha_i, \alpha_j)}{(\alpha_j, \alpha_j)} \Big/ \frac{2(\alpha_i, \alpha_j)}{(\alpha_i, \alpha_i)} = (\alpha_i, \alpha_i)/(\alpha_j, \alpha_j) = c, \tag{4}$$

i.e., the roots have only two possible lengths. Hence

$$(\alpha_j, \alpha_j) = \lambda, \quad (\alpha_i, \alpha_i) = c\lambda, \tag{5a}$$

$$(\alpha_i, \alpha_j) = \begin{cases} -c\lambda/2 & \text{if } \alpha_i, \alpha_j \text{ are connected roots,} \\ 0 & \text{if } \alpha_i, \alpha_j \text{ are not connected roots.} \end{cases}$$

Normalizing α_j such $\lambda = 2/c$ yields the following relations:

$$(\alpha_j, \alpha_j) = \lambda = 2/c, \quad (\alpha_i, \alpha_i) = 2,$$

$$(\alpha_i, \alpha_j) = \begin{cases} -1 & \text{if } \alpha_i, \alpha_j \text{ are connected roots} \\ 0 & \text{if } \alpha_i, \alpha_j \text{ are not connected roots} \end{cases} \tag{5b}$$

2. To every given irreducible representation (denoted IR) corresponds a unique vector L (in the idempotent $\mathcal{D}$) called the highest weight vector (denoted h.w.v.) of the

given IR. From this h.w.v. L all the properties of the IR can be deduced; for instance the Weyl formula giving the dimension N is well known:

$$N=\prod_{\mu\in\Sigma_+}(L+R,\mu)/(R,\mu)=\prod_{\mu\in\Sigma_+}[(L,\mu)/(R,\mu)+1], \tag{6}$$

Σ_+ being the subset of positive root and

$$R=\tfrac{1}{2}\sum_{\mu\in\Sigma_+}\mu. \tag{7}$$

From the h.w.v. L, a set of N ordinary weight vectors $\{\lambda_1,\ldots,\lambda_r,\ldots,\lambda_N\}$ can be deduced (all distinct if there is no degeneracy) and used in turn to calculate matrices of the IR: diagonal ones,

$$(F_\mu)^r_r=(\mu,\lambda_r),\quad \mu\in\Sigma_+, \tag{8}$$

and nondiagonal ones,

$$(E_\alpha)^t_r=\pm\{(F_\alpha)^r_r+[(E_\alpha)^r_s]^2\}^{1/2}$$

where

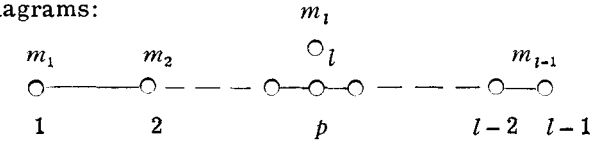
$$(E_\alpha)^r_s\neq 0 \quad \text{if } \lambda_s=\lambda_r+\alpha, \tag{9}$$

using

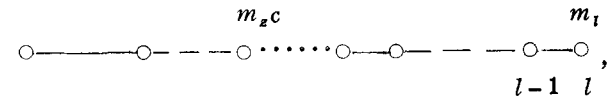
$$(E_{-\alpha})^r_s=-(E_\alpha)^s_r. \tag{10}$$

II. CALCULATIONS OF THE HIGHEST WEIGHT VECTOR

Having emphasized the importance of the h.w.v., it seems natural to calculate its expression for each of the two types of IR given by the following Dynkin diagrams:



(Type I)



(Type II)

where

$$m_i=L_{\alpha_i}=2(L,\alpha_i)/(\alpha_i,\alpha_i)$$
$$(m_i\in Z>0,\ i=1,2,\ldots,l-1,l). \tag{11}$$

Writing $L=\sum_{k=1}^{l}a_k\alpha_k$ and using (5), we get the a_k's as solution of a system of l linear equations:

$$m_i=[2/(\alpha_i,\alpha_i)][a_{i-1}(\alpha_{i-1},\alpha_i)+a_i(\alpha_i,\alpha_i)$$
$$+a_{i+1}(\alpha_{i+1},\alpha_i)+a_l(\alpha_l,\alpha_i)\,\delta_{ip}], \tag{12}$$

the last term occurring only for diagrams of Type I when $i=p$. The system (12) has been solved for each of the two types of diagrams (I) and (II). The corresponding results are given in Tables I and II for diagrams (I) and (II) respectively. If one writes $a_k=(1/\Delta)\sum_{i=1}^{l}\xi^i_k m_i$, then one gets two different expressions of Δ according to the type of diagram, say Δ_p for (I) and Δ_z for (II). These expressions will be analyzed in Sec. III to give the reason for the limitation of the number of simple Lie algebras. As a consequence of Chevalley's theorem[7,14] the classification of Dynkin diagrams is equivalent to that of simple algebraic groups over algebraically closed fields of zero characteristic.

III. ANALYSIS OF RESULTS AND APPLICATIONS

A. The two sets of algebras $W_{lpc=1}$ and $W_{lzc\neq 1}$

As the h.w.v. has been written

$$L=\sum_{k=1}^{l}a_k\alpha_k \quad \text{with} \quad a_k=\frac{1}{\Delta}\sum_{i=1}^{l}\xi^i_k m_i, \tag{13}$$

we must have $\Delta\neq 0$ and $\Delta>0$. In the case of diagrams of Type I, i.e., of $W_{lpc=1}$ we have

$$\Delta\equiv\Delta_p=p^2+(2-p)l=l+1+(p-1)(p-l+1)>0,$$

$p=l-1$ (or 1), $\Delta=l+1$ >0 for all l A_l,

$p=2$ (or $l-2$) $\Delta=p^2=4$ >0 for all l D_l,

$p=3$ (or $l-3$) $\Delta=9-l$ >0 for $l=6,7,8$ E_6,E_7,E_8,

p big, $\Delta\sim p(p-l)$ >0 for $p>l$ nonsense.

In the case of diagrams of Type II, i.e., of W_{lzc} we have

$$\Delta\equiv\Delta_z=l+1+(1-c)(l-z)z:$$

$c=1$: we come back to the previous case where all the roots have the same length with a linear diagram ($l=p-1$), i.e., to

$c=2$: $\Delta=l+1-(l-z)z=2+(z-1)(z-l+1)>0$, A_l

$z=l-1$, $\Delta=2$ >0 for all l B_l,

TABLE I. ξ_k^i for Type I (algebras $W_{lpc=1}$): $\Delta_p \equiv \Delta = p^2 + (2-p)l$, $\delta = l - p - 2$, $\Delta + p\delta = 2(l-p)$.

ξ_k^i	$1 \le i \le p-1$	$p \le i \le l-1$	l
1			
⋮	$(\Delta + i\delta)k$		
k		$2k(l-i)$	$k(l-p)$
⋮	$(\Delta + k\delta)i$		
$p-1$			
p			
⋮		$(l-i)[p^2 + (2-p)k]$	
k	$(l-k)2i$		$(l-k)p$
⋮		$[p^2 + (2-p)i](l-k)$	
$l-1$			
l	$(l-p)i$	$p(l-i)$	l

TABLE II. ξ_k^i for Type II (algebras W_{lzc}): $\Delta_z(j) = l + 1 - j + (1-c)(l-z)(z-j)$, $j = i, k, z$ or 0.

ξ_k^i	$1 \le i \le z-1$	$z \le i \le l$
1		
⋮	$k\Delta_z(i)$	
k		$k(l+1-i)$
⋮	$i\Delta_z(k)$	
$z-1$		
z	$i(l+1-z)$	$z(l+1-i)$
$z+1$		
⋮		$(l+1-i)[k + (1-c)z(k-z-1)]$
k	$ic(l+1-k)$	
⋮		$(l+1-k)[i + (1-c)z(i-z-1)]$
l		

$z = 1$, $\Delta = 2 > 0$ for all l C_l,

$z = 2$, $\Delta = 5 - l > 0$ for $l = 4$ F_4,

$c = 3$: $\Delta = l + 1 - 2z(l-z)$

$z = 1$, $\Delta = 3 - l > 0$ for $l = 2$ G_2

$c > 1$: z big $\Delta \sim z(z-l) > 0$ for $z > l$ nonsense.

When it is written, for instance, $9 - l > 0$, of course, one can take $l = 5$ (or 4) which gives D_5 (or A_4), already seen; similarly for $5 - l > 0$ $l = 3$ gives B_3 already seen.

As all other diagrams lead to a null h.w.v., one is left with the only nine s.L.a., already known and widely used by physicists; these nine s.L.a. can be classified in two sets:

$$W_{lpc=1} = \{A_l, D_l, E_l \text{ with } l = 6, 7, 8\}$$

$$W_{lzc\neq 1} = \{B_l, C_l, F_4, G_2\}.$$

B. Power of weight vector

(Freudenthal[10] and Jacobson[11] use equivalently the word "level".) By definition the power $\delta(\lambda_r)$ of weight vector $\lambda_r = \sum_{k=1}^{l} \lambda_r^k \alpha_k$ is

$$\delta(\lambda_r) = \sum_{k=1}^{l} \lambda_r^k . \tag{14}$$

1. Power of the h.w.v.

The power $\delta(L)$ of the h.w.v. $L = \sum_{k=1}^{l} a_k \alpha_k$ is

$$\delta(L) = \sum_{k=1}^{l} a_k = \frac{1}{\Delta} \sum_{k=1}^{l} \sum_{i=1}^{l} \xi_k^i m_i$$

$$= \frac{1}{\Delta} \sum_{i=1}^{l} \sum_{k=1}^{l} \xi_k^i \; m_i . \tag{15}$$

Let us write $\sum_{k=1}^{l} \xi_k^i = \Lambda^i$ so that in general

$$\delta(L) = \frac{1}{\Delta} \sum_{i=1}^{l} \Lambda^i m_i . \tag{16}$$

For $W_{lpc=1}$ the calculation of $\delta(L)$ implies three steps (and, of course, $\Delta = \Delta_p$): $1 \le i \le p-1$:

$$\Lambda^i = (i/2)[2(l+1)(l-p) + (p-i)\Delta_p]. \tag{17}$$

$p \le i \le l-1$:

$$\Lambda^i = \lfloor (l-i)/2 \rfloor [2(l+1)p + (i-p)\Delta]. \tag{18}$$

$i = l$:

$$\Lambda^i = (l/2)[2(l+1) - \Delta]. \tag{19}$$

Hence we get for the power of the h.w.v. of $W_{lpc=1}$ algebras:

$$\delta(L_p) = \frac{1}{\Delta_p} \left\{ \sum_{i=1}^{p-1} \frac{i}{2}[2(l+1)(l-p) + (p-i)\Delta_p] m_i \right.$$
$$+ \sum_{i=p}^{l-1} \frac{l-i}{2}[2(l+1)p + (i-p)\Delta_p] m_i$$
$$\left. + \frac{l}{2}[2(l+1) - \Delta_p] m_l \right\}. \tag{20a}$$

Specializing p to $l-1, 2, 3$ we get Λ^i [hence $\delta(L_p)$] for A_l, D_l, E_l $(l = 6, 7, 8)$ respectively; the results are given in Table III.

It is remarkable that due to the symmetry in i and k of Table I the h.w.v. R of the IR given by the Dynkin

TABLE III. $\delta(L_p)=(1/\Delta_p)\sum_{i=1}^{l}\Lambda_p^i m_i$ for $W_{lpc=1}$, $\Delta_p=p^2+(2-p)l$ $=l+1+(1-p)(l-1-p)$.

			$1\le i\le p-1$	$p\le i\le l-1$	$i=l$
p	W_{lp_1}	Δ_p	$\frac{\Lambda_p^i}{\Delta_p}=\frac{i}{2\Delta_p}\{2(l+1)(l-p)+(p-i)\Delta_p\}$	$\frac{\Lambda_p^i}{\Delta_p}=\frac{l-i}{2\Delta_p}[2(l+1)p+(i-p)\Delta_p\}$	$\frac{\Lambda_p^l}{\Delta_p}=\frac{l}{2\Delta_p}[2(l+1)-\Delta_p]$
$l-1$	A_l	$l+1$	$\frac{i(l+1-i)}{2}$	$\frac{(l-i)(l+1-i)}{2}$	$\frac{l}{2}$
2	D_l	4	$\frac{i(l^2-l+2-2i)}{4}$	$\frac{(l-i)(l+i-1)}{2}$	$\frac{l(l-1)}{4}$
3	E_l	$9-l$ $(l=6,7,8)$	$\frac{i}{2}\left\{\frac{2l^2-7l+21-i(9-l)}{9-l}\right\}\Rightarrow\begin{cases}l=6;\ \frac{i}{2}(17-i)\\ l=7;\ \frac{i}{2}(35-i)\\ l=8;\ \frac{i}{2}(93-i)\end{cases}$	$\frac{l-i}{2}\ \frac{9l-21+i(9-l)}{9-l}\Rightarrow\begin{cases}l=6;\ \frac{6-i}{2}(11+i)\\ l=7;\ \frac{7-i}{2}(21+i)\\ l=8;\ \frac{8-i}{2}(51+i)\end{cases}$	$\frac{l}{2}\ \frac{3l-7}{9-l}\Rightarrow\begin{cases}l=6;\ 11\\ l=7;\ \frac{49}{2}\\ l=8;\ 68\end{cases}$

diagram for which $m_i=1$ for all $i=1,2,\ldots,l$ have the same coefficients as $\delta(L_p)$, i.e.,

$$R=\frac{1}{\Delta_p}\left\{\sum_{k=1}^{p-1}\frac{k}{2}[2(l+1)(l-p)+(p-k)\Delta_p]\alpha_k+\sum_{k=p}^{l-1}\frac{l-k}{2}[2(l+1)p+(k-p)\Delta_p]\alpha_k+\frac{l}{2}[2(l+1)-\Delta_p]\alpha_l\right\}. \tag{20b}$$

It will be seen later (Theorem I, Sec. 3B2) that R is also the half sum of the positive roots.

For W_{lzc} the calculation of $\delta(L)$ implies only two steps (and, of course, $\Delta=\Delta_z$): $1\le i\le z-1$:

$$\Lambda^i=(i/2)\{c(l+1)(l-z)+(z+1-i)\Delta_z\}. \tag{21}$$

$z\le i\le l$:

$$\Lambda^i=\lfloor(l+1-i)/2\rfloor\{(l+1)(z+1)+(i-z-1)\Delta_z\}. \tag{22}$$

Hence we get for the power of the h.w.v. of W_{lzc} algebras

$$\delta(L_z)=\frac{1}{\Delta_z}\left\{\sum_{i=1}^{z-1}\frac{i}{2}[c(l+1)(l-z)+(z+1-i)\Delta_z]m_i+\sum_{i=1}^{l}\left(\frac{l+1-i}{2}\right)[(l+1)(z+1)+(i-z-1)\Delta_z]m_i\right\}. \tag{23a}$$

Due to the properties of Table II the h.w.v. R of the IR given by the Dynkin diagram for which $m_i=1$ for all $i=1,2,\ldots,l$ will be

$$R=\frac{1}{\Delta_z}\left\{\sum_{k=1}^{z-1}\frac{k}{2}[c(l+1)(l-z)+(z+1-k)\Delta_z]\alpha_{l+1-k}+\sum_{i=z}^{l}\left(\frac{l+1-k}{2}\right)[(l+1)(z+1)+(k-z-1)\Delta_z]\alpha_{l+1-k}\right\}. \tag{23b}$$

The connection of R and $\delta(L_z)$ is so established; that R is the half sum of the positive roots will be seen in Theorem I as before. The formulas obtained for A_l from $W_{lzc=1}$ as well as from $W_{lpc=1}$ are evidently the same for $p=z=l-1$.

Now for $c=2$ we get the Λ^i's for B_l when $z=l-1$, for C_l when $z=1$, for F_4 when $z=2$; for $c=3$ we get the Λ^i's for G_2 when $z=1$ and $l=2$. The results are given in Table IV.

It is worthwhile writing the formulas for $l=2$ and $z=1$ considering the frequent use of algebras of order 2. In that case, we get $\Delta_z=4-c$, and Table II gives $\xi_1^1=2$, $\xi_1^2=1$, $\xi_2^1=c$, $\xi_2^2=2$; hence for the h.w.v.

$$L_{2,c}=\frac{1}{4-c}\{(2m_1+m_2)\alpha_1+(cm_1+2m_2)\alpha_2\} \tag{24}$$

and its power

$$\delta(L_{2,c})=\frac{1}{4-c}\{(2+c)m_1+3m_2\} \tag{25}$$

which checks with Table IV.

These formulas can be used for A_2 $[c=1,\ \delta(L_{2,1})=m_1+m_2]$, for B_2 or for C_2 $[c=2,\delta(L_{2,2})=2m_1+\frac{3}{2}m_2]$, and for G_2 $[c=3,\delta(L_{2,3})=5m_1+3m_2]$.

The most important fact which comes out from Tables III and IV is that $\delta(L)$ is either integer or half-integer so that $2\delta(L)+1=T$ is always an integer either odd or even respectively. As we shall see below, T is the number of layers of the weight system constituted by all the weight vectors; the dimension N of the representation is equal to the cardinal of the set of weight vectors denoted by $\{\text{w.v.}\}=\{\lambda_1=L,\lambda_2,\ldots,\lambda_N\}$. Ordinary weight vectors are obtained by subtracting simple roots one by one from the h.w.v. L subject to rule (I):

If α_k is a simple root and $\lambda_s\in\{\text{w.v.}\}$ then $\lambda_r=\lambda_s-\alpha_k\in\{\text{w.v.}\}$ if and only if the integer $Q(\lambda_s,\alpha_k)$ determined by the two conditions

$$\lambda_s+Q(\lambda_s,\alpha_k)\alpha_k\in\{\text{w.v.}\},\qquad \lambda_s+(Q+1)\alpha_k\notin\{\text{w.v.}\} \tag{26a}$$

is such that

$$\frac{2(\lambda_s,\alpha_k)}{(\alpha_k,\alpha_k)}+Q(\lambda_s,\alpha_k)>0. \tag{26b}$$

TABLE IV. $\delta(L_z) = (1/\Delta_z)\sum_{i=1}^{l}\Lambda_z^i m_i$ for W_{lzc}, $\Delta_z = l+1+(1-c)(l-z)z$.

				$1 \le i \le z-1$	$z \le i \le l$
c	z	W_{lzc}	Δ_z	$\frac{\Lambda_z^i}{\Delta_z} = \frac{i}{2\Delta_z}\{(z+1-i)\Delta_z + c(l+1)(l-z)\}$	$\frac{\Lambda_z^i}{\Delta_z} = \frac{l+1-i}{2\Delta_2}\{(i-z-1)\Delta_z + (l+1)(z+1)\}$
1	$l-1$	A_l	$l+1$	$\frac{i}{2}(l+1-i)$	$\left(\frac{l+1-i}{2}\right)i$
2	$l-1$	B_l	2	$\frac{i}{2}(2l+1-i)$	$\left(\frac{l+1-i}{2}\right)\left\{i+\frac{l(l-1)}{2}\right\}$
	1	C_l	2	$\frac{i}{2}(l^2+1-i)$	$\frac{l+1-i}{2}(l-1+i)$
	2	$F_{l=4}$	$5-l=1$	$\frac{i}{2(5-l)}\{(3-i)(5-l)+2(l+1)(l-2)\}$ only valid for $l=4$ and $i=1$	$\frac{l+1-i}{2(5-l)}\{(i-3)(5-l)+(l+1)3\}$ which gives for F_4
				i.e., $\frac{\Lambda_z^1}{\Delta_z}(F_4) = 11$	$\frac{\Lambda_z^i}{\Delta_z}(F_4) = \frac{5-i}{2}(i+12)$, $i=2,3,4$
3	1	$G_{l=2}$	$3-l=1$	$\frac{i}{2(3-l)}\{(2-i)(3-l)+3(l+1)(l-1)\}$ gives for $l=2$ and $i=1$	$\frac{l+1-i}{2(3-l)}\{(i-2)(3-l)+(l+1)2\}$ which gives for G_2
				$\frac{\Lambda_z^1}{\Delta_z}(G_2) = 5$	$\frac{\Lambda_z^i}{\Delta_z}(G_2) = \frac{3-i}{2}(i+4)$, $i=1,2$

N.B.: The formulas obtained for A_l from $W_{lpc=1}$ as well as from $W_{lzc=1}$ are evidently the same for $p=z=l-1$.

One can define the vector

$$S_r = \sum_{j=1}^{l} i_r^j \alpha_j, \tag{27}$$

where i_r^j are l positive or null integers ($j=1,2,\ldots,l$) such that if

$$\lambda_r = \lambda_1 - S_r \in \{\text{w.v.}\}, \tag{28}$$

then

$$\delta(\lambda_r) = \delta(\lambda_1) - \delta(S_r), \tag{29}$$

i.e., the power of λ_r differs from the power of $\lambda_1 = L$ by the integer

$$\delta(S_r) = \sum_{j=1}^{l} i_r^j = i_r^1 + \cdots + i_r^l = r-1, \tag{30}$$

which is the number of simple roots subtracted from λ_1 to give λ_n.

In other words for any $\lambda_r \in \{\text{w.v.}\}$, $\delta(L)$ and $\delta(\lambda_r)$ are either both integers or both half-integers so that for a given representation all the powers of the weight system are of the same nature (corresponding to Wigner's integer or half-integer representations).

Now Eq. (30) might have many independent solutions, say q_r solutions satisfying conditions (26a), (26b); in that case the q_r ordinary weight vectors (in brief o.w.v.) $\lambda_r^{(1)}, \ldots, \lambda_r^{(q_r)}$ form the rth layer of o.w.v. all with the same power $\delta(\lambda_r) = \delta(\lambda_1) - (r-1)$. The rth layer is said to be power degenerate of order q_r. In particular for the first layer corresponding to the unique h.w.v. $\lambda_1 = L$ one has $r=1$, $S_1 = 0$, $\delta S_1 = 0$, $q_r = 1$ and the first layer is never degenerate. If $\delta(\lambda_1)$ is an integer, after $(m_0 - 1)$ subtracting steps such that

$$\delta(\lambda_{m_0}) = \delta(\lambda_1) - (m_0 - 1) = 0$$

we have an m_0th layer of w.v. with power equal to zero; here $m_0 = \delta(\lambda_1) + 1$. If $\delta(\lambda_1)$ is a half-integer, after $(m_{1/2} - 1)$ subtracting steps such that

$$\delta(\lambda_{m_{1/2}}) = \delta(\lambda_1) - (m_{1/2} - 1) = \tfrac{1}{2},$$

we have a $m_{1/2}$th layer of w.v. with power equal to $\frac{1}{2}$; here $m_{1/2} = \delta(\lambda_1) + \frac{1}{2}$. In both cases due to the symmetry of the process the total number T of layers (called the height of the w.v. system) is then

$$T = 2\delta(\lambda_1) + 1.$$

As we shall see the power degeneracy cannot diminish as the number of subtracting steps grows (up to $m-1$ steps) and consequently the degeneracy is maximum either for the m_0th layer if $\delta(\lambda_1)$ is integer, say q_{m_0}, or for the $m_{1/2}$th layer if $\delta(\lambda_1)$ is half-integer, say $q_{m_{1/2}}$. This maximum power degeneracy q_m is called the width of the w.v. system. So that finally we have for the dimension N of the representation (counting each w.v. with its multiplicity)

if $\delta(\lambda_1)$ is integer, $N = 2(q_1 + \cdots + q_i + \cdots q_{m_0-1}) + q_{m_0}$,

if $\delta(\lambda_1)$ is half-integer, $N = 2(q_1 + \cdots + q_i + \cdots + q_{m_{1/2}})$.

(with $q_1 = 1$ and $q_{i+1} \ge q_i$). In both cases we have $T = 2\delta(\lambda_1) + 1 \le N$, the equal sign corresponding to the case of no degeneracy.

2. Effective determination of o.w.v.

The first layer being occupied by the unique h.w.v. $\lambda_1 = L$, let us look for the w.v.'s of the second layer.

According to rule (I), since $\lambda_1 \in \{\text{w.f.}\}$, $\lambda_1 + \alpha_i \notin \{\text{w.v.}\}$, we have $Q(\lambda_1, \alpha_i) = 0$ for $i = 1, 2, \ldots, l$. As $2(\lambda_1, \alpha_i)/(\alpha_i, \alpha_i) = m_i$, for $\lambda_1 - \alpha_i$ to be a w.v. we have the condition

$$2(\lambda_1, \alpha_i)/(\alpha_i, \alpha_i) + Q(\lambda_1, \alpha_i) = m_i > 0. \tag{31}$$

If there are q_2 values of $m_i \neq 0$, we obtain a second layer of q_2 different w.v. $\lambda_2 = \{\lambda_2^{(1)}, \ldots, \lambda_2^{(q_2)}\}$ with the same power $\delta(\lambda_2) = \delta(\lambda_1) - 1$. Similarly the w.v. of the third

layer are obtained by determining first $Q(\lambda_2^{(i)}, \alpha_j)$:

$$\lambda_2^{(i)} + \alpha_j = \lambda_1 - \alpha_i + \alpha_j \in \{\text{w.v.}\} \text{ if and only if } \alpha_i = \alpha_j,$$

$$\lambda_2 + 2\alpha_j = \lambda_1 - \alpha_i + \alpha_j + \alpha_j \notin \{\text{w.v.}\}$$

so that $Q(\lambda_2^{(i)}, \alpha_j) = \delta_{i,j}$ and the condition for $\lambda_2^{(i)} - \alpha_j$ to be a w.v. is

$$\frac{2(\lambda_2, \alpha_j)}{(\alpha_j, \alpha_j)} + Q(\lambda_2^{(i)}, \alpha_j) = \frac{2(\lambda_1 - \alpha_j, \alpha_j)}{(\alpha_j, \alpha_j)} + \delta_{i,j} > 0$$

$$= m_j - \frac{2(\alpha_j, \alpha_j)}{(\alpha_j, \alpha_j)} + \delta_{i,j} > 0$$

If $j = i$, we get

$$m_i > 1; \tag{32}$$

so that for $m_i \geq 2$, $\lambda_1 - 2\alpha_i$ is a w.v. of power $\delta(\lambda_1) - 2$. If $j \neq i$ we get in the case where α_j and α_i are not connected,

$$m_j > 0, \tag{33a}$$

[so that for $m_j \geq 1$, $\lambda_1 - \alpha_i - \alpha_j$ (with $|i-j| \geq 2$) is a w.v. with power $\delta(\lambda_1) - 2$]; in the case where α_j and α_i are connected, i.e., $j = i \pm 1$, then $(\alpha_i, \alpha_j) = -1$ and the condition

$$m_j + 2/(\alpha_j, \alpha_j) > 0 \tag{33b}$$

is fulfilled even if $m_j = 0$; then $\lambda_2 - \alpha_j = \lambda_1 - \alpha_i - \alpha_j$ with $|i-j| = 1$ is a w.v. of power $\delta(\lambda_1) - 2$. So that with each w.v. of the second layer $\lambda_2^{(i)}$ we get at least one w.v. of the third layer.

To study the rth layer, let us write now a w.v. of the $(r-1)$th layer as

$$\lambda_{r-1}^{(i)} = \lambda_1 - S_{r-1}^{(i)}, \tag{34}$$

where

$$S_{r-1}^{(i)} = \sum_{j=1}^{l} i_{r-1}^j \alpha_j \tag{35}$$

with $\delta(S_{r-1}^{(i)}) = r-2$ and $\delta(\lambda_{r-1}^{(i)}) = \delta(\lambda_1) - r + 2$; (all i_{r-1}^j are positive or nul integers; $j = 1, 2, \ldots, l$). For $\lambda_r^{(i)} = \lambda_{r-1}^{(i)} - \alpha_s$ to be a weight vector of the rth layer, we determine $Q(\lambda_{r-1}^{(i)}, \alpha_s)$

$$\lambda_{r-1}^{(i)} + Q\alpha_s \in \{\text{w.v.}\},$$

$$\lambda_{r-1}^{(i)} + (Q+1)\alpha_s \notin \{\text{w.v.}\}.$$

So that $Q = \sum_{j=1}^{l} i_{r-1}^j \delta_{j,s}$ and the condition for $\lambda_{r-1}^{(i)} - \alpha_s$ to be a w.v. is

$$\frac{2(\lambda_{r-1}^{(i)}, \alpha_s)}{(\alpha_s, \alpha_s)} + Q = \frac{2(\lambda_1 - S_{r-1}^{(i)}, \alpha_s)}{(\alpha_s, \alpha_s)} + Q > 0,$$

$$m_s - \frac{2(S_{r-1}^{(i)}, \alpha_s)}{(\alpha_s, \alpha_s)} + \sum_{j=1}^{l} i_{r-1}^j \delta_{j,s} > 0.$$

If $s \neq j$ and $|s-j| \geq 2$ for all j's such that $i_{r-1}^j \neq 0$, that is to say, if α_s is none of the α_j involved in $S_{r-1}^{(i)}$ and if α_s is not connected with any one of them, then rule (I) gives

$$m_s > 0, \text{ i.e., } m_s \geq 1. \tag{36}$$

If $s \neq j$ and α_s is connected with at least one of the α_j's involved in $S_{r-1}^{(i)}$, then for that value of j, $(\alpha_j, \alpha_s) = -1$; rule (I) is fulfilled even if $m_s = 0$. (Notice that this conclusion remains true if α_s is connected with two α_j's, or exceptionally three α_j's in D_l or in E_l. If $\alpha_s = \alpha_j$, i.e., If $\alpha_s = \alpha_j$, i.e., if α_s is a particular α_j, then rule (I) gives

$$m_j - i_{r-1}^j > 0, \text{ i.e., } m_j \geq i_{r-1}^j + 1. \tag{37}$$

In particular among the q_{r-1} solutions of Eq. (30) applied to the $(r-1)$th layer there is the maximal one $i_{r-1}^s = r - 2$ (with $i_{r-1}^j = 0$ for all other j's) and correspondingly $\lambda_1 - (r-1)\alpha_s$ will be a w.v. of the rth layer with power

$$\delta(\lambda_r) = \delta(\lambda_1) - r + 1 \text{ if } m_s \geq r - 1.$$

The conditions $m_s \geq 1$ for the second layer, $m_s \geq 2$ for the third layer, etc., $\cdots$, $m_s \geq r-1$ for the rth layer become obvious in terms of Young diagrams; also we can see that power degeneracy cannot diminish as the number of subtracting steps grows as stated previously.

Due to the action of the Weyl group the w.v. system takes a spindle shape. Within a given layer $\{\lambda_r\}$ a certain w.v. M can occur more than once as soon as $r \geq 3$; indeed we have

$$M = \lambda_r^{(i)} = \lambda_1 - \sum_{j=1}^{l} i_r^j \alpha_j = \lambda_{r-1}^{(i_1)} - \alpha_{s_1} = \lambda_{r-1}^{(i_2)} - \alpha_{s_2} = \cdots. \tag{38}$$

For example, the w.v. system of the representation $\overset{1}{\circ}\!\!-\!\!\!-\!\!\overset{1}{\circ}$ of A_2 is

$$\{\text{w.v.}\} = \{\alpha_1 + \alpha_2; \alpha_1, \alpha_2; 0, 0; -\alpha_2, -\alpha_1; -(\alpha_1 + \alpha_2)\} \tag{39}$$

and the null w.v. of the third layer is obtained in two ways from the second layer; so that the nul w.v. is degenerate and its multiplicity is two.

3. Width of a weight diagram; Freudenthal's formula. Casimir operators

In general, if M appear n_M times, then M is said to be degenerate and n_M is its multiplicity (or the dimension of the corresponding degenerate subspace of the w.v. space); it means that each w.v. such as M has to be counted n_M times to maintain the fact that the dimension N of the representation space is equal to the total number of w.v.

Freudenthal's recursion formula[10] gives the multiplicity n_M of M as

$$[(L+R, L+R) - (M+R, M+R)]n_M$$

$$= 2 \sum_{\mu > 0} \sum_{k=1}^{\infty} (M + k\mu, \mu) n_{M+k\mu}, \tag{40}$$

where R as for the Weyl's formula is given by Eq. (7). To calculate dimensions of representations by Weyl's formula [Eq. (6)], one does not need $L+R$ but R. As roots and weights are dual forms[1-14,16] with respect to the fundamental Killing quadratic form of the algebra, the power $\delta(L)$ of the h.w.v. in the weight space corresponds to R in the root space.

Theorem I: $\delta(L) = \sum_{i=1}^{l} (\Lambda^i/\Delta) m_i$ and $R = \frac{1}{2}\sum_{\mu>0} \mu^i \alpha_i$ are dual elements. The ordering of the roots is important for the use of this theorem; for W_{lp1} the order is given in Table V; for W_{lzc} one has to interchange m_i and α_{l+1-i} (B_l and C_l also since they are dual too). With these precautions R can be built up out of Tables III

TABLE V. Eigenvalues of Casimir operator for W_{lp1}. Ordering of the roots for W_{lp1}: For A_l as the coefficients of $\delta(L)$ are symmetric in i and $l+1-i$ the interchange has no effect and it is just as well not do it. (See Ref. 16, p. 27.) For D_l from an orthonormal basis $\{e_i\}$ of $\mathbb{R}^l$ all roots are defined as $\pm e_i \pm e_j$ $(i \neq j)$. As we notice that Table III gives the same coefficient for $i=1$ and for $i=l$ we define the simple roots in the following order: $\alpha_1 = e_{l-1} - e_l, \ldots, \alpha_{l-i} = e_i - e_{i+1}, \ldots, \alpha_{l-1} = e_1 - e_2$, $\alpha_l = e_{l-1} + e_l$. For E_l from an orthogonormal basis $\{e_i\}$ of $\mathbb{R}^8$ all roots are defined as $\pm e_i \pm e_j$ $(i \neq j)$, and the vectors $\frac{1}{2}\sum_{i=1}^{l}(-1)^{m(i)}e_i$, with $\sum m(i)$ = even; we define the simple roots as $\alpha_l = e_1 + e_2$, $l = 6,7,8$ only, $\alpha_1 = \frac{1}{2}(e_1 + e_8 - \sum_{i=2}^{7} e_i)$, $\alpha_2 = e_2 - e_1$, $\alpha_3 = e_3 - e_2$, $\alpha_4 = e_4 - e_3, \ldots$, $\alpha_{l-1} = e_{l-1} - e_{l-2}$. With the above ordering of the simple roots of W_{lp1}, if, using Table III, we write $\delta(L) = \sum_{k=1}^{l} b_k m_k$, then we get simply $R = \frac{1}{2}\sum_{\rho>0}\mu = b_k\alpha_k$, with $b_k = \Lambda^k/\Delta_p$.

$$
\begin{aligned}
C_p = \frac{1}{\Delta_p}\Bigg[& \sum_{k=1}^{p-1}\Bigg(\sum_{i=1}^{i=k}(\Delta_p + k\delta)i m_i + \sum_{i=k+1}^{p-1}(\Delta_p + i\delta)k m_i + \sum_{i=p}^{l-1} 2k(l-i)m_i + k(l-p)m_l + k[2(l+1)(l-p) + (p-k)\Delta_p]\Bigg) m_k \\
& + \sum_{k=p}^{l-1}\Bigg(\sum_{i=1}^{p-1}(l-k)2i m_i + \sum_{i=p}^{i=k}[p^2 + (2-p)i(l-k)]m_i + \sum_{i=k+1}^{l-1}[p^2 + (2-p)k(l-i)\, m_i + (l-k)p m_l \\
& + (l-k)[2(l+1)p + (k-p)\Delta_p]\Bigg) m_k + \Bigg(\sum_{i=1}^{p-1}(l-p)i m_i + \sum_{i=p}^{l-1} p(l-i)m_i + l m_l + l[2(l+1) - \Delta_p]\Bigg) m_l\Bigg].
\end{aligned}
$$

and IV for W_{lp1} and W_{lzc}.

Let us give two examples easy to check in no time.

For G_2 Table IV gives $\delta(L(G_2)) = 5m_1 + 3m_2$; (41a)

then Theorem I: $R(G_2) = 3\alpha_1 + 5\alpha_2$. (41b)

For F_4 Table IV gives $\delta(L(F_4)) = 11m_1 + 21m_2 + 15m_3 + 8m_4$; (41c)

then Theorem I: $R(F_4) = 8\alpha_1 + 15\alpha_2 + 21\alpha_3 + 11\alpha_4$. (41d)

Now that we have $L = \sum_{k=1}^{l} a_k\alpha_k$ (Tables I and II) and $R = \sum_{k=1}^{l} b_k\alpha_k$ (Tables III and IV) using universally adopted Racah's notation[17b] it is easy to build $K = L + R$ and, consequently, K^2, which is involved in Freudenthal's formula as well as in the second order Casimir operator whose eigenvalues are $K^2 - R^2 = L(L+2R) = C$ for the representation defined by a given Dynkin diagram.

Using Eq. (5) and properties of the Cartan matrix involved in Eq. (12), we obtain

$$
\begin{aligned}
C = K^2 - R^2 &= \sum_{k=1}^{l}(a_k + 2b_k)m_k \frac{(\alpha_k, \alpha_k)}{2} \\
&= \sum_{k=1}^{l} a_k(m_k + 2)\frac{(\alpha_k, \alpha_k)}{2}\,,
\end{aligned} \tag{42}
$$

given in Table V for W_{lp1} and in Table VI for W_{lzc}. (The trivial exercise of specialization to particular values of p, z, and c is left to the reader.)

The width of the weight diagram can be deduced easily now from Freudenthal's formula. We have seen that this width is the degeneracy n_0 of the null weight vector when $\delta(L)$ is integer and $n_{\alpha_i/2}$ of the w.v. $M = \alpha_i/2$ when $\delta(L)$ is half-integer. In the first case we get

$$
(K^2 - R^2)n_0 = 2\sum_{\mu>0}\sum_{k=1}^{\infty}(k\mu, \mu)n_{k\mu} \tag{40'a}
$$

and in the second case

$$
\begin{aligned}
&[(K^2 - R^2 - \tfrac{5}{4}(\alpha_i, \alpha_i) - 2]n_{\alpha_i/2} \\
&\quad = 2\sum_{\mu>0}\sum_{k=1}^{\infty}(\alpha_i/2 + k\mu, \mu)n_{\alpha_i/2+k\mu},
\end{aligned} \tag{40'b}
$$

where μ is a positive root and $\alpha_i/2 + k\mu$ must be a weight. In the Appendix examples of application of these formulas are given.

Of course, the use of Tables V and VI can be avoided if one use the second form of Eq. (42) that we write again as

$$
C = L(L+2R) = \sum_{k=1}^{l} a_k(m_k + 2)\left(\frac{\alpha_k, \alpha_k}{2}\right), \tag{42}
$$

where only the coefficients of the h.w.v. L given in Tables I and II are involved. Anyway the ordering of the roots is still necessary to go from $\delta(L)$ to R.

C. Matrices of IR of semisimple Lie algebras

1. On Weyl's formula and dimensions of IR of semisimple Lie algebras

Weyl's formula[1,11] gives the dimension N of an IR

TABLE VI. Eigenvalues of Casimir operator for W_{lzc}. (The order of the roots is as given in Ref. 16, Chap. V, pp. 28–29.)

$$
\begin{aligned}
C_z = \frac{1}{\Delta_z}\Bigg[& \sum_{k=1}^{z}\Bigg(\sum_{i=1}^{i=k} i\Delta_z(k)m_i + \sum_{i=k+1}^{z-1} k\Delta_z(i)m_i + \sum_{i=z}^{l} k(l+1-i)m_i + k[(l-k-z)\Delta_z + (l+1)(z+1)]\Bigg) m_k \\
& + \sum_{k=z+1}^{l-z+1}\Bigg(\sum_{i=1}^{z-1} ic(l+1-k)m_i + \sum_{i=z}^{i=k}(l+1-k)[i + (1-c)z(i-z-1)]m_i + \sum_{i=k+1}^{l}(l+1-i)[k + (1-c)z(k-z-1)]m_i \\
& + k[(l-k-z)\Delta_z + (l+1)(z+1)]\Bigg) m_k \frac{(\alpha_K, \alpha_K)}{2} \\
& + \sum_{k=l-z+2}^{l}\Bigg(\sum_{i=1}^{z-1} ic(l+1-k)m_i + \sum_{i=z}^{i=k}(l+1-k)[i + (1-c)z(i-z-1)]\Bigg) m_i + \sum_{i=k+1}^{l}(l+1-i)[k + (1-c)z(k-z-1)]m_i \\
& + (l+1-k)[(z-l+k)\Delta_z + c(l+1)(l-z)]\Bigg) m_k \frac{(\alpha_k, \alpha_k)}{2}\Bigg].
\end{aligned}
$$

as

$$N=\prod_{\mu>0}\left[\frac{(L,\mu)}{(R,\mu)}+1\right]. \tag{43}$$

This formula implies the knowledge of L and of all the positive roots μ [$R=\frac{1}{2}\sum_{\mu>0}\mu$ being deduced either directly from the μ's or from $\delta(L)$]. What follows shows that the knowledge of the positive roots is enough. As $\mu=\sum_{i=1}^{l}\mu^i\alpha_i$ ($\mu^i\in Z^+$), using Eq. (11), we have

$$(L,\mu)=\sum_{i=1}^{l}\mu^i m_i\,\frac{(\alpha_i,\alpha_i)}{2}\ .$$

As $R=\frac{1}{2}\sum_{\mu>0}\mu$ is also the highest weight of the IR corresponding to the Dynkin diagram such that all $m_i=1$ ($i=1,2,\ldots,l$), we have

$$(R,\mu)=\sum_{i=1}^{l}\mu^i\,\frac{(\alpha_i,\alpha_i)}{2}=\delta(L,\mu),$$

where $\delta(L,\mu)$ is the power (i.e., the sum of the m_i's coefficients) of (L,μ). Formula (43) becomes

$$N=\prod_{\mu>0}\left[\frac{\sum_{i=1}^{l}\mu^i m_i(\alpha_i,\alpha_i)/2}{\sum_{i=1}^{l}\mu^i(\alpha_i,\alpha_i)/2}+1\right]. \tag{44}$$

For W_{lp1} for which $(\alpha_i,\alpha_i)/2=1$ we get

$$N(W_{lp1})=\prod_{\mu>0}\left[\frac{\sum_{i=1}^{l}\mu^i m_i}{\delta(\mu)}+1\right], \tag{44a}$$

where $\sum_{i=1}^{l}\mu^i m_i$ is obtained from $\mu=\sum_{i=1}^{l}\mu^i\alpha_i$ by interchanging α_i and m_i and $\delta(\mu)=\sum_{i=1}^{l}\mu^i$ is the power of the positive root μ.

For W_{lzc} using previous notations, we have

$$N(W_{lzc})=\prod_{\mu>0}\left[\frac{\sum_{i=1}^{g}\mu^i m_i+\sum_{i=g+1}^{l}\mu^i m_i/c}{\sum_{i=1}^{g}\mu^i+\sum_{i=g+1}^{l}\mu^i/c}+1\right]. \tag{44b}$$

The number of positive roots μ of a given Lie algebra being called n_p the Coxeter index h is then $h=2n_p/l$ and the maximum power $\delta(\mu_m)$ of the positive roots is $\delta(\mu_m)=h-1$. When expressing (44a) and (44b) it is useful to give the factors of N in increasing order of $\delta(\mu)$ with

$$1\leq\delta(\mu)\leq\delta(\mu_m)=h-1,$$

where $\delta(\mu)=1$ corresponds to simple roots.

It follows from Eq. (44) that to write the dimension of the IR of a given Lie algebra corresponding to a Dynkin diagram (or to a Young diagram) only the positive roots of that algebra are needed. The families of positive roots are build up out of an orthonormal basis $\{e_i\}$ of a vector space E according to Tables VII and VIII for W_{lp1} and W_{lzc} respectively. The dimensions are then deduced according to the above method and given in Tables IX and X for W_{lp1} and W_{lzc} respectively.

At that stage it is useful to establish at least for A_l, B_l, C_l the connection between Young and Dynkin dia-

TABLE VII. Families of positive roots for W_{lp1} algebras.

p	W_{lp1}	E	o.n. basis $\{e_i\}$ of E.	n_p	$\delta(\mu_m)$	first family $e_{i+p}-e_i=\alpha_{i+1}+\cdots+\alpha_{i+p}$	Other families of positive roots
$l-1$	A_l	$\mathbb{R}^{l+1}$	$i=0,1,\ldots,l$	$\frac{l(l+1)}{2}$	l	$i=0,1,2,\ldots,l-1$ $\mu_m=e_l-e_0=\alpha_1+\alpha_2+\cdots+\alpha_l$	
2	D_l	$\mathbb{R}^l$	$i=0,1,\ldots,l-1$	$l(l-1)$	$2l-3$	$i=0,1,2,\ldots,l-2.$	second family $e_{i+p}+e_i=\alpha_1(1-\delta_{i,0})+\alpha_l+\alpha_2+\alpha_3+\cdots+\alpha_i+\alpha_{i+1}+\cdots+\alpha_{i+p}$ $+(\alpha_2+\alpha_3+\cdots+\alpha_i)(1-\delta_{i,0})(1-\delta_{i,1})$ yields $\alpha_l=e_1+e_0$ for $p=1$, $i=0$ as well as $\mu_m=e_{l-1}+e_{l-2}=\alpha_1+2\alpha_2+2\alpha_3+\cdots+2\alpha_{l-2}$ $+\alpha_{l-1}+\alpha_l$
3	E_l	$\mathbb{R}^8$	$i=1,2,\ldots,8$ $n_p=(l-1)(l-2)+(l-6)[6(l-7)+1]+2^{l-2}$				second family
	$l=6,7$ $l=8$					$i=1,2,\ldots,l-2$ $i=1,2,\ldots,l-1$	$e_{i+p}+e_i=\alpha_2(1-\delta_{i,1})+\alpha_l+\alpha_3+\alpha_4+\cdots+\alpha_i+\alpha_{i+1}$ $+\cdots+\alpha_{i+p}$ $+(\alpha_3+\alpha_4+\cdots+\alpha_i)(1-\delta_{i,1})(1-\delta_{i,2})$
							third family $\mu(j)=\frac{1}{2}\sum_{i=1}^{8}(-1)^{m(i)}e_i$ with $m(i)=0$ or 1 and $\sum' m(i)=$ even
	$l=6$		$2\times10+16$		11		$\alpha_1=\mu(1)=\frac{1}{2}(e_1+e_8-e_7-e_6-e_5-e_4-e_3-e_2)$; $\mu(i)=\alpha_1+\alpha_2+\cdots+\alpha_i,\quad 1\leq i\leq 4$ $\mu(5)=\mu(4)+\alpha_l,\ \mu(6)=\mu(5)+\alpha_3,\ \mu(7)=\mu(6)+\alpha_2,$ $\mu(8)=\mu(6)+\alpha_4+\alpha_5$ $\mu(9)=\mu(3)+\alpha_l;\ \mu(6+i)=\mu(i)+\alpha_5,\ 4\leq i\leq 7;\ \mu(14)=\mu(13)+\alpha_4$ $\mu(15)=\mu(14)+\alpha_3,\ \mu_m(16)=\mu(15)+\alpha_l$
	$l=7$		$2\times15+32+1$		17		$\mu(16+i)=\mu(i)+\begin{cases}\alpha_2+\alpha_3+\alpha_4+\alpha_5+\alpha_6 \text{ for } i=1,5,6,8,9,11,12,16\\ \alpha_3+\alpha_4+\alpha_5+\alpha_6+\alpha_l \text{ for } i=2,3,4,7,10,13,14,1\end{cases}$ $\mu_m(E_7)\equiv\mu_{33}=e_8-e_7=2\alpha_1+3\alpha_2+4\alpha_3+3\alpha_4+2\alpha_5+\alpha_6+2\alpha_7$
	$l=8$		$2\times28+64$		29		$\mu(32+i)=\mu(16+i)+\alpha_7$ $\mu(48+i)=\mu(i)+\alpha_2+2\alpha_3+2\alpha_4+2\alpha_5+2\alpha_6+\alpha_7+\alpha_l$ $1\leq i\leq 16$ $\qquad$ $1\leq i\leq 16$

TABLE VIII. Families of positive roots of W_{lzc} algebras.

c	z	W_{lzc}	E	o.n. basis $\{e_i\}$ of E	n_P	$\delta(\mu_m)$	first family	second family	third family
2	$l-1$	B_l	$\mathbb{R}^l$	$i=1,2,\ldots,l$	l^2	$2l-1$	$e_i-e_{i+p}=\alpha_{i+1}+\cdots+\alpha_{i+p-1}$ $i=1,2,\ldots,l-1$ yields $\alpha_1,\ldots,\alpha_{l-1}$	$e_i=\alpha_i+\alpha_{i+1}+\cdots+\alpha_l$ $i=1,2,\ldots,l$ yields $\alpha_l=e_l$	$e_i+e_{i+p}=\alpha_i+\alpha_{i+1}+\cdots+\alpha_{i+p-1}+2\alpha_{i+p}+\cdots+2\alpha_l$ $\mu_m=e_1+e_2=\alpha_1+2\alpha_2+\cdots+2\alpha_l$
2	1	C_l	$\mathbb{R}^l$	$i=1,2,\ldots,l$	l^2	$2l-1$	$e_{i+p}-e_i=\alpha_{i+1}+\cdots+\alpha_{i+p}$ $i=1,2,\ldots,l-1$ yields the simple roots $\alpha_2,\ldots,\alpha_l$	$2e_i=\alpha_1+2\alpha_2+\cdots+2\alpha_l$ $i=1,2,\ldots,l$ yields $\alpha_1=2e_1$ and $\mu_m=2e_l=\alpha_1+2\alpha_2+\cdots+2\alpha_l$	$e_{i+p}+e_i=\alpha_1+\alpha_2+\cdots+\alpha_i+\alpha_{i+1}+\cdots+\alpha_{i+p}+(\alpha_2+\cdots+\alpha_l)(1-\delta_{i,1})$
2	2	F_4	$\mathbb{R}^4$	$i=1,2,3,4$	24	11	$e_2-e_3=\alpha_1$ $e_3-e_4=\alpha_2$ $e_2-e_4=\alpha_1+\alpha_2$ $e_1-e_2=2\alpha_4+2\alpha_3+\alpha_2$ $e_1-e_3=2\alpha_4+2\alpha_3+\alpha_2+\alpha_1$ $e_1-e_4=2\alpha_4+2\alpha_3+2\alpha_2+\alpha_1$ fourth family $\frac{1}{2}(e_1\pm e_2\pm e_3\pm e_4)$ yields	$e_4=\alpha_3$ $e_3=\alpha_2+\alpha_3$ $e_2=\alpha_1+\alpha_2+\alpha_3$ $e_1=\alpha_1+2\alpha_2+3\alpha_3+2\alpha_4$ $\alpha_4=\frac{1}{2}(e_1-e_2-e_3-e_4)$ $\mu(i)=\alpha_4+\cdots+\alpha_{5-i}$, $1\le i\le 4$ $\mu(5)=\mu(3)+\alpha_3$; $\mu(6)=\mu(4)+\alpha_3$ $\mu(7)=\mu(6)+\alpha_2$; $\mu(8)=\mu(7)+\alpha_3$	$\mu_m=e_1+e_2=2\alpha_1+3\alpha_2+4\alpha_3+2\alpha_4$ $e_1+e_3=\alpha_1+3\alpha_2+4\alpha_3+2\alpha_4$ $e_1+e_4=\alpha_1+2\alpha_2+4\alpha_3+2\alpha_4$ $e_2+e_3=\alpha_1+2\alpha_2+2\alpha_3$ $e_2+e_4=\alpha_1+\alpha_2+2\alpha_3$ $e_3+e_4=\alpha_2+2\alpha_3$
3	1	G_2	$\mathbb{R}^3$	$i=1,2,3$	6	5	$e_i-e_{i+p}=\alpha_i+\alpha_{i+1}+\cdots+\alpha_{i+p-1}$ $i=1,2$ yields $\alpha_1,\alpha_2,\alpha_1+\alpha_2$	$e_1+e_2+e_i-3e_3$ $i=3,2,1$ yields respectively $\alpha_1+2\alpha_2;\alpha_1+3\alpha_2$; and $\mu_m=2\alpha_1+3\alpha_2$	

grams. For Young diagrams (as oppose to Dynkin diagrams) one has to say in which Lie algebra they have to be considered. Then, if λ_i is the length of the ith line, one has for A_l

$$m_i=\lambda_i-\lambda_{i+1} \quad (\text{for } i=1,2,\ldots,l);$$

however, for B_l [algebra of SO$(2l+1)$] one has

$$m_i=\lambda_i-\lambda_{i+1} \quad (\text{for } i=1,2,\ldots,l-1) \quad \text{and} \quad m_l=2\lambda_l;$$

for C_l (algebra of Sp$(2l)$) one has

$$m_i=\lambda_{l+1-i}-\lambda_{l+2-i} \quad (\text{for } i=2,\ldots,l) \quad \text{and} \quad m_1=\lambda_l.$$

with these precautions taken, Tables IX and X can be used, for instance, to help the reduction of the IR of a group w.r.t. its invariant subgroups as for the decomposition of SU(n) into representations of SO(3) and the studies of the chain SU$(2l+1)\supset$ SO$(2l+1)\supset$ SO(3) for l integer and of the chain SU$(2j+1)\supset$ Sp$(2j+1)\supset$ SO(3) for j half-integer, which are the root of the seniority concept so widely used by physicists (cf. Hamermesh,[18] Chap. 11).

An IR is called *basic* if all components m_j of the h.w.v. are zero except one, $m_i=\delta_{i,j}$ for $j=1,2,\ldots,i,\ldots,l$; such a representation is denoted $(W_l)\beta_i$ and its dimension $N(W_l)\beta_i$ is obtained by doing $m_j=\delta_{i,j}$ in Tables IX (for W_{lp1}) and X (for W_{lzc}). The results listed in Tables XI and XII have already been obtained[19] by the L dependent method; they cannot be used to compute the dimension of any general IR since the dimension formulas are very far from being linear in the m_i's; they are only an example as a test of Tables IX and X.

The basic IR of smallest dimension will be called the *elementary* IR as it corresponds to the dimension of the smallest vector space of representation and according to our coherent notation (cf. Sec. III, Type I and II) corresponds to a terminal simple root, i.e., $i=1,l$, or $l-1$, this last value being specially valid for D_l and for E_l $(=6,7,8)$.

A representation of particular interest is also the one whose dimension is equal to the number r of parameters of the associated group; such a representation will be called the regular representation and denoted RR; we have

$$r=2n_p+l=l[(2n_p/l)+1]=l(h+1)=l[\delta(\mu_m)+2].$$

In general the adjoint representation is a basic one except for the cases of: A_l for which the RR is the IR $m_1=m_l=1$, $m_i=0$ for $i=2,3,\ldots,l-1$; C_l for which the RR is the following reducible representation:

$$\mathrm{RR}(C_l)=(C_l)\beta_{l-1}\oplus(C_l)\beta_l\oplus(C_l)\beta_0,$$

where $(C_l)\beta_0$ is the scalar identity representation for which all $m_i=0$. As for E_l one has $n_p=(l-1)(l-2)+(l-6)[6(l-7)+1]+2^{l-2}$, which is not a simple function of l to handle all the basic representations are computed directly using Tables IX and X.

TABLE IX. Dimensions of IR of W_{lp1} algebras.

P	
$l-1$	$N(A_l)=\prod_{i=1}^{l}(m_i+1)\cdots\prod_{i=1}^{l+1-k}\left[\frac{m_i+m_{i+1}+\cdots+m_{i+k-1}}{k}+1\right]\cdots\left[\frac{m_1+m_2+\cdots+m_l}{l}+1\right]$.
2	$N(D_l)=\Pi_1(D_l)\Pi_2(D_l)$ corresponds to the two families of roots $\Pi_1(D_l)=\Pi(A_{l-1})$ $\Pi_2(D_l)=(m_p+1)\prod_{i=2}^{l-1}\left[\frac{m_1+m_2+m_3+\cdots+m_l}{i}+1\right]\prod_{i=3}^{l}\left[\frac{m_1+m_1+m_2+\cdots+m_{i-1}}{i}+1\right]$ $\times\prod_{2i+k=5}^{2l-3}\left[\frac{m_1+m_1+2(m_2+\cdots+m_i)+m_{i+1}+\cdots+m_{i+k}}{2i+k}+1\right]$.
3	$N(E_l)=\Pi_1(E_l)\Pi_2(E_l)\Pi_3(E_l)$ corresponds to the three families of roots $\Pi_1(E_l)\Pi_2(E_l)=\begin{cases}\Pi_1(D_{l-1}) \text{ for } l=6,7 \text{ contains } (l-1)(l-2) \text{ factors}\\ \Pi_1(D_l) \text{ for } l=8 \text{ contains } l(l-1)=56 \text{ factors}\end{cases}$ $\Pi_3(E_l)$ for $l=6,7,8$ contains respectively $16,(32+1),64$ factors $\left[\frac{f(i)}{\delta(f(i))}+1\right]$ built up with the roots of the third family $N(E_l)=\left\{\prod_{i=2}^{l-1}(m_l+1)\prod_{i=2}^{l-k}\left[\frac{m_i+\cdots+m_{i+k-1}}{k}+1\right]\right\}\left\{(m_l+1)\prod_{i=2,3}^{4\le i+k\le l}\left[\frac{m_i+\cdots+m_{i+k-1}+m_l}{k+1}+1\right]\right.$ for $l=6$ $k=2,3,4$ $\left.\left[\frac{m_2+2m_3+m_4+m_l}{5}+1\right]\left[\frac{m_2+2m_3+m_4+m_5+m_l}{6}+1\right]\left[\frac{m_2+2m_3+2m_4+m_5+m_l}{7}+1\right]\right\}$ $\left\{\prod_{i=1}^{4}\left[\frac{m_1+\cdots+m_i}{i}+1\right]\left[\frac{m_1+m_2+m_3+m_4+m_l}{5}+1\right]\left[\frac{m_1+m_2+2m_3+m_4+m_l}{6}+1\right]\right.$ $\left.\left[\frac{m_1+2m_2+2m_3+m_4+m_l}{7}+1\right]\left[\frac{m_1+m_2+2m_3+2m_4+m_5+m_l}{8}+1\right]\prod_{i=9}^{16}\left[\frac{f(i)}{\delta(f(i))}+1\right]\right\}$ where $f(9)=f(3)+m_l$; $f(6+i)=f(i)+m_5$ for $4\le i\le 7$; $f(14)=f(13)+m_4$ $f(15)=f(14)+m_3$; $f(16)=f(15)+m_l$ To build $\Pi_3(E_7)\begin{cases}f(16+i)=f(i)+m_2+m_3+m_4+m_5+m_6 & \text{for } i=1,5,6,8,9,11,12,16\\ f(16+i)=f(i)+m_3+m_4+m_5+m_6+m_l & \text{for } i=2,3,4,7,10,13,14,15\\ f_{33}=2m_1+3m_2+4m_3+3m_4+2m_5+m_6+2m_7\end{cases}$ are needed To build $\Pi_3(E_8)\begin{cases}f(32+i)=f(16+i)+m_7\\ f(48+i)=f(i)+m_2+2m_3+2m_4+2m_5+2m_6+m_7+m_8\end{cases}1\le i\le 16$ are needed

The process of alternation: Starting from the representation space of the elementary IR of dimension, say, n for A_l, D_l, B_l, we can represent the Dynkin diagram given by $m_i=1$, $m_i=0$ for $i=2,3,\ldots,l$ by a Young diagram consisting of a single box. Then the Young diagram corresponding to the Dynkin diagram given by $m_i=\delta_{i,j}$ is a column of i boxes, i.e., a skew-symmetric tensor of rank i in $E^{\Delta n}$ and the number of linearly independent components of that tensor is equal to $\binom{n}{i}$; consequently the dimension of the IR given by the Dynkin diagram of the basic representation $m_i=\delta_{i,j}$ is also $\binom{n}{i}$ as a direct calculation using Tables IX and X yields (see Tables XI and XII). To make the above reasoning obvious for D_l, a relabelling of the roots interchanging i and $l-i$ has been used so that $\binom{2l}{l-i}$ becomes $\binom{2l}{i}$. The alternation process applied to E_6, E_7, E_8, C_l, F_4, G_2 yields reducible representations (except for few cases of E_6). In Tables XI and XII whenever possible Dynkin diagrams have been displayed with the dimension of the basic representation written below the corresponding simple root; possible reduction of the alternation process have also been expanded.

2. Construction of the representation matrices of s.s.L.a.

2. 1. Diagonal matrices: To each weight vector $\lambda_r^{(i)}$ $\binom{i=1,\ldots,q_r}{r=1,\ldots,T}$, corresponds a unique vector $v_r^{(i)}$ in the representation space E_N such that[1-14,16]

$$H_\mu v_r^{(i)}=(\mu,\lambda_r^{(i)})v_r^{(i)}$$

TABLE X. Dimensions of IR of W_{lzc} algebras.

c	z	
2	$l-1$	$N(B_l)=\prod_{\mu>0}\left[\frac{2\sum_{i=1}^{l-1}\mu^i m_i+\mu^l m_l}{2\sum_{i=1}^{l-1}\mu^i+\mu_l}+1\right]=\Pi_1(B_l)\Pi_2(B_l)\Pi_3(B_l)$ $\Pi_1(B_l)=\prod_{i=1}^{l-1}(m_i+1)\cdots\prod_{i=1}^{l+1-k}\left[\frac{m_i+m_{i+1}+\cdots+m_{i+k-1}}{k}+1\right]\cdots\left[\frac{m_1+m_2+\cdots+m_{l-1}}{l-1}+1\right]$ $\Pi_2(B_l)=(m_l+1)\prod_{i=1}^{l-1}\left[\frac{2(m_i+\cdots+m_{l-1})+m_l}{2(l-i)+1}+1\right]$ $\Pi_3(B_l)=\prod_{i=1}^{l-1}\left[\frac{m_i+\cdots+m_{l-1}+m_l}{l-i+1}+1\right]\prod_{i=1}^{i=l-2}\left[\frac{m_i+\cdots+m_{i+k-1}+2(m_{i+k}+\cdots+m_{l-1})+m_l}{2(l-i)-k+1}+1\right]$
2	1	$N(C_l)=\prod_{\mu>0}\left[\frac{2\mu^1 m_1+\sum_{i=2}^{l}\mu^i m_i}{2\mu^1+\sum_{i=2}^{l}\mu^i}+1\right]=\Pi_1(C_l)\Pi_2(C_l)\Pi_3(C_l)$ $\Pi_1(C_l)=\prod_{i=2}^{l}(m_i+1)\cdots\prod_{i=2}^{l+1-k}\left[\frac{m_i+\cdots m_{i+k-1}}{k}+1\right]\cdots\left[\frac{m_2+m_3+\cdots+m_l}{l-1}+1\right]$ $\Pi_2(C_l)=(m_1+1)\prod_{i=2}^{l}\left[\frac{m_1+m_2+\cdots+m_i}{i}+1\right]$ $\Pi_3(C_l)=\prod_{i=2}^{l}\left[\frac{2m_1+m_2+\cdots+m_i}{i+1}+1\right]\prod_{i=2}^{i=l-1}\left[\frac{2(m_1+\cdots+m_i)+m_{i+1}+\cdots+m_{i+k}}{2i+k}+1\right]$
2	2	$N(F_4)=\prod_{\mu>0}\left[\frac{2(\mu^1m_1+\mu^2m_2)+\mu^3m_3+\mu^4m_4}{2(\mu^1+\mu^2)+\mu^3+\mu^4}+1\right]$ $=\prod_{i=1}^{4}(m_i+1)\cdots\prod_{i=1}^{5-k}\left[\frac{m_i+m_{i+1}+\cdots+m_{i+k-1}}{k}+1\right]\left[\frac{2m_2+m_3}{3}+1\right]$ $\times\prod_{i,j,k=1,2}\left\{\left[\frac{im_1+2m_2+m_3}{3+i}+1\right]\left[\frac{m_1+jm_2+m_3+m_4}{3+j}+1\right]\left[\frac{2m_2+km_3+m_4}{3+k}+1\right]\right\}$ $\times\prod_{i=2,3}\left[\frac{m_1+im_2+2m_3+m_4}{4+i}+1\right]\prod_{j=3,4}\left[\frac{2m_1+jm_2+2m_3+m_4}{5+j}+1\right]$ $\times\prod_{i,j=1,2}\left\{\left[\frac{2m_1+2m_2+im_3+m_4}{5+i}+1\right]\left[\frac{2m_1+4m_2+3m_3+jm_4}{9+j}+1\right]\right\}$
3	1	$N(G_2)=(m_1+1)(m_2+1)\left(\frac{m_1+m_2}{2}+1\right)\left(\frac{2m_1+m_2}{3}+1\right)\left(\frac{3m_1+m_2}{4}+1\right)\left(\frac{3m_1+2m_2}{5}+1\right)$

with $\mu=\sum_{k=1}^{l}\mu^k\alpha_k$ being a positive root (all $\mu^k\in Z^+$). Hence

$$(H_\mu)^{r,\,i}_{r,\,i}=(\mu,\lambda_r^{(i)})=\sum_{k=1}^{l}\mu^k(\alpha_k,\lambda_r^{(i)}),$$

and $(H_\mu)^{r,\,i}_{r,\,i}$ is known when the $(H_{\alpha_k})^{r,\,i}_{r,\,i}=(\alpha_k,\lambda_r^{(i)})$, $k=1,\ldots,l$ are known. Due to the symmetry of the weight vector system, it suffices to write down its positive part only, i.e., the $\delta(\lambda_1)$ first layers if $\delta(\lambda_1)$ is an integer (as the following one gives the degeneracy of the null w.v.) or the $\delta(\lambda_1)+\frac{1}{2}$ first layers if $\delta(\lambda_1)$ is a half-integer. The complete matrix of order N can then be filled up with the opposite numbers (to get a zero trace matrix as expected).

From Eq. (34), using (5b) and (11), we get first

$$(S_r^{(i)},\alpha_k)=\sum_{j=1}^{l}[i_r^j(\alpha_j,\alpha_k)-i_r^{j-1}-i_r^{j+1}]\delta_{j,k};\tag{45}$$

hence for W_{lzc} with $z+1\leq k\leq l$

$$(H_{\alpha_k})^{r,\,i}_{r,\,i}=m_k/c-[(2i_r^k/c)-i_r^{k-1}-i_r^{k+1}];\tag{46}$$

for W_{lzc} with $1\leq k\leq z$, or for W_{lp1} one has to make $c=1$ in Eq. (46).

In case the w.v. $M=\lambda_r^{(i)}$ presents a degeneracy of order n_M, we get just as many identical diagonal elements.

A relatively general example of application of the formula (46) is given in the Appendix.

2.2. Nondiagonal matrices: The relation[11]

$$E_{-\mu}=-{}^tE_\mu\tag{47}$$

allows the study of E_μ for μ being only a positive root. As $E_\mu v_s\propto v_r$ $(v_s,v_r\in E_N)$, we have $\lambda_r=\lambda_s+\mu\in\{\text{w.v.}\}$, and the nonnull elements $(E_\mu)^r_s$ are such that

$$\lambda_r=\lambda_s+\mu,\tag{48a}$$

i.e., are situated in the lower half of the matrix (E_μ) and connect w.v. of layers whose power differ by $\delta(\mu)$; in other words,

$$r=s-\delta(\mu).\tag{48b}$$

[Of course, if μ is a simple root, $\delta(\mu)=1$ and $r=s-1$.]

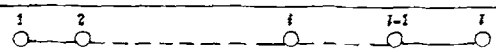
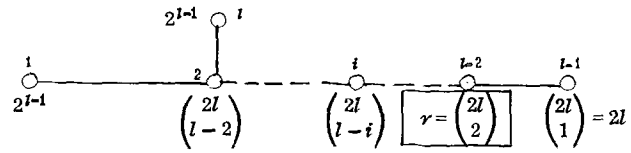
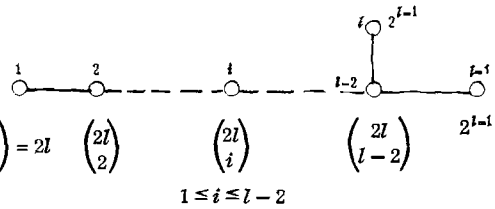
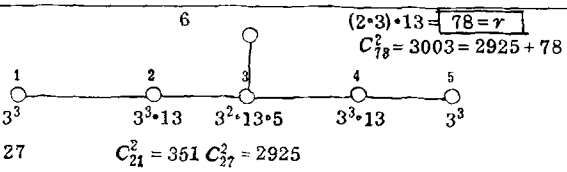
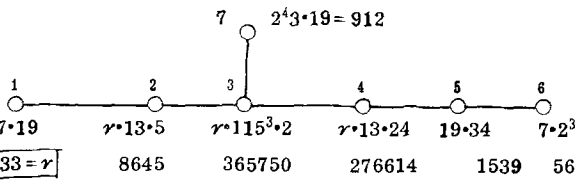
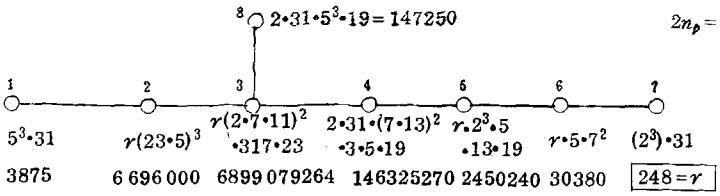
The proof of (48a) is well known; for any other positive root ν the commutation relation

$$[H_\nu,E_\mu]=(\nu,\mu)E_\mu\tag{48c}$$

yields

$$(H_\nu)^r_r(E_\mu)^r_s-(E_\mu)^r_s(H_\nu)^s_s=(\nu,\mu)(E_\mu)^r_s$$

TABLE XI. Dimensions of particular representation of W_{lp1}.

		$r=2n_p+l=l(h+1)$	Comments
A_l		$r=l(l+1)+l=l(l+2)$	*The R.R. corresponds to* the nonbasic IR
$l\geq 1$	$\binom{l+1}{1}=l+1 \qquad \binom{l+1}{i} \qquad \binom{l+1}{l}=l+1$		$m_1=1, m_l=1$, all other m_i being zero.
D_l	 $2\leq i\leq l-1$ 2^{nd} labeling		The ordering of the roots can be reversed because we have seen (IV.A) that we can have $p=2$ or $p=l-2$.
$l\geq 2$	 $\binom{2l}{1}=2l \quad \binom{2l}{2} \quad \binom{2l}{i} \quad \binom{2l}{l-2} \quad 2^{l-1}$ $1\leq i\leq l-2$		
E_6	 $(2\cdot 3)\cdot 13=78=r$ $C_{78}^2=3003=2925+78$ $3^3 \quad 3^3\cdot 13 \quad 3^2\cdot 13\cdot 5 \quad 3^3\cdot 13 \quad 3^3$ 27 $C_{21}^2=351$ $C_{27}^2=2925$	$2n_p=72$ $h=12$ $r=78$	For E_l as for D_l (cf. IV.A) the ordering of the roots can be reversed because we can have $p=3$ or $p=l-3$. Notice the symmetry of E_6.
E_7	 $2^4\cdot 3\cdot 19=912$ $7\cdot 19 \quad r\cdot 13\cdot 5 \quad r\cdot 115^3\cdot 2 \quad r\cdot 13\cdot 24 \quad 19\cdot 34 \quad 7\cdot 2^3$ $133=r \quad 8645 \quad 365750 \quad 276614 \quad 1539 \quad 56$ $C_{133}^2=8645+133$ $C_{133}^3=365750+2(8645+133)$ $C_{56}^4=365778071540$, $C_{56}^3=27664+56$ $C_{56}^2=1539+1$	$2n_p=126$ $h=18$ $r=133$	
E_8	 $2\cdot 31\cdot 5^3\cdot 19=147250$ $5^3\cdot 31 \quad r(23\cdot 5)^3 \quad r(2\cdot 7\cdot 11)^2\cdot 317\cdot 23 \quad 2\cdot 31\cdot(7\cdot 13)^2\cdot 3\cdot 5\cdot 19 \quad r\cdot 2^3\cdot 5\cdot 13\cdot 19 \quad r\cdot 5\cdot 7^2 \quad (2^3)\cdot 31$ 3875 6 696 000 6899 079264 146325270 2450240 30380 $248=r$ For E_8 (as shown above for E_7) the alternation process yields reducible representations.	$2n_p=240$, $h=30$, $r=248$	

or

$$(\nu,\lambda_r-\lambda_s-\mu)(E_\mu)_s^r=0;$$

hence, (48a).

If there is no degeneracy of the w.v. system, one has

$$(E_\mu)_s^r=\pm\{(H_\mu)_s^s+[(E_\mu)_t^s]^2\}^{1/2}. \tag{49a}$$

Indeed the commutation relation

$$[E_\mu,E_{-\mu}]=H_\mu \tag{49b}$$

yields

$$(E_\mu)_t^s(E_{-\mu})_s^t-(E_{-\mu})_r^s(E_\mu)_s^r=(H_\mu)_s^s \tag{49c}$$

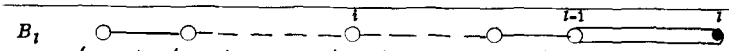
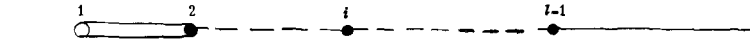
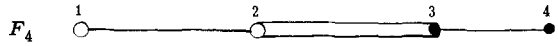
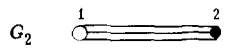
and, using (47), we get

$$[(E_\mu)_s^r]^2-[(E_\mu)_t^s]^2=(H_\mu)_s^s; \tag{49d}$$

hence (49a) which gives the elements of the nondiagonal matrices in terms of the elements of the diagonal matrices given by formula (46).

Notice that, (49a) being not linear, one cannot expect to get (E_μ) as a linear combination of the (E_α) with α as a simple root. Each matrix has to be calculated for

TABLE XII. Dimensions of particular representations of $W_{l\,sc}$.

		$r=2n_p+l=l(h+1)$	Comments
B_l 	$l\geq 2$ $\binom{2l+1}{1}$ $\binom{2l+1}{2}=r$ $\binom{2l+1}{i}$ $\binom{2l+1}{l-1}$ 2^l $1\leq i\leq l-1$	$r=2l^2+l=l(2l+1)=\binom{2l+1}{2}$	The R.R. corresponds to the basic I.R. $m_j=\delta_{2,j}$
C_l 	1st labeling $2\left(\frac{1}{l+2}\right)\binom{2l+1}{l}$ $2\left(\frac{i}{l+i+1}\right)\binom{2l+1}{l-i+1}$ $\left(2\,\frac{l-1}{2l}\right)\binom{2l+1}{2}$ $2l$ $1\leq i\leq l$ 2nd labeling $2l$ $(l-1)(2l+1)$ $2\left(\frac{l-i+1}{2l-i+2}\right)\binom{2l+1}{i}$ $2\left(\frac{1}{l+2}\right)\binom{2l+1}{l}$	$r=2l^2+l=l(2l+1)$	As well known $N(C_l)$ is always even (e.g., the factor 2 in $N(C_l)\beta_i$ for all i's). When l is odd, r is odd and no basic IR can be a regular one. For C_l the RR is reducible and using the scalar identity representation denoted $(C_l)\beta_0$ one has $(C_l)\mathrm{RR}=(C_l)\beta_{l-1}\oplus(C_l)\beta_l\oplus(C_l)\beta_0$
F_4  $(2^2)\cdot 13$ $2\cdot 13\cdot 7^2$ $13\cdot 7\cdot 3$ $13\cdot 2$ $52=r$ 1274 273 26 $C_{52}^2=1274+52$ $C_{26}^2=273+26+26$		$r=2\cdot 24+4=52$	The RR corresponds to the basic IR $m_j=\delta_{i,j}$
G_2  $2\cdot 7$ 7 $14=r$ $C_7^2=14+7$		$r=2\cdot 6+2=14$	The RR corresponds to the basic IR $m_j=\delta_{i,j}$

its own sake.

From (48b) we see that $r=s-\delta(\mu)=[t-\delta(\mu)]-\delta(\mu)$; so the calculation starts from $\lambda_s=-\lambda_1$ (the lowest w.v.) which yields $\lambda_t=0$, i.e., $(E_\mu)_t^s=0$; then $(E_\mu)_s^r=\pm[(H_\mu)_s^s]^{1/2}$ is known from Sec. 3A, Eq. (46), and the procedure is carried over by ascending along a parallel to the diagonal as $\delta(\mu)$ is fixed.

The commutation relation

$$[E_\mu,E_\nu]=N_{\mu,\nu}E_{\mu+\nu},\quad \mu,\nu,\mu+\nu\in\{\text{positive roots}\},\tag{50}$$

is used to obtain some coherence in signs.

If there is a degeneracy of the w.v. system, i.e., if in the same layer a certain w.v. M occurs with the multiplicity n_M, then the terms of the left-hand side of Eq. (49c) would be summed over the repeated indices t and r respectively.

Furthermore, as we have now n_M values $(H_\mu)_s^s$ which are identical (for $s=1,2,\ldots,n_M$) the number of independent equations is no more sufficient to determine all the matrix elements; the last commutation relation [Eq. (50)] is than a useful complement. One can also choose arbitrarily the values of the relevant matrix elements of one of the operators which is tantamount to choosing arbitrarily a basis in the degenerate subspace (of the w.v. system) of dimension n_M; but the values so obtained will depend on this choice.

Degeneracy is often met and complicates apparently simple problems as, for instance, the study[20] of the chain $G_2\supset A_2$.

CONCLUSIONS

The following results have been obtained in three steps:

1. In contrast to the point of view recently discussed in Refs. 17 and 21 consisting in breaking a given algebra into subalgebras, we have considered here the building of two classes of algebras out of known algebras.[15]

$$W_{l\,sc=1}=\{A_l,D_l,E_l \text{ for } l=6,7,8\},$$

$$W_{l\,sc\neq 1}=\{B_l,C_l,F_4,G_2\}.$$

This classification as well as this whole paper is based on Eq. (1) and on Chevalley's theorem,[7,14] stating that

the classification of Dynkin diagrams is equivalent to that of simple algebraic groups over closed fields of characteristic zero.

2. A study of the w.v. system has been performed using the results of Tables I and II of the first part. For the highest weight vector L, we have calculated its power $\delta(L)$ and shown, for $W_{l\rho c=1}$ (Table III) as well as for $W_{l\rho c\neq 1}$ (Table IV), that $\delta(L)$ is either integer or half-integer in agreement with the fact that $2\delta(L)+l=T$ is the integral number of layers (or shells) of the w.v. system whether this system is degenerate or not.

In case of the degeneracy of a particular weight vector M, Freudenthal's recursion formula gives the multiplicity n_M of M. In that formula as in Weyl's formula [Eq. (6)] comes in the form $R=\frac{1}{2}\sum_{\mu>0}\mu$ which can be deduced from Tables III and IV according to Theorem I; hence the eigenvalues of the Casimir operator (given in Tables V and VI) and width of weight diagrams are deduced.

3. The results obtained above have been used to build up the matrices of zero trace (diagonal and nondiagonal) representations for the two classes of algebras.

In the Appendix two examples are briefly studied to illustrate this paper.

ACKNOWLEDGMENTS

It is a pleasure to express my gratitude to the board of the Summer school (June 1976) of l'Université de Montréal, particularly to its Director, Professor A. Daigneault, for his kind hospitality at the Département de Mathématiques where this work has been done, as well as to Professor Hans Zassenhaus for his enlightening lectures on Lie groups. My thanks are also due to Miss J. Reggiori for her patient and careful typing.

APPENDIX

Example 1. $\underset{m_1}{\circ}\!\!-\!\!-\overset{c}{-}\!\!-\!\!-\underset{m_2}{\otimes} \in W_{21c}$

$$L=\lambda_1=[1/(4-c)][(2m_1+m_2)\alpha_1+(cm_1+2m_2)\alpha_2],$$

$$\delta(\lambda_1)=[1/(4-c)][(2+c)m_1+3m_2],\quad R=[1/(4-c)][3\alpha_1+(2+c)\alpha_2],$$

$$C=L(L+2R)=[1/(4-c)]\{(2m_1+m_2+6)m_1+[(cm_1+2m_2)+2(2+c)]m_2/c\}$$

or, equivalently, $C=[1/(4-c)]\{(2m_1+m_2)(m_1+2)+(cm_1+2m_2)[(m_2+2)/c]\}$.

According to Sec. 2D we can write:

for the second layer:

$\lambda_2^{(1)}=\lambda_1-\alpha_1\in\{\text{w.v.}\}$ if and only if $m_1\geq 1$;

$\lambda_2^{(2)}=\lambda_1-\alpha_2\in\{\text{w.v.}\}$ if and only if $m_2\geq 1$;

if $m_1m_2\neq 0$, then $\lambda_1^{(1)}$ and $\lambda_2^{(2)}\in\{\text{w.v.}\}$ with the same power $\delta(\lambda_2)=\delta(\lambda_1)-1$;

if $m_i=0$ $(i,j=1,2)$, then $\lambda_1^{(j)}\in\{\text{w.v.}\}$ but $\lambda_2^{(i)}\notin\{\text{w.v.}\}$ $(j\neq i)$;

for the third layer

$\lambda_3^{(1)}=\lambda_2^{(1)}-\alpha_1=\lambda_1-2\alpha_1\in\{\text{w.v.}\}$ if and only if $m_1\geq 2$;

$\lambda_3^{(2)}=\lambda_2^{(1)}-\alpha_2=\lambda_1-\alpha_1-\alpha_2$ w.v. even if $m_2=0$;

$\lambda_3^{(3)}=\lambda_2^{(2)}-\alpha_2=\lambda_1-2\alpha_2\in\{\text{w.v.}\}$ if and only if $m_2\geq 2$;

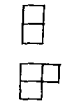
$\lambda_3^{(4)}=\lambda_2^{(2)}-\alpha_1=\lambda_1-\alpha_2-\alpha_1=\lambda_3^{(2)}\in\{\text{w.v.}\}$.

As $l=2$, there are no disconnected roots, and the third layer contains at least the degenerated w.v. $\{\lambda_3^{(2)}=\lambda_3^{(4)}\}$ and at most the four above w.v. with the same power $\delta(\lambda_3)=\delta(\lambda_1)-2$.

Particular cases can be considered:

for $c=1$, take $m_1=1$, $m_2=0$, corresponding to the Young diagram  of SU(3)

or $m_1=0$, $m_2=1$,

or $m_1=1$, $m_2=1$ 

with

$$\delta(\lambda_1)=2 \text{ and } \{\text{w.v.}\}=[\alpha_1+\alpha_2, \alpha_1, \alpha_2, 0, 0, -\alpha_2, \alpha_1, -\alpha_2-\alpha_1]$$

so that the dimension of the representation is 8 as forseen by Weyl's formula $(6, G_2)$; for $c=2$, L and R are obvious and $L(L+2R)=\frac{1}{2}[(2m_1+m_2+6)m_1+(2m_1+2m_2+8)m_2/2]$; for $c=3$, Weyl's formula (6) gives using (41b) for the dimension N

$$N(G_2)=(m_1+1)(m_2+1)[(m_1+m_2)/2+1][(2m_1+m_2)/3+1][(3m_1+m_2)/4+1][(3m_1+2m_2)/5+1]. \qquad (6, G_2)$$

For $m_1=0$, $m_2=1$ we have $N(G_2)=7$ and Freudenthal's formula gives $n_0=1$.
According to Sec. C.2.1. and summarizing what we know from before, we have

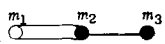
$$\{\text{w.v.}\}=\{\lambda_1; \lambda_1-\alpha_1, \lambda_1-\alpha_2; \lambda_1-\alpha_1-\alpha_2, \lambda_1-\alpha_2-\alpha_1, \lambda_1-2\alpha_1, \lambda_1-2\alpha_2; \cdots\}$$

$$H_{\alpha_1}=\begin{pmatrix} m_1 & & & & & & & & \\ & m_1-2 & & & & & & & \\ & & m_1+1 & & & & & & \\ & & & m_1-1 & & & & & \\ & & & & m_1-1 & & & & \\ & & & & & m_1-4 & & & \\ & & & & & & m_1+2 & & \\ & & & & & & & \ddots & \\ & & & & & & & & \ddots \end{pmatrix},$$

$$H_{\alpha_2}=\begin{pmatrix} m_2/c & & & & & & & & \\ & m_2/c & & & & & & & \\ & & (m_2-2)/c & & & & & & \\ & & & m_2/c+1-2/c & & & & & \\ & & & & m_2/c-2/c+1 & & & & \\ & & & & & m_2/c+2 & & & \\ & & & & & & (m_2-4)/c & & \\ & & & & & & & \ddots & \\ & & & & & & & & \ddots \end{pmatrix}.$$

If μ is a positive root such that $\mu=\sum_{k=1}^{l}\mu^k\alpha_k$, we get for this example

$$H_\mu=\mu^1 H_{\alpha_1}+\mu^2 H_{\alpha_2}.$$

Example II: Representations of C_3-algebra of group Sp(6):  $\in W_{312}$

Using Table II, we get

$$\lambda_1=L_{C_3}=\tfrac{1}{2}(3m_1+2m_2+m_3)\alpha_1+(2m_1+2m_2+m_3)\alpha_2+(m_1+m_2+m_3)\alpha_3,$$

$$\delta(L_{C_3})=\tfrac{1}{2}(9m_1+8m_2+5m_3), \quad \delta(L_{B_3})=3m_1+5m_2+3m_3, \quad R(C_3)=3\alpha_1+5\alpha_2+3\alpha_3,$$

$$C=\tfrac{1}{2}(3m_1+2m_2+m_3)(m_1+2)+(2m_1+2m_2+m_3)[(m_2+2)/2]+(m_1+m_2+m_3)[(m_3+2)/2].$$

Dimension:

$$N(C_3)=(m_1+1)(m_2+1)(m_3+1)[(m_1+m_2)/2+1][(m_2+m_3)/2+1][(2m_1+m_2)/3+1]$$
$$\times[(m_1+m_2+m_3)/3+1][(2m_1+m_2+m_3)/4+1][(2m_1+2m_2+m_3)/5+1]. \qquad (6, C_3)$$

Second layer: conditions for $\lambda_2^{(i)}$ to be a weight vector:

$\lambda_2^{(i)}=\lambda_1-\alpha_i$ if and only if $m_i\geq 1$, for $i=1,2$, $\delta(\lambda_2)=\delta(\lambda_1)-1$.

Third layer: conditions for the following vectors to be w.v. provided $\lambda_2^{(i)}\in\{\text{w.v.}\}$:

$\lambda_3^{(1)}=\lambda_2^{(1)}-\alpha_1=\lambda_1-2\alpha_1$ if and only if $m_1\geq 2$,

$\lambda_3^{(2)}=\lambda_2^{(1)}-\alpha_2=\lambda_1-\alpha_1-\alpha_2$ even if $m_2=0$,

$\lambda_3^{(3)}=\lambda_2^{(1)}-\alpha_3=\lambda_1-\alpha_1-\alpha_3$ if and only if $m_3\geq 1$,

$\lambda_3^{(4)}=\lambda_2^{(2)}-\alpha_1=\lambda_1-\alpha_2-\alpha_1=\lambda_3^{(2)}$ even if $m_1=0$,

$\lambda_3^{(5)}=\lambda_2^{(2)}-\alpha_2=\lambda_1-2\alpha_2$ if and only if $m_2\geq 2$,

$$\lambda_3^{(6)} = \lambda_2^{(2)} - \alpha_3 = \lambda_1 - \alpha_2 - \alpha_3 \qquad \text{even if } m_3 = 0,$$

$$\lambda_3^{(7)} = \lambda_2^{(3)} - \alpha_1 = \lambda_1 - \alpha_3 - \alpha_1 = \lambda_3^{(3)} \quad \text{if and only if } m_1 \geq 1,$$

$$\lambda_3^{(8)} = \lambda_2^{(3)} - \alpha_2 = \lambda_1 - \alpha_3 - \alpha_2 = \lambda_3^{(6)} \qquad \text{even if } m_2 = 0,$$

$$\lambda_3^{(9)} = \lambda_2^{(3)} - \alpha_3 = \lambda_1 - 2\alpha_3 \qquad \text{if and only if } m_3 \geq 2,$$

all with power $\delta(\lambda_3) = \delta(\lambda_1) - 2$.

Suppose $m_1 = m_2 = 0$, $m_3 = 1$, then $\delta(\lambda_1) = 5/2$, and we are left with the nondegenerate w.v. system:

$$\{\text{w.v.}\} = \{\tfrac{1}{2}\alpha_1 + \alpha_2 + \alpha_3;\ \tfrac{1}{2}\alpha_1 + \alpha_2;\ \tfrac{1}{2}\alpha_1;\ -\tfrac{1}{2}\alpha_1;\ -\tfrac{1}{2}\alpha_1 - \alpha_2;\ -\tfrac{1}{2}\alpha_1 - \alpha_2 - \alpha_3\},$$

so that the dimension of the corresponding representation is 6, as forseen by Weyl's formula $(6, C_3)$.

*On leave from Université de Nantes; permanent address: 24, rue de la Distillerie, 44000 Nantes, France.

N.B. The first seven references are given in order to justify the initials used to designate the two classes of algebras with possibilities of alternative interpretation.

[1]H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, N.J., 1946).
[2]S. Lie and F. Engel, *Theorie der Transformationsgruppen* (Leipzig, 1893).
[3]L.S. Pontrjagin, *Topological Groups* (Princeton U.P., Princeton, N.J., 1939).
[4]J. Patera and D. Sankoff, *Tables of Branching Rules* (Presses de l'Université de Montréal, Montréal, 1973).
[5]H. Zassenhaus, Lecture Notes of the Summer School (Université de Montréal, June 1976) on Lie Groups, Lie algebras and Representation Theory; *The Theory of Groups* (Chelsea, New York, 1958).
[6]E. Cartan, Thèse (Paris, 1894) [2nd ed. (Vuibert, Paris, 1933)].
[7]C. Chevalley, *Théorie des groupes de Lie* (Hermann, Paris, 1968).
[8]H. Bacry, *Leçons sur la théorie des groupes* (Gordon & Breach, New York, 1968).
[9]E.B. Dynkin, Am. Math. Soc. Transl. No. 17 (1950) and Ser. 2, 6 (1957), pp. 111–244.
[10]H. Freudenthal and H. de Vries, *Linear Lie Groups* (Academic, New York, 1969).
[11]N. Jacobson, *Lie Algebras* (Wiley-Interscience, New York, 1962).
[12]J.E. Humphreys, *Introduction to Lie Algebras* (Springer, New York, 1972).
[13]P.A. Rowlatt, *Group Theory and Elementary Particles* (Longmans Green, New York, 1966).
[14]I. Satake, *Classification Theory of Semi-Simple Algebraic Groups* (M. Dekker, New York, 1971).
[15]W. Laskar, "Highest weight of semisimple Lie algegras" (Cinqième colloque international sur les méthodes de la théorie des groupes en Physique, Montréal, July 1976).
[16]J.P. Serre, *Algebres de Lie semisimples complexes* (Benjamin, New York, 1966).
[17](a) J. Patera, R. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 17, 6 (1976). (b) J. Patera, R. Sharp, and P. Winternitz, J. Math. Phys. 17, 11 (1976).
[18]M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley, Reading, Mass., 1962).
[19]M. Bourbaki, *Groupes et algèbres de Lie* (Hermann, Paris, 1975), Chap. 8, p. 214.
[20]M. Perroud, J. Math. Phys. 17, 10 (1976).
[21]H. Bacry, Ph. Combe, and P. Sorba, Rept. Math. Phys. 5, 2, 145 (1974).

On a topological problem arising in physics

J. -P. J. Lafon

Observatoire de Meudon, Département Recherches Spatiales, 92190—Meudon, France
(Received 5 May 1975; revised manuscript received 28 June 1976)

The investigation of a large class of problems in physics requires the determination of the ranges of some parameters ξ, E, ··· for which equations of the form $F(r;\ \xi,\ E,\cdots)=0$ have some or no roots in a given bounded or unbounded interval of r. The solution of this problem is given under the form of three general theorems. Examples of utilization in physics are also discussed.

I. INTRODUCTION

In this paper we shall investigate the roots of equations of the form

$$F(r;\xi,E,\cdots)=0, \tag{1}$$

where $F(r;\xi,E,\cdots)$ denotes some given continuous differentiable function of a variable r and some parameters $\xi,E,\cdots$. We shall be concerned with the determination of the ranges of the parameters $\xi,E,\cdots$ for which Eq. (1) has some or no root in a given interval $[p,R]$.

This problem is encountered when studying many problems in physics.[1-10] A typical problem of this kind is the classification of the solutions of some differential equations using known invariant integrals: Examples concerning systems of electrically charged particles are given and discussed in detail in Sec. X. Other examples are mentioned in Sec. XI.

The form of the function F and the parameters suitable for the analysis of different systems may vary widely, so that it is interesting to solve the problem for a large class of functions F. However, although such a problem arises in various fields of physics, such as plasma physics, stellar dynamics, etc., it has not been investigated from a general point of view. Authors dealing with problems in different fields have used various techniques more or less adapted to the particular problem that they considered and often requiring too much numerical computations.[1-4]

In any case, the determination of the set of parameters such that F has no root $r\in[p,R]$ is very difficult, even numerically; this set is characterized by the conditions

$$F(r;\xi,E,\cdots)>0 \quad \forall r\in[p,R], \tag{2}$$

or

$$F(r;\xi,E,\cdots)<0 \quad \forall r\in[p,R]. \tag{3}$$

However, under slightly restrictive conditions, this set is an open domain D of the $\xi,E,\cdots$ space delimited by some continuous curves, surfaces or hypersurfaces, depending on the dimension of the $\xi,E,\cdots$ space. The collective treatments of the parameters, such as integration over D for instance, or numerical computations, if any, would be much easier if D was characterized by its boundary.

As explained in Sec. IX, the problem is of a similar nature whatever the dimension n of the $\xi,E,\cdots$ space. For simplicity we shall discuss in detail only the case where $n=2$. The generalizations when $n>2$ are straightforward; they will be considered in Sec. IX together with some other possible generalizations.

The basic assumptions are stated at the beginning of Sec. II. They will be somewhat relaxed in Sec. IX. In Sec. I we characterize the topological structure of the regions of the ξE plane where, respectively $F(r,\xi,E)>0$ for any $r\in[p,R]$, $F(r,\xi,E)<0$ for any $r\in[p,R]$, $F(r,\xi,E)=0$ for some $r\in[p,R]$ (Theorem 1). Then we give the statement of two other main theorems of the paper; they are concerned with the determination of the boundary Γ_{pR} of the region where $F(r,\xi,E)>0$ for any $r\in[p,R]$ (Theorems 2 and 3). Of course similar theorems concerning the boundary Γ^-_{pR} of the region where $F(r,\xi,E)<0$ for any $r\in[p,R]$ may be stated in a similar way.

The proof of Theorem 2 is prepared by five lemmas (Sec. III—VII). In Lemmas 1 and 2 the points of Γ_{pR} are characterized. In Lemmas 3—5 it is proved that Γ_{pR} is a continuous curve made up of arcs of some well characterized simple curves. Theorem 2 is proved in Sec. VIII and generalized in Sec. IX in the cases where the function F satisfies conditions less restrictive than those stated in Sec. II or depends on more than two parameters. The case where there are $N>1$ functions F depending on more than two parameters in addition to N variables is also considered in Sec. IX. In Sec. X we give some examples showing how Theorems 1—3 can be used for solving physical problems. Section XI is devoted to the conclusions with mention of other papers in which we have used the present theory.

II. ASSUMPTIONS AND CONCLUSIONS

Hereafter we shall investigate in detail the problems stated in Sec. I in the case where Eq. (1) depends on two parameters ξ and E. We confine our consideration to this case for ease in the argumentation. The assumptions are simple but involve a large number of various cases. However all of them are not necessary. This point will be discussed, together with some straightforward generalizations of the theorems stated hereafter, in Sec. IX.

Throughout the paper, except in Sec. IX, the following notations will be used (see Fig. 1):

For any $r\in[p,R]$, P_r denotes the set of points of the ξE plane for which $F(r,\xi,E)=0$;

For any $r\in[p,R]$, W_{rr} and W^-_{rr} are the regions of the

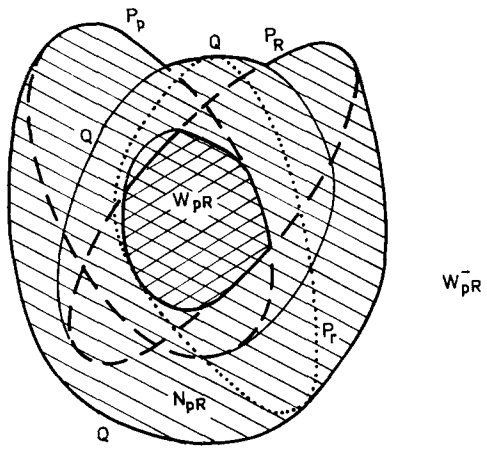


FIG. 1. Example of regions where $F(r,\xi,E)>0\ \forall\, r\in[p,R]$ $(\mathring{W}_{pR})$, $F(r,\xi,E)<0\ \forall\, r\in[p,R]$ $(\mathring{W}^{-}_{pR})$, and $F(r,\xi,E)=0$ for some $r\in[p,R]$ (N_{pR}) in a simple case where conditions 3 of Theorem 2 are satisfied for $p\leq r\leq R$.

ξE plane in which, respectively, $F(r,\xi,E)\geqslant 0$ and $F(r,\xi,E)\leqslant 0$;

For any α, β with $p\leqslant\alpha,\beta\leqslant R$, $W_{\alpha\beta}=\cap_{\alpha\leqslant r\leqslant\beta}W_{rr}$ $W^{-}_{\alpha\beta}=\cap_{\alpha\leqslant r\leqslant\beta}W^{-}_{rr}$. Of course $W_{\beta\alpha}=W_{\alpha\beta}$ and $W^{-}_{\beta\alpha}=W^{-}_{\alpha\beta}$;

For any α,β with $p\leqslant\alpha,\beta\leqslant R$, $\Gamma_{\alpha\beta}$ and $\Gamma^{-}_{\alpha\beta}$ are the boundaries of $W_{\alpha\beta}$ and $W^{-}_{\alpha\beta}$, respectively.

$$N_{pR}=\complement(\mathring{W}_{pR}\cup\mathring{W}^{-}_{pR}),\tag{4}$$

which means that N_{pR} is the complement of $\mathring{W}_{pR}\cup\mathring{W}^{-}_{pR}$.

Throughout Sec. II—VIII, unless otherwise stated, the assumptions are unchanged. Thus, in order to avoid tedious repetitions, we list them hereafter under the title "Assumptions (H)."

Assumptions (H)

(1) $F(r,\xi,H)$ is r,ξ,E continuously differentiable for any ξ, E and for $p\leqslant r\leqslant R$.

(2) $F'_{\xi}(r,\xi,E)$ and $F'_{E}(r,\xi,E)$ are never simultaneously zero.

(3) The interval $[p,R]$ can be divided into a finite number of subsequent subintervals for any r in which

(a) $F(r,\xi,E)=0$ and $F'_{r}(r,\xi,E)=0$ for any r and for a finite number of couples ξ, E depending or not on r, or

(b) $F(r,\xi,E)=0$ implies $F'_{r}(r,\xi,E)\neq 0$ except for a finite number of $r\in[p,R]$ for which $F'_{r}(r,\xi,E)\equiv 0$ for all ξ,E for which $F(r,\xi,E)=0$.

(4) All the roots of $F(r,\xi,E)=0$ and $F'_{r}(r,\xi,E)=0$ for $p\leqslant r\leqslant R$ are in bounded ξ and E intervals.

Remarks: These assumptions can be interpreted geometrically as follows:

(1) and (2) For any r, W_{rr} is the closed domain of the ξE plane in which $F(r,\xi,E)\geqslant 0$; $\mathring{W}_{rr}$ is the open domain in which $F(r,\xi,E)>0$; W_{rr} is bounded by a smooth curve P_r without singular point; P_r is the closed set of points for which $F(r,\xi,E)=0$, it separates $\mathring{W}_{rr}$ from the region in which $F(r,\xi,E)<0$, which is $\mathring{W}^{-}_{rr}$.

(1)—(3) The interval $[p,R]$ can be divided into a finite number of subintervals for any r in which:

(a) the curves P_r have an envelope Q, with arcs touching each curve P_r at a finite number of points and with a finite number of isolated points, or

(b) the curves P_r have no envelope, except for a finite number of isolated $r\in[p,R]$ for which there are stationary curves P_r.

(4) All the points of Q are contained in a bounded domain of the ξE plane.

In the sequel we shall currently use the following other consequences of Assumptions (H):

(1) $F(r,\xi,E)$ is ξ,E continuous uniformly over $[p,R]$.

We do not give a detailed proof of this property of $F(r,\xi,E)$. It is a well known consequence of both the r,ξ,E continuity of $F(r,\xi,E)$ and the Borel—Lebesgue topological theorem.

$$(2)\quad \mathring{W}_{pR}=\cap_{p\leqslant r\leqslant R}\mathring{W}_{rr}.\tag{5}$$

Proof: of course

$$\mathring{W}_{pR}\subset\bigcap_{p\leqslant r\leqslant R}\mathring{W}_{rr}.\tag{6}$$

Then if some Π_0 with coordinates ξ_0, E_0 is in all the regions $\mathring{W}_{rr}$ for $p\leqslant r\leqslant R$, $F(r,\xi_0,E_0)>0\ \forall\ r\in[p,R]$. Therefore, since $F(r,\xi,E)$ is ξ,E continuous uniformly over $[p,R]$, there is an open neighborhood θ of Π_0 in the ξE plane for points in which $F(r,\xi,E)>0\ \forall\ r\in[p,R]$. Of course $\theta\subset\cap_{p\leqslant r\leqslant R}W_{rr}=W_{pR}$. Since θ is an open domain $\theta\subset\mathring{W}_{pR}$ Now, if $\cap_{p\leqslant r\leqslant R}\mathring{W}_{rr}=\phi$, from (6) it follows that $\mathring{W}_{pR}=\phi$. This completes the proof of (5).

$$(3)\quad \text{For } p\leqslant\alpha\leqslant\gamma\leqslant\beta,\quad W_{\alpha\beta}=W_{\alpha\gamma}\cap W_{\gamma\beta}.\tag{7}$$

This is a straightforward consequence of the definition of $W_{\alpha\beta}$.

Now we prove a theorem concerning the topological structure of the regions $\mathring{W}_{pR}$, $\mathring{W}^{-}_{pR}$, and N_{pR}, and two others concerning their delimitation.

Theorem 1: When Assumptions (H) are satisfied, $\mathring{W}_{pR}$ and $\mathring{W}^{-}_{pR}$ are the regions of the ξE plane where respectively $F(r,\xi,E)>0$ and $F(r,\xi,E)<0$ for any $r\in[p,R]$; N_{pR} contains only the points of all the curves P_r for $p\leqslant r\leqslant R$. Moreover,

$$N_{pR}=\left(\bigcup_{p\leqslant r\leqslant R}W_{rr}\right)\cap\complement\mathring{W}_{pR}=\left(\bigcup_{p\leqslant r\leqslant R}W^{-}_{rr}\right)\cap\complement\mathring{W}^{-}_{pR}\tag{8}$$

and

$$\mathring{N}_{pR}\neq\phi.\tag{9}$$

Finally the regions $\mathring{W}_{pR}$, $\mathring{W}^{-}_{pR}$, N_{pR} cover the ξE plane and two of them have no common point (Fig. 1).

Proof: For any $r\in[p,R]$, W_{rr} is the only region of the ξE plane where $F(r,\xi,E)>0$ [Assumptions (H)]. Then from (5) it follows that $\mathring{W}_{pR}$ is that region in which $F(r,\xi,E)>0$ for any $r\in[p,R]$. Of course, similarly, $\mathring{W}^{-}_{pR}$ is that region in which $F(r,\xi,E)<0$ for any $r\in[p,R]$.

Now from (4) it follows that, for any point Π_0 with coordinates ξ_0, E_0 in N_{pR}, $F(r,\xi_0,E_0)\geqslant 0$ for some $r\in[p,R]$ and $F(r,\xi_0,E_0)\leqslant 0$ for some other $r\in[p,R]$;

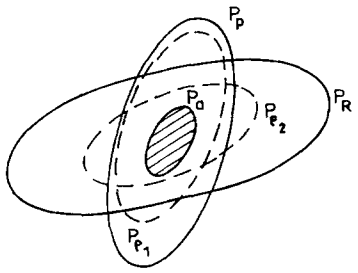


FIG. 2. Example of region W_{pR} when conditions of Lemma 3 are satisfied, $p<\rho_1<a<\rho_2<R$. P_a is a stationary curve. This figure illustrates also the case where conditions 1 of Theorem 2 are satisfied.

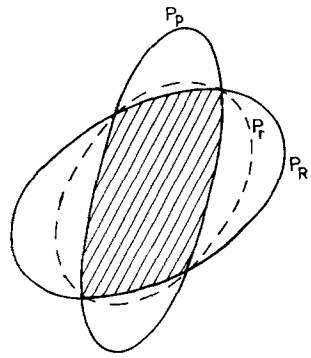


FIG. 3. Example of region W_{pR} when conditions 2 of Theorem 2 are satisfied for $p \leq r \leq R$.

since $F(r,\xi,E)$ is continuous there is at least one $r\in[p,R]$, say δ, and a corresponding curve P_δ such that $p\leq\delta\leq R$, $F(\delta,\xi_0,E_0)=0$ and so $\Pi_0\in P_\delta$.

Conversely, for any $r\in[p,R]$, $P_r\subset C\mathring{W}_{rr}$. Thus from (6) it follows that $P_r\subset C\mathring{W}_{pR}$. Of course similarly $P_r\subset C\mathring{W}^-_{rr}$ and $P_r\subset C\mathring{W}^-_{pR}$. Therefore, $P_r\subset N_{pR}$.

Finally N_{pR} contains only the points of all the curves P_r for $p\leq r\leq R$.

Then define K as $K=\cup_{p\leq r\leq R}W_{rr}$. K is the set of points with coordinates ξ, E for which there is always some $r\in[p,R]$ for which $F(r,\xi,E)\geq 0$. Consequently $K=C\mathring{W}_{pR}$. Of course similarly $\cup_{p\leq r\leq R}W^-_{rr}=C\mathring{W}^-_{pR}$. This proves (8).

Finally there is at least one point Π_0 with coordinates ξ_0, E_0 and two values of r in $[p,R]$, r_1 and r_2, for which

$$\Pi_0\in\mathring{W}_{r_1r_1} \quad\text{and}\quad \Pi_0\notin W_{r_2r_2}.$$

Otherwise for any α, β with $p\leq\alpha,\beta\leq R$, $\mathring{W}_{\alpha\alpha}\subset W_{\beta\beta}$ and, since $F(r,\xi,E)$ is continuous $W_{\alpha\alpha}=W_{\beta\beta}$, which is not consistent with the assumption that W_{rr} varies with r.

Therefore, $F(r_1,\xi_0,E_0)>0$, $F(r_2,\xi_0,E_0)<0$. Since $F(r,\xi,E)$ is ξ,E continuous, uniformly over $[p,R]$ there is an open neighborhood θ of Π_0 where $F(r_1,\xi,E)>0$ and $F(r_2,\xi,E)<0$ so that $\theta\subset N_{pR}$. Thus $\Pi_0\in\mathring{N}_{pR}\neq\emptyset$.

Theorem 2: There is some r, say L, such that $p<L\leq R$ and for $R>r\geq L$, $W_{pr}=\emptyset$, whereas for $p\leq r<L$, $W_{pR}\neq\emptyset$. The interval $[p,L]$ can be divided into a finite number of subintervals (r_j,r_{j+1}) $(j=1,2,\ldots,m;$ $r_1=p$; $r_m=L)$ for r in which $W_{pR}\neq\emptyset$ and can be delimited as follows:

(1) If for $r>r_j$, in the neighborhood of r_j, there is no arc and no isolated point of Q, let r_{j+1} denote the lowest $r>r_j$ for which an arc or an isolated point of Q appears. Then for $r_j<r\leq r_{j+1}$, W_{pr} can be delimited using Lemma 3 (Sec. V) (Fig. 2).

(2) If for $r>r_j$, in the neighborhood of r_j, there are only isolated points of Q, in particular there is no stationary curve and no arc of Q, let r_{j+1} denote the upper bound of the $r>r_j$ such that this assumption is valid for the whole open interval $]r_j,r[$. Then $W_{pr}=W_{pr_j}\cap W_{rr}$ for $p<r\leq r_2$ (Fig. 3).

(3) If for $r>r_j$ in the neighborhood of r_j there are some arcs of Q, let r_{j+1} denote the lowest $r>r_j$ for which P_r is a stationary curve, or some arc of Q reaches an isolated point of Q or has a singular point or a common point with another arc of Q (Figs. 1 and 4). Then, for $r_j<r\leq r_{j+1}$ each subarc of these arcs of Q contained in the interior of $W_{pr_j}\cap W_{rr}$ divides $W_{pr_j}\cap W_{rr}$ into two regions one of which contains W_{pr}; W_{pr} is the part of $W_{pr_j}\cap W_{rr}$ common to all these regions.

Proof: The proof of this theorem is detailed in Sec. VIII; it is based on five lemmas proved in Secs. III—VII. Lemmas 1 and 2 are concerned with the properties of the points of Γ_{pR}. Lemmas 3—5 are concerned with the topological structure of W_{pR} and Γ_{pR}: Γ_{pR} is a continuous curve made up of arcs of some curves which are data of the problem.

One may also state and prove the following similar theorem:

Theorem 3: In Theorem 2 replace p by R and reverse all the relations of order. One obtains the statement of a new theorem which can be proved in the same way as Theorem 2.

III. LEMMA 1

Statement: Under Assumptions (H) W_{pR} is a closed domain of the ξE plane. If $W_{pR}\neq\emptyset$, Γ_{pR} is made up of points of P_p, P_R, Q and some stationary curves P_{a_1}, $P_{a_2},\ldots,P_{a_n}$ $(p<a_1<a_2<\cdots<a_n<R)$.

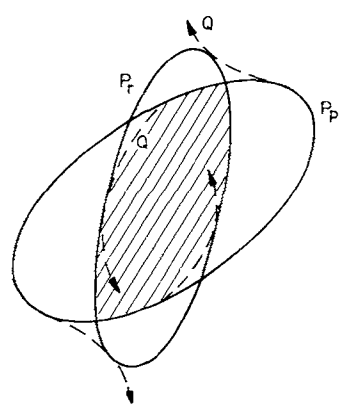


FIG. 4. Example of region W_{pr} in a simple case where conditions 1 of Theorem 2 are satisfied.

Proof: W_{pR} is the intersection of the closed domains W_{rr} for $p \leq r \leq R$; consequently it is also a closed domain of the ξE plane. Now, let Π_0 denote a point of Γ_{pR} with coordinates ξ_0, E_0.

$$F(r, \xi_0, E_0) \geq 0 \quad \forall\, r \in [p, R] \tag{10}$$

Moreover there is at least one r, say r_0, for which $p \leq r_0 \leq R$ and

$$F(r_0, \xi_0, E_0) = 0. \tag{11}$$

Otherwise, since $F(r, \xi, E)$ is ξ, E continuous uniformly over $[p, R]$ (Sec. II), there would be an open neighborhood of Π_0 in the ξE plane for points in which

$$F(r, \xi_0, E_0) > 0 \quad \forall\, r \in [p, R]$$

which is not consistent with assumption that Π_0 is a point of the boundary Γ_{pR} of the closed domain W_{pR}. Then from (10) it follows that three cases are possible:

$$r_0 = p$$

and then Π_0 is on P_p, or

$$r_0 = R$$

and then Π_0 is on P_R, or

$$p < r_0 < R$$

and from (10) and (11) it results that $F(r_0, \xi_0, E_0)$ is a local minimum of $F(r, \xi_0, E_0)$ so that

$$F'_r(r_0, \xi_0, E_0) = 0. \tag{12}$$

Then from (11) and (12) it follows that Π_0 is either a point of the envelope Q of the curves P_r or a point of some stationary curve P_r (here P_{r_0}).

IV. LEMMA 2

Statement: Under Assumptions (H) any point on Q or on a stationary curve is either on Γ_{pR} or not in W_{pR}.

Proof: Let Π_0 denote a point of Q, or a point of some stationary curve, with coordinates ξ_0, E_0. There is at least one r, say r_0, for which

$$F(r_0, \xi_0, E_0) = 0 \quad \text{and} \quad F'_r(r_0, \xi_0, E_0) = 0.$$

Now any open neighborhood of Π_0 in the ξE plane is divided by P_{r_0} into two parts, one of which contains points for which $F(r_0, \xi, E) < 0$, i.e., points of the exterior of W_{pR}. Consequently, Π_0 cannot be in the interior of W_{pR}.

V. LEMMA 3

Statement: In addition to Assumptions (H) assume that for $p < r < R$ there is no arc of Q, in other words for the points of the curves P_r, $F'_r(r, \xi, E) \neq 0$, except for all the points of some stationary curves $P_{a_1}, P_{a_2}, \ldots, P_{a_n}$ for which $F'_r(r, \xi, E) = 0$.

Then, if $W_{pR} \neq \emptyset$, it is the intersection of W_{pp}, W_{RR}, $W_{a_1 a_1}$, $W_{a_2 a_2}, \ldots, W_{a_n a_n}$ (Fig. 2). Moreover, if there is no stationary curve and the regions W_{rr} are connected for $p < r < R$, $W_{pp} \subset W_{RR}$ or $W_{RR} \subset W_{pp}$ and $W_{pR} \neq \emptyset$ and is equal to that of the regions W_{pp} and W_{RR} which is contained in the other.

Proof: First assume that there is no stationary curve for $p < r < R$. Of course, if $W_{pR} \neq \emptyset$, $W_{pR} \subset W_{pp} \cap W_{RR}$; besides from Lemma 1 it follows that Γ_{pR} is made up of only arcs of P_p and P_R. Consequently, $W_{pR} = W_{pp} \cap W_{RR}$. Now, if there are some r, $a_1, a_2, \ldots, a_i, \ldots, a_n$ for which

$$p < a_1 < a_2 < \cdots < a_i < \cdots < a_n < R$$

and

$$F'_r(r, \xi, E) = 0$$

for the points of all the curves P_{a_i}, similar arguments can be applied to the case where r varies between two successive values a_i, a_{i+1} or between p and a_1, or between a_n and R. Thus, if $W_{pR} \neq \emptyset$,

$$W_{pR} = W_{pa_1} \cap \left(\bigcap_i W_{a_i a_{i+1}} \right) \cap W_{a_n R}$$

$$= W_{pp} \cap \left(\bigcap_i W_{a_i a_i} \right) \cap W_{RR}$$

which proves the first assertion.

Now assume that there is no stationary curve for $p < r < R$ and that the regions W_{rr} are connected. Define M_{pR} as $M_{pR} = \cup_{p<r<R} P_r$. Any point Π of M_{pR} has coordinates ξ, E such that $F(r, \xi, E) = 0$ for some r for which $p < r < R$ and $F'_r(r, \xi, E) \neq 0$. Moreover, $F(r, \xi, E)$ is continuously differentiable. Thus for any point $\Pi \in M_{pR}$ there is an open neighborhood of Π included in M_{pR}, so that M_{pR} is an open domain. Besides there is a continuously differentiable function $\varphi(\xi, E)$ such that, for all the points of M_{pR}, the equation $F(r, \xi, E) = 0$ is equivalent to $r = \varphi(\xi, R)$. Then, since $\varphi(\xi, E)$ is continuous, it can be continued for the points of M_{pR} using the closure of the values of $\varphi(\xi, E)$, i.e., the closed interval $[p, R]$. Of course, for the points of $\overline{M_{pR}}$, $p \leq r = \varphi(\xi, E) \leq R$ and $F(\varphi(\xi, E), \xi, E) = 0$; thus $\overline{M_{pR}} = N_{pR}$; moreover, if all the regions W_{rr} are connected for $p \leq r \leq R$, the condition $F(r, \xi, E) > 0$ is equivalent to one of the conditions $r < \varphi(\xi, E)$ and $r > \varphi(\xi, E)$ and the condition $F(r, \xi, E) < 0$ is equivalent to the other. Consequently $W_{pp} \subset W_{RR}$ or $W_{RR} \subset W_{pp}$. In any case, $W_{pR} \neq \emptyset$ and $W_{pR} = W_{pp}$ or W_{RR}.

VI. LEMMA 4

Statement: In addition to Assumptions (H) assume that some point Π_0 of Q is not strictly isolated. Then one of the following conditions is satisfied:

$\Pi_0 \notin W_{pR}$ and there is an open curvilinear neighborhood of Π_0 on Q out of W_{pR}; or

$\Pi_0 \in \Gamma_{pR}$ and there is an open curvilinear neighborhood of Π_0 on Q which is an arc of Γ_{pR}; or

$\Pi_0 \in \Gamma_{pR}$ and Π_0 is a singular point of Q, or a common point of two or many arcs of Q, or a common point of Q and P_p, P_R or some stationary curve, or an isolated point of Q on an arc of Q.

Proof: Let ξ_0, E_0 denote the coordinates of Π_0 in the ξE plane. There is at least one r, say r_0, for which

$$F(r_0, \xi_0, E_0) = 0 \quad \text{and} \quad F'_r(r_0, \xi_0, E_0) = 0.$$

Assume that Π_0 is not isolated.

If Π_0 is out of W_{pR} there are some r for which

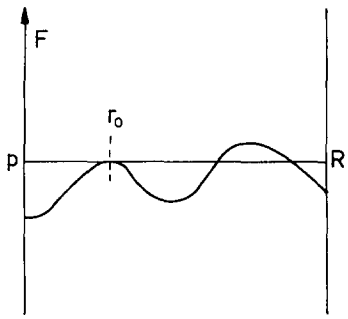


FIG. 5. Example of function $F(r,\xi,E)$ for $\xi=\xi_0$, $E=E_0$ coordinates of a point Π_0 at which a curve P_{r_0} touches Q. $\Pi_0 \notin W_{pR}$.

$F(r,\xi_0,E_0)<0$, either in a neighborhood of r_0 (Figs. 5 and 6) or not (Fig. 7). Now since $F(r,\xi,E)$ is ξ,E continuous uniformly over $[p,R]$ (Sec. II) there is always an open neighborhood θ of Π_0 in the ξE plane for points in which $F(r,\xi,E)<0$ for some r, in the neighborhood of those for which $F(r,\xi_0,E_0)<0$. Of course

$$(Q\cap\theta)\cap W_{pR}=\phi.$$

If Π_0 is not out of W_{pR}, from Lemma 2 it results that $\Pi_0\subset\Gamma_{pR}$. Thus condition (10) is satisfied. Now assume that $F(p,\xi_0,E_0)\neq 0$ and $F(R,\xi_0,E_0)\neq 0$, i.e., $\Pi_0\notin P_p$, $\Pi_0\notin P_R$. Assume also that no open curvilinear neighborhood of Π_0 on Q is contained in Γ_{pR} in its entirety. Then in any curvilinear neighborhood θ_c of Π_0 on Q there is at least one point $\Pi_1(\theta_c)$, with coordinates ξ_1, E_1 at which some curve P_{r_1} touches Q and for which $F(r_1,\xi_1,E_1)=0$, $F'_r(r_1,\xi_1,E_1)=0$, and $F(r,\xi_1,E_1)<0$ for some r for which $p<r<R$. Now, since Π_0 is neither on P_p nor on P_R, θ_c can be reduced in such a way that $\theta_c\subset W_{pp}\cap W_{RR}$; then $F(p,\xi_1,E_1)>0$, $F(R,\xi_1,E_1)>0$, and $F(r,\xi_1,E_1)$ has a negative minimum reached for some r, say r_2, for which

$$F'_r(r_2,\xi_1,E_1)=0,\quad r_2\neq p,\quad r_2\neq R.$$

Since $F(r,\xi,E)$ is ξ,E continuous, uniformly over $[p,R]$ (Sec. II), it is uniformly continous in particular for ξ, E coordinates of points of Q. Consequently $F(r,\xi_1,E_1)$ can be negative only in neighborhoods of the r for which $F(r,\xi_0,E_0)=0$. Let r^i $(i=1,2,\ldots,n)$ denote the roots of $F(r,\xi_0,E_0)$ in neighborhoods of which $F(r,\xi_1,E_1)<0$ for Π_1 in some neighborhoods of Π_0 on Q. For any open neighborhood $]r^i-h,r^i+h[$ of an r^i one can find an open curvilinear neighborhood of Π_0 on Q,

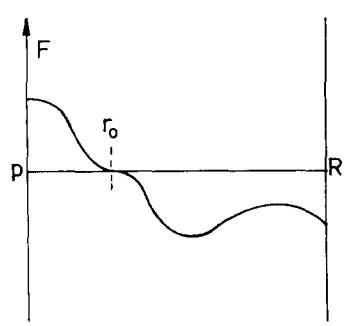


FIG. 6. Example of function $F(r,\xi,E)$ for $\xi=\xi_0$, $E=E_0$ coordinates of a point Π_0 at which a curve P_{r_0} touches Q. $\Pi_0\notin W_{pR}$.

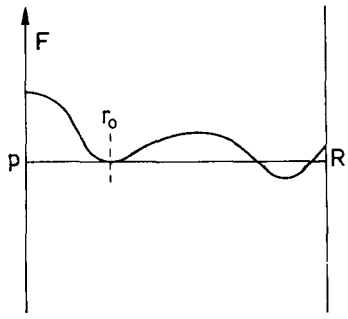


FIG. 7. Example of function $F(r,\xi,E)$ for $\xi=\xi_0$, $E=E_0$ coordinates of a point Π_0 at which a curve P_{r_0} touches Q. $\Pi_0\notin W_{pR}$.

θ_h, for points in which all the values lower than or equal to zero that $F(r,\xi,E)$ reaches near r^i correspond to values of r for which $r^i-h<r<r^i+h$ (Figs. 8 and 9).

(*) If there are one or more $r^i\neq r_0$, then since, for all the r^i, $r^i\neq p$, $r^i\neq R$, and $F(r^i,\xi_0,E_0)=0$, condition (10) implies that, for any r^i, $F'_r(r^i,\xi_0,E_0)=0$ (Fig. 8). Consequently, Π_0 may be an isolated point of Q, not strictly isolated but contained in an arc of Q; otherwise it is a common point of two or many special curves: Π_0 may be simultaneously on two or many arcs of Q corresponding to r in neighborhoods of r_0 and some r^i, or on some arcs of Q and stationary curves corresponding to different r^i.

(*) If some $r^i=r_0$, for any h one can choose $\theta_c\subset\theta_h$ in such a way that (Fig. 9)

$$r_0-h<r_{1,2}<r_0+h.$$

Finally for any k and any open neighborhood θ of Π_0 in the ξE plane, one can find $h\leq k$ and $\theta_c\subset\theta_h\subset\theta\cap Q$ in such a way that there is at least one point Π_1 with coordinates ξ_1, E_1 in θ_c, and so in θ, for which the equation $F'_r(r,\xi_1,E_1)=0$ has at least two different roots r_1 and r_2 for which

$$r_0-k\leq r_0-h<r_{1,2}<r_0+h\leq r_0+k.$$

Thus Π_0 cannot be a regular point on Q.

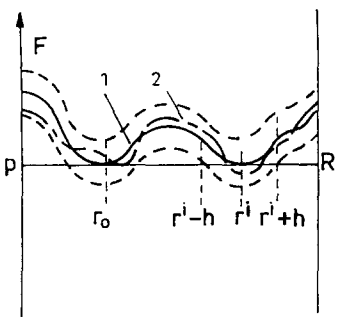


FIG. 8. Example of function $F(r,\xi,E)$ for $\xi=\xi_0$, $E=E_0$ coordinates of a point Π_0 where two curves P_{r_0} and P_{r^i} touch Q. $\Pi_0\in\Gamma_{pR}$ and is the intersection of two arcs of Q corresponding to r in the neighborhood of r_0 and r^i respectively (curve 1). Curve 2 shows an example of function $F(r,\xi,E)$ for ξ, E coordinates of a point in a neighborhood of Π_0 on Q where $F(r,\xi,E)<0$ only for $r^i-h\leq r\leq r^i+h$.

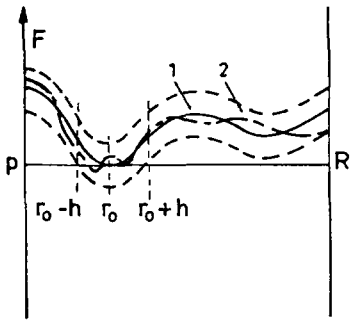


FIG. 9. Example of function $F(r,\xi,E)$ for $\xi=\xi_0$, $E=E_0$ coordinates of a singular point of Q (curve 1). Curve 2 shows an example of function $F(r,\xi,E)$ for ξ, E coordinates of a point in a curvilinear neighborhood of Π_0 on Q where $F(r,\xi,E)<0$ only for $r_0-h\leq r\leq r_0+h$.

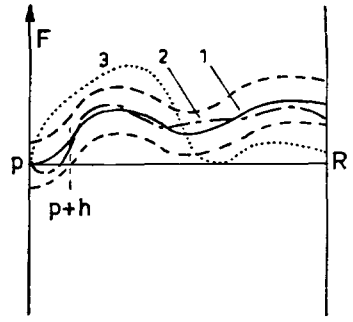


FIG. 10. Example of function $F(r,\xi,E)$ for $\xi=\xi_0$, $E=E_0$ coordinates of a point $\Pi_0\in P_p\cap Q$ (curves 1 and 3). Curve 2 shows an example of function $F(r,\xi,E)$ for ξ, E coordinates of a point in a curvilinear neighborhood of Π_0 on P_p where $F(r,\xi,E)<0$ only for $p\leq r\leq p+h$.

VII. LEMMA 5

Statement: Under Assumptions (H) let Π_0 denote a point of $P_p\cup P_R$. One of the following conditions is satisfied:

$\Pi_0\notin W_{pR}$ and there is an open curvilinear neighborhood of Π_0 on P_p or P_R out of W_{pR}; or

$\Pi_0\in\Gamma_{pR}$ and there is an open curvilinear neighborhood of Π_0 on P_p or P_R which is an arc of Γ_{pR}; or

$\Pi_0\in Q$; or

$\Pi_0\in P_p\cap P_R$.

Proof: The arguments to be used are very similar to those developed in the immediately preceding section. Thus we do not go into all the details.

If $\Pi_0\notin W_{pR}$, using the uniform ξ,E continuity of $F(r,\xi,E)$ over $[p,R]$ as in Sec. VI, we may prove that there is an open curvilinear neighborhood of Π_0 on P_p or P_R out of W_{pR}.

If $\Pi_0\in W_{pR}$, since $\Pi_0\in P_p$ or $\Pi_0\in P_R$, any open neighborhood of Π_0 in the ξE plane contains some points of the exterior of W_{pp} or W_{RR} and so $\Pi_0\in\Gamma_{pR}$.

Then assume that $\Pi_0\in P_p$, $\Pi_0\notin P_R$ and that in any curvilinear neighborhood θ_c of Π_0 on P_p there is at least one point Π_1 with coordinates ξ_1, E_1 for which $F(r,\xi_1,E_1)<0$ for some r different from p. Since $F(r,\xi,E)$ is uniformly ξ,E continuous over $[p,R]$ (Sec. II) $F(r,\xi_1,E_1)$ can be negative only in neighborhoods of r for which $F(r,\xi_0,E_0)=0$. Since $\Pi_0\in\Gamma_{pR}$, condition (10) is satisfied; it follows that two cases can occur:

For any h there is a θ_c with a point Π_1 in θ_c for which $F(r,\xi_1,E_1)<0$ for $p<r<p+h$ so that $F'_r(p,\xi_1,E_1)<0$. Since $F'_r(r,\xi,E)$ is continuous $F'_r(p,\xi_0,E_0)=0$ and $\Pi_0\in Q$ (for instance see Fig. 10, curve 1); or

For any h there is a θ_c with a point Π_1 in θ_c for which $F(r,\xi_1,E_1)<0$ for $r_r-h<r<r_r+h$, where r_r denotes some r for which $r_r\neq p$, $F(r_r,\xi_0,E_0)=0$. Then from condition (10) it follows that $F'_r(r_r,\xi_0,E_0)=0$ and $\Pi_0\in Q$ (for instance see Fig. 10, curve 3).

Of course the case where the assumption $\Pi_0\in P_p$, $\Pi_0\notin P_R$ is replaced by $\Pi_0\notin P_p$, $\Pi_0\in P_R$ can be investigated in a similar way. Obvious results completing the proof of Lemma 5 are then obtained.

VIII. PROOF OF THEOREM 2 (STATED IN SEC. II)

First assume that $\mathring{W}_{pR}\neq\emptyset$. From Lemmas 1—5 it follows that W_{pR} is a closed domain of the ξE plane bounded by a continuous curve Γ_{pR} made of arcs of P_p, P_R, Q and stationary curves P_a; the different arcs are limited by singular points of Q, isolated points of Q on some arcs of Q, or common points of some of the curves P_p, P_R, Q, P_{a_i}.

In order to delimit W_{pR} we shall divide the interval $[p,R]$ into a finite number of subintervals $(r_j,\ r_{j+1})$ $(j=1,2,\ldots,m;\ r_1=p,\ r_m=R)$ in which $\mathring{W}_{r_jr}\supset\mathring{W}_{pR}\neq\emptyset$ and W_{r_jr} can be delimited easily using Lemmas 3—5. Then, using relation (7) all the regions W_{pr} can be delimited for p increasing from p to R. Of course one can also determine the regions W_{Rr} for r decreasing from R to p in a similar way.

Three cases can occur:

(1) If for $r>p$ in the neighborhood of $p=r_1$ there is no arc and no isolated point of Q, let r_2 denote the lowest $r>p$ for which an arc of Q or an isolated point of Q appears. Then for $p<r\leq r_2$, W_{pr} can be delimited using Lemma 3 (Sec. V) (Fig. 2).

(2) If for $r>p$ in the neighborhood of $p=r_1$, there are only isolated points of Q (there is no stationary curve and no arc of Q), let r_2 denote the upper bound of the $r>p$ such that this assumption is valid for the whole interval $]p,r[$. From Lemmas 1—5 it follows that $W_{pr}=W_{pp}\cap W_{rr}$ for $p<r\leq r_2$ (Fig. 3).

(3) If for $r>p$ in the neighborhood of $p=r_1$ there are some arcs of Q, let r_2 denote the lowest $r>p$ for which P_r is a stationary curve or some arc of Q reaches an isolated point of Q, or has a singular point or a common point with another arc of Q (Figs. 1 and 4). Then, from Lemmas 1, 4, and 5 it follows that, for $p\leq r\leq r_2$ each subarc of these arcs of Q contained in the interior of $W_{pp}\cap W_{rr}$ divides $W_{pp}\cap W_{rr}$ into two regions, one of which contains W_{pr}; W_{pr} is the part of $W_{pp}\cap W_{rr}$ common to all these regions.

Now assume that there is some r_j such that the

regions W_{pr} have been delimited for all the $r \in [p, r_j]$. Three cases similar to those just mentioned and numbered (1), (2), (3) can occur. They can be described by replacing $p = r_1$, W_{pp}, P_p, r_2, respectively, by r_j W_{pr_j}, Γ_{pr_j}, r_{j+1} in the three precedent statements, other notations remaining unchanged. Thus there is some $r_{j+1} > r_j$ such that W_{pr} can be delimited for any $r \in [p, r_{j+1}]$.

Of course for $j > 1$, any r_j is a value of $r \leq R$ for which, if $r_j \neq R$, one of the following conditions is satisfied:

(1) r_j separates two subintervals of $[p, R]$ for r in which there are some points of Q or not; or

(2) r_j separates two subintervals of $[p, R]$ for r in which Q is made up of only isolated points or not; or

(3) for $r = r_j$, P_{r_j} is a stationary curve; or

(4) for $r = r_j$ some arc of Q has a singular point or reaches an isolated point of Q or has a common point with another arc of Q.

Since it was assumed that there is a finite number of $r \in [p, R]$ for which any of these conditions can be satisfied [Assumptions (H)], the interval $[p, R]$ can be covered by a finite number of intervals $[r_j, r_{j+1}]$.

Assume that $\mathring{W}_{pR} = \emptyset$. If $W_{pp} \neq \emptyset$, for any point Π_0 with coordinates ξ_0, E_0 in W_{pp}, $F(p, \xi_0, E_0) > 0$. Since $F(r, \xi, E)$ is continuous there is some h such that $F(r, \xi_0, E_0) > 0$ for $p \leq r < p + h$. Since $F(r, \xi, E)$ is ξ, E continuous uniformly over $[p, R]$, there is an open neighborhood θ of Π_0 where $F(r, \xi, E) > 0$ for $p \leq r < p + h$. Therefore, for $p \leq r < p + h$, $\mathring{W}_{pr} \neq \emptyset$, so that $W_{pr} \neq \emptyset$. Of course, similarly, if for some r, say λ, $\mathring{W}_{p\lambda} \neq \emptyset$, there are some $r > \lambda$ for which $\mathring{W}_{pr} \neq \emptyset$ and so $W_{pr} \neq \emptyset$.

Now let L denote the upper bound of the r such that $W_{p\alpha} \neq \emptyset$ for any $\alpha \in [p, r]$. For any $r \in [p, L[$, W_{pr} can be delimited as indicated for W_{pR} when $\mathring{W}_{pR} \neq \emptyset$. Let J denote the greatest $r_j < L$. Of course for $r \in [J, L[$, W_{pr} is a region delimited in $W_{pJ} \cap W_{rr}$ by the arcs of Q contained in its interior, if any. Thus if

$$\overset{\circ}{\widehat{W_{pJ} \cap W_{rr}}} \neq \emptyset,$$

$\mathring{W}_{pr} \neq \emptyset$. Finally from the definition of L it follows that $\mathring{W}_{pL} = \emptyset$ so that

$$\overset{\circ}{\widehat{W_{pJ} \cap W_{LL}}} = \emptyset.$$

L is the lowest r for which

$$\overset{\circ}{\widehat{W_{pJ} \cap W_{rr}}} = \emptyset.$$

IX. DISCUSSION OF ASSUMPTIONS (H)—GENERALIZATION OF THE THEOREMS

Assumptions (H) are not very restrictive. However, all of them are not necessary for Theorems 1—3 to be usable.

First $F(r, \xi, E)$ should not necessarily be defined for any ξ, E. If all the regions W_{rr} or all the regions W_{rr}^- are contained in some bounded open domain θ of the ξE plane for $p \leq r \leq R$, the results are not changed except that W_{pR}, W_{pR}^-, and N_{pR} are three bounded regions covering θ.

Similarly, $F'_\xi(r, \xi, E)$ and $F'_E(r, \xi, E)$ need not necessarily never be zero simultaneously. In fact the argumentation developed in Secs. II—VIII requires only the following less restrictive conditions:

Each equation $F(r, \xi, E) = 0$ should characterize a continuous curve P_r separating two regions of the ξE plane, one in which $F(r, \xi, E) > 0$ and another in which $F(r, \xi, E) < 0$.

When the system

$$F'_r(r, \xi, E) = 0, \quad F(r, \xi, E) = 0 \tag{13}$$

has solutions in ξ, E depending on r for r in some subintervals of $[p, R]$, these solutions should be the coordinates of the points at which the curve P_r touches an envelope Q.

Of course these conditions are satisfied if Assumption (H) (2) is satisfied. However they may also be satisfied when Assumption (H) (2) is satisfied only for ξ, E roots of Eqs. (13).

Thus Theorems 1—3 can be used under assumptions less restrictive than Assumptions (H).

It is also easy to generalize Theorems 1—3 when the function F depends on three parameters ξ, E, A in addition to the variable r. For any r, the regions W_{rr} and W_{rr}^- are three-dimensional volumes separated by a surface P_r. W_{pR} and W_{pR}^- are also three-dimensional volumes bounded by a surface Γ_{pR}. Γ_{pR} is made up of sheets which are pieces of P_p, P_R, stationary surfaces P_{r_i} and an envelope Q of the surfaces P_r which is also a surface. Assumptions (H) and Theorems 1—3 can be adapted to this case straightforwardly.

A similar generalization is also straightforwardly possible when F depends on more than three parameters in addition to the variable r.

Finally, generalized Theorems 1—3 can be applied to the investigation of the roots of N equations of the form:

$$F(x, r, \ldots; \xi, E, A, \cdots) = 0,$$

$$G(x, r, \ldots; \xi, E, A, \cdots) = 0,$$

$$H(x, r, \ldots; \xi, E, A, \cdots) = 0,$$

$$\cdot$$

$$\cdot$$

$$\cdot$$

$$\cdot$$

where $x, r, \cdots$ denote N variables and $\xi, E, A, \cdots$ the parameters. For instance, consider the system

$$F(x, r; \xi, E) = 0, \tag{14}$$

$$G(x, r; \xi, E) = 0. \tag{15}$$

In order to delimit the regions of the ξE plane in which Eqs. (14) and (15) have some or no roots in x, r with $p_1 \leq x \leq R_1$, $p_2 \leq r \leq R_2$, one can apply Theorems 1—3 to the function F, or the function G, assuming that F (resp. G) depends implicitly on one variable and two parameters through two variables which, together with the parameters satisfy the condition $G = 0$ (resp. $F = 0$).

X. EXAMPLES OF USE OF THE THEOREMS IN PHYSICS

In this section we give some examples showing how the theorems proved in Secs. II–IX can be used for solving problems in physics.

In two papers[5,6] we have investigated the behavior of cylindrical and spherical metallic probes immersed in homogeneous isotropic collisionless plasmas. We have shown that, when the surface properties of the probe are known, all problems concerning sheath structure and particles collection can be solved, at least numerically, using iteration. At each step, for each radial distance, one has to find the particles which reach this radial distance and those which do not.

The classification of the particles is performed using the two constants of the motion which characterize each orbit of a particle with a given charge in this case. For instance, the condition for a particle to travel from the external sheath boundary with radius R to some point of the sheath at a radial distance r is

$$E > \text{u.b.}_{r<\rho<R} H(\rho, \xi), \tag{16}$$

with

$$H(r, \xi) = Z\varphi(r) + \xi/2r^2, \tag{17}$$

$$E \geq 0, \quad \xi \geq 0, \quad r \geq p, \tag{18}$$

where Z, r, p, $\varphi(r)$, ξ, and E denote, respectively, the number of unit charges that the particle carries with the sign of the charge, the radial distance, the probe radius, the normalized potential at distance r, the normalized square of the angular momentum of the particle, and its normalized energy.[5] R may be finite or infinite.

This is typically a problem which can be solved using our theorem. Hereafter we do not investigate this case further, since it is discussed in detail elsewhere,[5,6] and we consider that more complex case where the plasma is no longer isotropic.

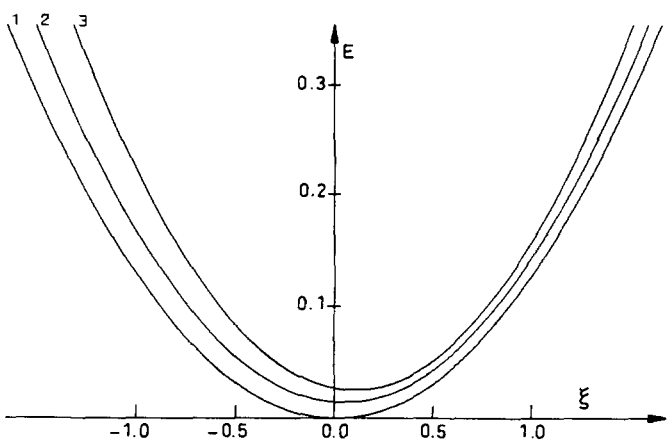


FIG. 11. Application of Theorem 3 to plasma sheath problems. Region common to all regions above the parabolas characterized by Eq. (19) for $2.4875 \leq r \leq 2.8$; $\beta(r) = 19.6 - 2.5r^2$; $\varphi(r)$ is given by the curve shown on Fig. 14. Curves 1,2,3 correspond respectively to $r = 2.8$, 2.675, 2.4875. Conditions 1 of Theorem 3 are satisfied.

For a cylindrical probe immersed in a plasma with an imposed magnetic field parallel to the axis, the problem is similar but, Eq. (17) must be replaced by

$$H(r, \xi) = Z\varphi(r) + \left(\frac{\xi - \beta(r)}{r}\right)^2 \tag{19}$$

with the same notation previously used except ξ; in (19) ξ denotes a constant of motion different from the angular momentum (which now is not constant); $\beta(r)$ is a linear function of $\int_r^R B(r) r\, dr$, where $B(r)$ denotes the magnetic field, everywhere parallel to the probe axis, at radial distance r.[7] Equation (16) characterizes the points of the ξE plane which lie above all parabolas of equation $E = Z\varphi(\rho) + [(\xi - \beta(\rho))/\rho]^2$ for $r < \rho < R$.

Figs. 11–13 illustrate the solution of Eq. (16) for decreasing values of r, in the case where this equation governs the motion of the ions through the sheath surrounding a cylindrical probe immersed in an initially neutral Maxwellian plasma with an imposed magnetic field parallel to the axis of the cylinder. The sheath model is that described by the author in a previous paper[7] [$p = 0.3$, $R = 2.8$ in units of the common Debye lengths of the plasma particles; $\beta(r) = 19.6 - 2.5r^2$]; the electric potential is that shown on Fig. 14; it illustrates the topology of a model with an overestimated sheath radius [this radius is determined by successive trials as indicated in Ref. 7 (Sec. II)].

Figure 11 illustrates the case where conditions 1 of Theorem 3 are satisfied; it corresponds to $2.4875 \leq r \leq 2.8$.

Figure 12 illustrates the next interval of r ($1.7375 \leq r \leq 2.4875$), in which conditions 3 are satisfied. For any r in this interval there is an arc of Q in $W_{2.8,2.4875} \cap W_{rr}$. Figure 13 illustrates these two intervals and also what happens for lower r: Condition 3 are satisfied but there is no arc of Q in $W_{2.8,1.7375} \cap W_{rr}$.

Let us go a little more into details concerning the use

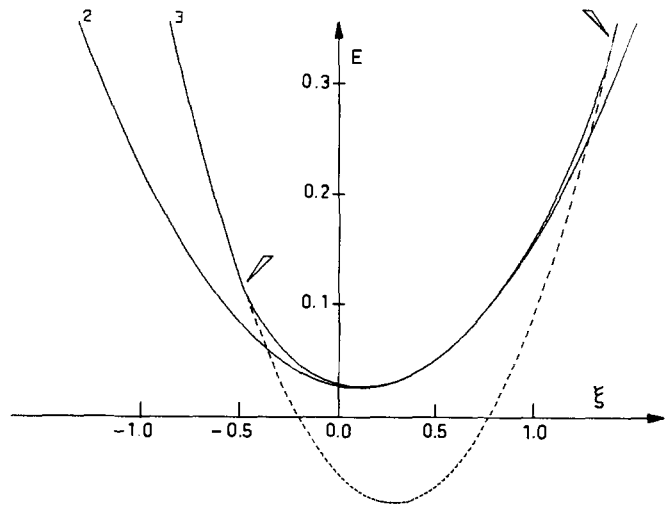


FIG. 12. Application of Theorem 3 to plasma sheath problems. Region common to all regions above the parabolas characterized by Eq. (19) for $1.7375 \leq r \leq 2.4875$; $\varphi(r)$ is given by the curve shown on Fig. 14. Curves 2 and 3 correspond respectively to $r = 2.4875$ and 1.7375. The arrows show δ_r^- and δ_r^+. Conditions 3 of Theorem 3 are satisfied. There are two arcs of Q in $W_{2.8,2.4875} \cap W_{rr}$.

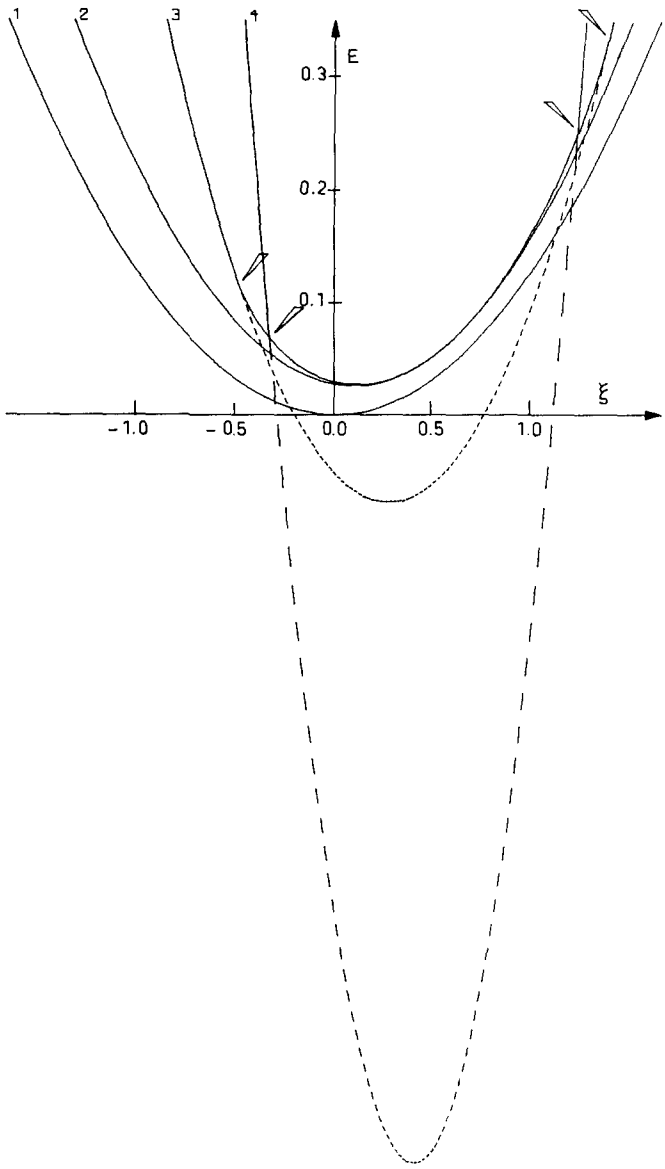


FIG. 13. Application of Theorem 3 to plasma sheath problems. Region common to all regions above the parabolas characterized by Eq. (19) for $0.8625 \leq r \leq 2.4875$; $\varphi(r)$ is given by the curve shown on Fig. 14. Curves 1, 2, 3, 4 correspond respectively to $r = 2.8$, 2.4875, 1.7375, 0.8625. The arrows show the successive positions of δ_r^- and δ_r^+. Conditions 1 and 3 with or without the arc of Q in $W_{P_r P_j} \cap W_{rr}$ are satisfied successively.

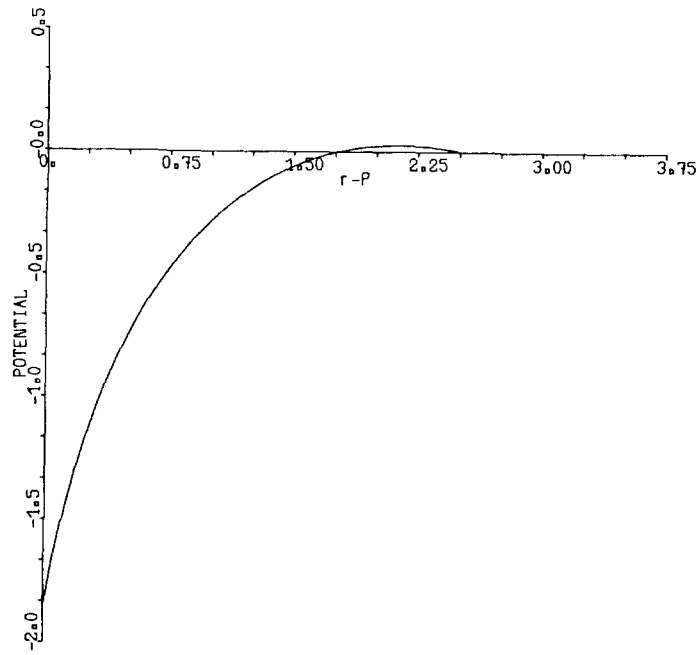


FIG. 14. Function $\varphi(r)$ for Figs. 11—13.

of Theorems 1—3 for the numerical investigation of the general case where the curves P_r are the parabolas characterized by Eq. (19). First note that if $B(r)$ and $\varphi(r)$ are bounded and twice continuously differentiable functions for $p \leq r \leq R$, Assumptions (H) are satisfied; moreover, in this case, for two arbitrary radial distances r_1 and r_2, the corresponding parabolas always have no common point or two common points with finite coordinates $\xi_\alpha(r_1, r_2)$, $E_\alpha(r_1, r_2)$, $\xi_\beta(r_1, r_2)$, and $E_\beta(r_1, r_2)$ $[\xi_\alpha(r_1, r_2) \leq \xi_\beta(r_1, r_2)]$. It is easy to prove that the sum and the product of $\xi_\alpha(r_1, r_2)$ and $\xi_\beta(r_1, r_2)$ have an upper and a lower bound for $p \leq r_1 \leq r_2 \leq R$; it follows that the $\xi_\alpha(r_1, r_2)$ and the $\xi_\beta(r_1, r_2)$ corresponding to r_1 and r_2 for which $p \leq r_1 < r_2 \leq R$ have an upper and lower bound.

Then, for any r and ρ such that $p \leq r < \rho \leq R$, either P_r and P_ρ have no common point and $W_{rr} \subset W_{\rho\rho}$ or P_r and P_ρ have two common points with abscissas $\xi_\alpha(r, \rho)$ and $\xi_\beta(r, \rho)$ and the points of P_r for which $E > H(\rho, \xi)$ are only those with abscissas out of the interval $[\xi_\alpha(r, \rho), \xi_\beta(r, \rho)]$. Thus, either, for all ρ between r and R, P_r and P_ρ have no common point and $W_{Rr} = W_{rr}$ ($\Gamma_{Rr} = P_r$); or there are some intervals $I_1, I_2, \cdots$ [enclosed in the interval (r, R)] for ρ in which P_r and P_ρ have common points and the points of P_r for which $E > H(\rho, \xi)$ for all ρ for which $r < \rho \leq R$ are only those with abscissas such that

$$\xi < \xi_r^-, \quad \text{or} \quad \xi > \xi_r^+, \tag{20}$$

where

$$\xi_r^- = \text{lower bound}_{r \in I_1 \cup I_2 \cup \ldots} \xi_\alpha(r, \rho)$$
$$\xi_r^+ = \text{upper bound}_{r \in I_1 \cup I_2 \cup \ldots} \xi_\beta(r, \rho). \tag{21}$$

Of course ξ_r^- and ξ_r^+ are always finite and between the absolute bounds of $\xi_\alpha(r, \rho)$ and of $\xi_\beta(r, \rho)$ for $p \leq r \leq \rho \leq R$. Moreover, ξ_r^- and ξ_r^+ are continuous functions of r.

Finally, for $\xi \leq \xi_r^-$ and $\xi \geq \xi_r^+$, W_{Rr} is bounded by P_r whereas, for $\xi_r^- \leq \xi \leq \xi_r^+$, it is bounded by arcs of other curves. In the sequel δ_r^- and δ_r^+ will denote, respectively, the points of P_r with abscissas ξ_r^- and ξ_r^+ when these points exist.

If, when r decreases from r_i to some lower value ρ, ξ_r^- increases and ξ_r^+ decreases, from (20) and (21) it follows that, for each r, Γ_{Rr} is made up of arcs of P_r for $\xi \leq \xi_r^-$ and $\xi_r^+ \leq \xi$ and of an arc of Γ_{Rr_i} for $\xi_r^- \leq \xi \leq \xi_r^+$.

By contrast, if ξ_r^- decreases and ξ_r^+ increases, of course for $\xi_{r_i}^- \leq \xi \leq \xi_{r_i}^+$, Γ_{Rr} is identical to Γ_{Rr_i}; this is a straightforward consequence of (20) and (21). However, the arcs of Γ_{Rr} corresponding to $\xi_r^- \leq \xi \leq \xi_{r_i}^-$ and $\xi_{r_i}^+ \leq \xi \leq \xi_r^+$ are different from those of Γ_{Rr_i} for

the same ξ; otherwise there would be points of $\Gamma_{R r_i}$ both in $W_{R r}$ and out of $W_{R\rho}$ for some $\rho \in]r, r_i[$. Then from Theorem 3 it follows that for $\xi_r^- \leq \xi \leq \xi_{r_i}^-$ and $\xi_{r_i}^+ \leq \xi \leq \xi_r^+$, $\Gamma_{R r}$ is made up of arcs of the envelope Q and these arcs are generated by δ_r^- and δ_r^+, which are the points of contact of P_r with Q, when r varies.

The cases where ξ_r^- and ξ_r^+ increase or decrease simultaneously are more complicated but can be investigated in a similar obvious way.

In any case these results can be used in a very simple way for progressive numerical delimitation of the regions $W_{R r}$, for r decreasing from R to p. For instance, ξ_r^- increases or ξ_r^+ decreases when r decreases if, for any r, δ_r^- is in the regions $\mathring{W}_{\rho\rho}$ corresponding to $\rho < r$.

XI. CONCLUSION

We have given a powerful method for finding parameters ξ, E for which equations of the form $F(r, \xi, E) = 0$ have some or no roots in given intervals, for a large class of functions F. This method is based on very general theorems with rather elementary and easily generalizable hypotheses (Secs. II and IX), and so it is suitable for solving a large class of problems. We have given examples of how it can be used in physics to solve some *a priori* very complicated problems rather easily and with the minimum of numerical operations (Sec. X). Of course, problems of other kinds than those mentioned in Sec. X can be solved with this method. Among them let us mention those concerning the photoelectron sheath around spatial probes[8] or the structure of flat galaxies.[9] There are also plasma sheath problems in plane symmetry. For instance, when there is a magnetic field $B(r)$ with fixed direction parallel to the planes where the electric potential $\varphi(r)$ and the particle densities are constant, the equations governing the motion of particles carrying Z unit electric charges ($Z \gtrless 0$) are of the form (16) with

$$H(r, \xi) = Z\varphi(r) + [\xi - \beta(r)]^2. \tag{22}$$

In (22) $\beta(r)$ is a linear function of $\int B(r)\,dr$ and ξ is a constant of the motion. The curves P_r are parabolas, two of which have in general one common point. These problems and others are introduced in our thesis[10] and will be discussed in further papers.

ACKNOWLEDGMENTS

I would like to express my gratitude to Dr. A. Mangeney, Dr. R. Hakim, Dr. J.L. Steinberg, and Dr. H. Weil for their useful advice, discussion, and criticism.

[1] I. B. Bernstein and I. N. Rabinowitz, Phys. Fluids **2**, 112 (1959).
[2] R. Buckley, Proc. R. Soc. London A **290**, 186 (1966).
[3] J. Laframboise, "Theory of spherical and cylindrical Langmuir probes in a collisionless Maxwellian plasma at rest," Report No. 100, University of Toronto, Canada, UTIAS 100 (1966).
[4] J. Lemaire and M. Scherer, Phys. Fluids **14**, 1683 (1971).
[5] J.-P.J. Lafon, Plasma Phys. **17**, 731, 1175 (1975).
[6] J.-P.J. Lafon, Plasma Phys. **17**, 741 (1975).
[7] J.-P.J. Lafon, J. Plasma Phys. **10**, 383 (1973).
[8] J.-P.J. Lafon, Radio Sci. **11**, (5), 483 (1976).
[9] J.-P.J. Lafon, Astron. Astrophys. **46**, 461 (1976).
[10] J.-P.J. Lafon, "Etude d'une classe de systèmes obéissant à un ensemble continu de conditions dan un ensemble continu d'états," Thèse de Doctorat d'Etat, Univ. de Paris VII, Rapport DESPA159bis, Observatoire de Meudon, 92190-Meudon (France).

Kerr's theorem and the Kerr–Schild congruences*

D. Cox

Department of Mathematics, Physics, and Astronomy, Sul Ross State University, Alpine, Texas
(Received 29 November 1976)

A simple flat spacetime derivation of the Kerr–Schild congruences is presented starting from Kerr's theorem.

It is well known that the Kerr—Schild spacetimes[1] are constructed from a null vector field in Minkowski space. This vector field is found to be geodesic and shear-free (GSF). In addition, the vector field is translationally invariant. In this paper, we present a simple derivation of the Kerr—Schild congruences starting from Kerr's theorem[2] and invoking translational invariance. We also show that the Kerr congruence is a Kerr—Schild congruence for which the translational Killing vector is timelike and, in addition, the congruence is rotationally invariant. We will approach the problem strictly from a flat spacetime viewpoint and will prove three theorems which will be stated later.

We emphasize that what follows does not show that the metric given by

$$g_{ab}=\eta_{ab}+h\xi_a\xi_b$$

satisfies the vacuum Einstein field equations. This can only be shown by the substitution of this metric into the field equations and the subsequent solution of those equations.

Before proceeding, we must discuss notation. The two references sited above use the same symbols u, v, ζ, $\bar{\zeta}$, Y, but these have slightly different meanings. The main difference arises from using different signature for the metrics. In this paper, the notation of Ref. 2 is used. To clarify the notation, we will now present the results which are needed from both papers. The metric of Minkowski space is given by

$$ds^2=2du\,dv-2d\zeta\,d\bar{\zeta}=(l_a n_b+n_a l_b-m_a\bar{m}_b-\bar{m}_a m_b)dx^a dx^b,$$

where l_a, n_a, m_a, $\bar{m}_a$ is any constant normed null tetrad. Then Kerr's theorem states that the most general analytic GSF null congruence ξ_a is given by $\xi_a=n_a$, or by

$$\xi_a=l_a+Y\bar{m}_a+\bar{Y}m_a+Y\bar{Y}n_a. \tag{1}$$

$\bar{Y}$ is a complex function of the coordinates u, v, ζ, $\bar{\zeta}$ defined implicitly by $F=0$, where

$$F=F(\bar{Y},\,X_1,\,X_2), \tag{2}$$

with

$$X_1\equiv u+\bar{Y}\zeta,\quad X_2\equiv\bar{\zeta}+\bar{Y}v.$$

Any constant null tetrad can be chosen and then, by coordinate transformations, coordinates can always be chosen such that

$$D\equiv l^a\partial_a=\partial_v,\quad \Delta\equiv n^a\partial_a=\partial_u,$$
$$\delta\equiv m^a\partial_a=-\partial_{\bar{\zeta}},\quad \delta\equiv\bar{m}^a\partial_a=-\partial_\zeta.$$

The Kerr—Schild space—times have metrices of the form

$$g_{ab}=\eta_{ab}+h\xi_a\xi_b,$$

where η_{ab} is a Minkowski metric and ξ_a is a GSF null congruence in Minkowski space of the form (1) with

$$F=\phi(\bar{Y})+(q+p\bar{Y})X_1-(c+\bar{q}\bar{Y})X_2=0.$$

The function h is given by

$$h=mP^{-3}(\rho+\bar{\rho}),$$

where ρ is the complex expansion of the congruence ξ_a and

$$P\equiv pY\bar{Y}+q\bar{Y}+\bar{q}\bar{Y}+c,$$

with m, p, c real constants and q a complex constant. ϕ is an arbitrary analytic function of $\bar{Y}$.

The Kerr congruence is that particular Kerr—Schild congruence which yields the Kerr metric.[3] It is the particular choice of $p=c=2^{-1/2}$, $q=0$, and

$$\phi=ia\bar{Y},$$

where a is a real constant.[1]

We will now state and prove the first of the three theorems.

Theorem I: Let $T^a=pl^a-\bar{q}\bar{m}^a-qm^a+cn^a$, with p and c real constants and q a complex constant, be a translational Killing vector in Minkowski space. The most general analytic GSF null congruence ξ_a which is invariant under dragging along by T^a is given by $\xi_a=n_a$, or by ξ_a of the form (1) with

$$F=\phi(\bar{Y})+(q+p\bar{Y})X_1-(c+\bar{q}\bar{Y})X_2. \tag{3}$$

ϕ is an arbitrary analytic function of Y.

We note, in passing, that P defined above is $T\cdot\xi$.

We demand that ξ_a be invariant under dragging along by T^a. This implies that the Lie derivative of ξ_a with respect to T^a is zero; that is,

$$\underset{T}{\pounds}\,\xi_a=\xi_{a,b}T^b+\xi_b T^b{}_{,a}=0. \tag{4}$$

Since T^a is a constant vector; $T^b{}_{,a}\equiv 0$. If $\xi_a=n_a$, we see that the condition is trivially satisfied. In the more general case we find

$$\begin{aligned}\underset{T}{\pounds}\,\xi_a&=\xi_{a,b}T^b\\&=Y_{,b}T^b\bar{m}_a+\bar{Y}_{,b}T^b m_a+(Y\bar{Y})_{,b}T^b n_a\\&=0.\end{aligned}$$

Since $\bar{m}_a$, m_a, and n_a are independent vectors, we see that

$$\underset{T}{\pounds}\,\xi_a=0 \text{ if and only if } TY\equiv T^bY_{,b}=0.$$

The directional derivative T is given by

$$T=p\partial_\nu+q\partial_\xi+\bar{q}\partial_\zeta+c\partial_u.$$

We will need the two derivatives

$$TX_1=c+\bar{q}\bar{Y}+\xi T\bar{Y}, \tag{5a}$$

$$TX_2=(q+p\bar{Y})+\nu T\bar{Y}. \tag{5b}$$

In Ref. 2, it was shown that the congruence is GSF provided

$$(\partial_\nu-\bar{Y}\partial_\xi)\bar{Y}=0, \tag{6a}$$

$$(\partial_\zeta-\bar{Y}\partial_u)\bar{Y}=0. \tag{6b}$$

We find that the congruence is translationally invariant if

$$(p\partial_\nu+q\partial_\xi+\bar{q}\partial_\zeta+c\partial_u)\bar{Y}=0. \tag{6c}$$

Now we consider the system of equations

$$(\partial_\nu-\bar{Y}\partial_\xi)X=0, \tag{7a}$$

$$(\partial_\zeta-\bar{Y}\partial_u)X=0, \tag{7b}$$

$$(p\partial_\nu+q\partial_\xi+\bar{q}\partial_\zeta+c\partial_u)X=0, \tag{7c}$$

where for the moment $\bar{Y}$ is considered to be known. This is a system of linear partial differential equations. A necessary and sufficient condition for this system to be completely integrable, i.e., for a solution to exist, is given by Eisenhart (Ref. 4, pp. 69–70). This integrability condition is simply that the commutators of the linear operators $(\partial_\nu-\bar{Y}\partial_\xi)$, $(\partial_\zeta-\bar{Y}\partial_u)$, and $(p\partial_\nu+q\partial_\xi+\bar{q}\partial_\zeta+c\partial_u)$ should be linear combinations of these operators. When these commutators are worked out, the condition is satisfied provided $\bar{Y}$ satisfies (6).

We now assume $\bar{Y}$ is a solution of (6). Then, Eqs. (7) are integrable. From the derivatives (5) we see that a solution is

$$(q+p\bar{Y})X_1-(c+\bar{q}\bar{Y})X_2.$$

The most general (analytic) solution is then

$$X=f((q+p\bar{Y})X_1-(c+\bar{q}\bar{Y})X_2),$$

where f is an arbitrary analytic function of its argument.

In particular, the solution $\bar{Y}$ must be of this form:

$$\bar{Y}=g((q+p\bar{Y})X_1-(c+\bar{q}\bar{Y})X_2) \iff \bar{Y}-g((q+p\bar{Y})X_1-(c+\bar{q}\bar{Y})X_2)=0.$$

This can also be written in the form

$$F=\phi(\bar{Y})+(q+p\bar{Y})X_1-(c+\bar{q}\bar{Y})X_2=0,$$

where $\phi=-g^{-1}$. Thus, Theorem I is proven. We also see that the Kerr–Schild congruences are those GSF congruences which are translationally invariant as previously stated.

Before we state the second theorem, we recall the definition of the null coordinate system in terms of the usual Cartesian coordinate system,

$$u=2^{-1/2}(t-z),\quad \nu=2^{-1/2}(t+z),\quad \zeta=2^{-1/2}(x+iy).$$

Let R^a be a space-like rotational Killing vector. Since we are free to choose any null tetrad and the associated coordinates, coordinates can always be chosen such that R^a points along the z axis. The directional derivative $R^a\partial_a$ is then well known and given by

$$R\equiv R^a\partial_a=x\partial_y-y\partial_x.$$

It is quite easy to show from the above transformation and its inverse that in the null coordinates, this is

$$R=i(\zeta\partial_\zeta-\bar{\zeta}\partial_{\bar{\xi}}).$$

The vector R^a is given by

$$R^a=i(\bar{\zeta}m^a-\zeta\bar{m}^a).$$

Theorem II: Let $R^a=i(\bar{\zeta}m^a-\zeta\bar{m}^a)$ be a spacelike rotational Killing vector in Minkowski space. The most general analytic GSF null congruence ξ_a which is invariant under dragging along by R^a is given by $\xi_a=n_a$, or by ξ_a of the form (1) with

$$F=\alpha(X_1)\bar{Y}+\beta(X_1)X_2, \tag{8}$$

where α and β are arbitrary analytic functions of X_1.

We demand that ξ_a be invariant under dragging along by R^a. This implies that the Lie derivative of ξ_a with respect to R^a is zero; that is,

$$\underset{R}{\pounds}\,\xi_a=\xi_{a,b}R^b+\xi_bR^b{}_{,a}=0.$$

The case $\xi_a=n_a$ is identically satisfied. The more general case gives

$$\underset{R}{\pounds}\,\xi_a=(RY-iY)\bar{m}_a+(R\bar{Y}+i\bar{Y})m_a+R(Y\bar{Y})n_a=0,$$

using the fact that

$$\bar{\zeta}_{,a}=\bar{m}_a.$$

Since $\bar{m}_a$, m_a, n_a are independent

$$RY-iY=0, \tag{9a}$$

$$R\bar{Y}+i\bar{Y}=0, \tag{9b}$$

$$R(Y\bar{Y})=0. \tag{9c}$$

(9a) and (9c) follow from (9b) so we have

$\underset{R}{\pounds}\,\xi_a=0$ if and only if $R\bar{Y}=-i\bar{Y}$.

We will need the derivatives

$$RX_1=i\bar{Y}\zeta+\zeta R\bar{Y}, \tag{10a}$$

$$RX_2=-i\bar{\zeta}+\nu R\bar{Y}. \tag{10b}$$

In order for a GSF null congruence to be rotationally invariant then $\bar{Y}$ must satisfy the system of equations:

$$(\partial_\nu-\bar{Y}\partial_\xi)\bar{Y}=0, \tag{11a}$$

$$(\partial_\zeta-\bar{Y}\partial_u)\bar{Y}=0, \tag{11b}$$

$$(\zeta\partial_\zeta-\bar{\zeta}\partial_{\bar{\xi}}+1)\bar{Y}=0. \tag{11c}$$

Thus, we are lead to consider the system of equations:

$$(\partial_\nu-\bar{Y}\partial_\xi)X=0, \tag{12a}$$

$$(\partial_\zeta-\bar{Y}\partial_u)X=0, \tag{12b}$$

$$(\zeta\partial_\zeta-\bar{\zeta}\partial_{\bar{\xi}}+1)X=0, \tag{12c}$$

where for the moment we consider $\bar{Y}$ to be known. This system is found to be completely integrable provided $\bar{Y}$ satisfies (11).

We now assume $\bar{Y}$ is a solution of (11). Then Eqs. (12) are integrable. Using (9b) in (10) we find

$$RX_1 = 0,$$

$$RX_2 = -iX_2.$$

The most general solution is therefore

$$X = f(X_1)X_2.$$

where f is an arbitrary analytic function of its argument.

In particular, the solution $\bar{Y}$ must be of the form:

$$\bar{Y} = g(X_1)X_2 \iff \bar{Y} - g(X_1)X_2 = 0.$$

This can also be written in the form:

$$F = \alpha(X_1)\bar{Y} + \beta(X_1)X_2 = 0,$$

where α and β are arbitrary analytic functions of X_1. Thus, Theorem II is proven.

The last theorem to be proved follows directly from the first two.

Theorem III: The most general analytic GSF null congruence ξ_a which satisfies Theorems I and II is given by $\xi_a = n_a$, or by ξ_a of the form (1) with

$$F = ia\bar{Y} + p\bar{Y}X_1 - cX_2,$$

where a is a real constant.

If $\bar{Y}$ is to satisfy both theorems, then the commutator of the two operators R and T must be a linear combination of R, T, $(\partial_\nu - \bar{Y}\partial_\zeta)$, and $(\partial_{\bar{\zeta}} - \bar{y}\partial_u)$. Working out this commutator, we find

$$[T, R]\psi = i[\bar{q}\partial_\zeta - q\partial_{\bar{\zeta}}]\psi.$$

This is of the desired form provided $q = 0$. If we now compare (3), with $q = 0$, with (8) we see that they are compatible provided $\beta = -c$, $\phi = A\bar{y}$ and $\alpha = pX_1 + A$, where A is a complex constant. Thus,

$$F = A\bar{Y} + p\bar{Y}X_1 - cX_2.$$

If we examine the terms linear in $\bar{Y}$ in this expression we find they are

$$[A + pu - c\nu]\bar{Y}.$$

Thus, the real part of A can be eliminated by a translation of the origin of either u or ν or both. Therefore, F can always be written in the form

$$F = ia\bar{Y} + p\bar{Y}X_1 - cX_2,$$

proving Theorem III.

If the vector T^a is timelike, then a tetrad can be chosen such that

$$T^a = 2^{-1/2}(l^a + n^a)$$

or $p = c = 2^{-1/2}$, proving the statement about the Kerr congruence.

ACKNOWLEDGMENTS

This work was completed while the author was at the Relativity Center, University of Texas at Austin, and was presented in a seminar there in the Spring of 1975. I wish to thank Ed Flaherty and Paul Sommers for many helpful discussions.

*Supported in part by NSF Grant GP-43844-X.

[1]G.C. Debney, R.P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
[2]D. Cox and E.J. Flaherty, Commun. Math. Phys. **47**, 75 (1976).
[3]R.P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).
[4]L.P. Eisenhart, *Riemannian Geometry* (Princeton University, Princeton, N.J., 1966).

Composite next nearest neighbor degeneracy

Richmond B. McQuistan

Department of Physics and Laboratory for Surface Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201
(Received 8 October 1976)

Expressions are derived which yield, exactly, the composite degeneracy of those arrangements of simple, indistinguishable particles on a one-dimensional lattice space, which are characterized by the number of nearest and next nearest neighbor pairs.

I. INTRODUCTION

For one-dimensional systems in which the next nearest neighbor approximation is assumed, the most general expression for E_i, the interaction energy is written (see Fig. 1)

$$E_i = n_{11}V_{11} + n_{01}V_{01} + n_{00}V_{00} + n_{111}V_{111} + n_{110}V_{110} + n_{101}V_{101} + n_{000}V_{000} + n_{001}V_{001} + n_{010}V_{010}, \quad (1)$$

where n_{11}, n_{01}, and n_{00} are, respectively, the numbers of occupied, mixed, and vacant nearest neighbor pairs on the lattice space; n_{111}, n_{110}, n_{101}, n_{000}, n_{001}, and n_{010} are the numbers of the various possible kinds of next nearest neighbor pairs; and V_{11}, V_{01}, V_{11}, V_{111}, V_{110}, V_{101}, V_{000}, V_{001}, and V_{010} are the corresponding potential energies of interaction.

This expression, Eq. (1), for the interaction energy does not preclude the special situations in which $V_{111} = V_{101}$, $V_{000} = V_{010}$, and $V_{001} = V_{011}$, i.e., in which the interaction between two next nearest neighbor pairs is independent of the state of occupation of the intervening site.

In the present paper we develop expressions which describe, for simple, indistinguishable particles distributed on a one-dimensional lattice space which consists of equivalent sites, the degeneracy of arrangements containing a specified number of occupied, mixed, and vacant nearest neighbor pairs *and* of next nearest neighbor pairs of the 111, 110, 101, 000, 001, and 010 types. It should be clear that the subscript 1 refers to occupation and 0 to vacancy.

We discuss the respective degeneracies in terms of the vacancy and occupation of lattice sites on the space. Clearly, the results are valid for any binary variable such as spin or $A-B$ atoms in a binary alloy for which a one-dimension lattice is appropriate.

Previous papers[1–4] concerned with the degeneracy for one-dimensional lattice spaces, have treated special cases in which only one kind of nearest neighbor pair (i.e., either occupied, mixed, or vacant) is involved. Recently, another article[5] had dealt with the composite degeneracy for all three kinds of nearest neighbor pairs 11, 01, and 00. Specifically, the present paper extends these results to include next nearest neighbor pairs of all possible kinds.

We note, first, that the quantities represented by n_{11}, n_{01}, and n_{00} are not independent but are related by the fact that on any particular arrangement there are a total of $N-1$ nearest neighbor pairs of all kinds, i.e.,

$$N-1 = n_{11} + n_{01} + n_{00}. \quad (2)$$

Similary, the next nearest neighbor pairs represented by n_{111}, n_{110}, n_{010}, n_{000}, n_{001}, and n_{010} are not independent but are subject to the constraint that their sum is the total number of next nearest neighbor pairs on the array, i.e.,

$$N-2 = n_{111} + n_{110} + n_{101} + n_{000} + n_{001} + n_{010}. \quad (3)$$

In addition to these relationships, there are other equations relating members of the sets $\{n_{ij}\}$ and $\{n_{ijk}\}$ to each other and to the particles and N lattice sites. We derive such relationships in the following section.

II. RELATIONSHIPS AMONG $\{n_{ij}\}$, $\{n_{ijk}\}$, q AND N

For purposes of the following discussion we consider the sites of the lattice space to be numbered consecutively from left to right, $1, 2, \ldots, N-1, N$. The state of occupation of sites numbered 1, 2, $N-1$, and N determine whether n_{01}, n_{110}, and n_{001} are even or odd and thereby specify the form of the relationships we seek.

The ten possible arrangements involving the state of occupation of sites 1, 2, $N-1$, and N are shown in Fig. 2. The state of occupation of the remaining $N-4$ sites is not germane to the discussion in this section.

(8) Sites 1, 2, N_{01}, and N are all vacant [see Fig. 2(a)]

In the case depicted in Fig. 2(a) each individual, isolated particle and each group of contiguous particles are surrounded on both the right and left by at least one vacancy. Consequently, 01-type nearest neighbor pairs and 110-type next nearest neighbor pairs appear in sets of two. Except for the two vacancies or group of contiguous vacancies incorporating sites 1, 2, and $N-1$, N, each individual vacancy and each group of contiguous vacancies are surrounded on the right *and* left by at least one particle. Consequently, 001-type next nearest

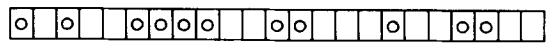


FIG. 1. In this figure N, the number of sites is 23 and q, the number of particles, is 11. The number of the various kinds of nearest and next nearest neighbor pairs is:

$n_{11}=5$,	$n_{111}=2$,	$n_{000}=1$,
$n_{01}=11$,	$n_{110}=6$,	$n_{001}=9$,
$n_{00}=6$,	$n_{101}=1$,	$n_{010}=2$.

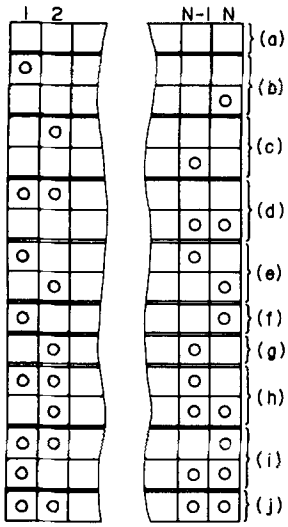


FIG. 2. In this figure we show the ten possible arrangements involving the state of occupation of sites 1, 2, N-1, and N. In those cases, e.g., (g), where reflective symmetry can arise we have shown it as a single case.

neighbor pairs appear either in sets of two or in two single pairs, one at each end of the space. Thus we may write

$$2q = n_{01} + 2n_{11}, \tag{4}$$

$$2(N-q) - 2 = n_{01} + 2n_{00}, \tag{5}$$

$$2n_{11} = n_{110} + 2n_{111}, \tag{6}$$

$$2n_{00} - 2 = n_{001} + 2n_{000}. \tag{7}$$

Equation (4) can be derived by drawing, whenever possible, from each occupied site, two lines, one to the right and one to the left (see Fig. 3) across the separations between the sites. Thus, a single line across a separation represents a mixed nearest neighbor pair and a double line signifies an occupied nearest neighbor pair. An accounting of the lines leads to Eq. (4).

If, in a manner similar to that outlined above, lines are drawn, when possible, from a vacant site to its adjacent sites, then there will be a total of $2(N-q)-2$ lines. In this case the occurrence of a double line across a separation indicates a vacant nearest neighbor pair, and a single line represents a mixed nearest neighbor pair. An accounting of the lines leads to Eq. (5).

If, from each pair of occupied nearest neighbor sites, lines are drawn, when possible, to adjacent sites there will be $2n_{11}$ lines; $2n_{111}$ of these lines are associated with next nearest neighbor pairs of the 111-type and n_{110} lines are representative of 110-type next nearest neighbor pairs. Eq. (6) follows. A similar construction for vacant nearest neighbor pairs leads to Eq. (7).

An examination of Eqs. (4)—(7) indicates that n_{01} is even, as are n_{110} and n_{001}.

(b) Site 1 is occupied and sites 2, $N-1$, and N are vacant [see Fig. 2(b)].

n_{01} is odd in this case because each occupied site or group of contiguous occupied sites (except the one involving site 1) has two mixed nearest neighbor pairs associated with it; site 1 has a single mixed nearest neighbor pair associated with it.

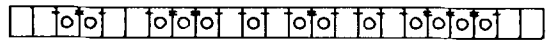


FIG. 3. This figure is used in deriving Eq. (4). A line is drawn from an occupied site to its adjacent nearest neighbor sites. In the case where a particle resides in site 1 and/or N, a single line is drawn from the occupied site to the adjacent site.
$N=23$, $q=13$ and $n_{11}=7$ (double lines), $n_{01}=12$ (single lines), $n_{00}=3$.

Because all occupied nearest neighbor pairs or contiguous groups of occupied nearest neighbor pairs are surrounded by at least a single vacancy at each end, both on the right and on the left, there are always an even number of next nearest neighbor pairs of the 110 type.

n_{001} is always odd because all vacant nearest neighbor pairs occur in sets of two except the one that is farthest to either end of the space (when site 1 or N is occupied).

We conclude that n_{01} and n_{001} are odd and n_{110} is even. By constructions similar to those performed in connection with Eqs. (4)—(7) we may write

$$2q - 1 = n_{01} + 2n_{11}, \tag{8}$$

$$2(N-q) - 1 = n_{01} + 2n_{00}, \tag{9}$$

$$2n_{11} = n_{110} + 2n_{111}, \tag{10}$$

$$2n_{00} - 1 = n_{001} + 2n_{000}. \tag{11}$$

(c) Sites 1, $N-1$, and N are vacant and site 2 is occupied [see Fig. 2(c)].

In this case, by means of the reasoning outlined above we conclude that n_{01} is even, n_{110} is even, and n_{001} is odd. Then the relationships we seek are

$$2q = n_{01} + 2n_{11}, \tag{12}$$

$$2(N-q) - 2 = n_{01} + 2n_{00}, \tag{13}$$

$$2n_{11} = n_{110} + 2n_{111}, \tag{14}$$

$$2n_{00} - 1 = n_{001} + 2n_{000}. \tag{15}$$

(d) Sites 1 and 2 are occupied and sites $N-1$ and N are vacant [see Fig. 2(d)].

n_{01}, n_{110}, and n_{001} are all odd and we write

$$2q - 1 = n_{01} + 2n_{11}, \tag{16}$$

$$2(N-q) - 1 = n_{01} + 2n_{00}, \tag{17}$$

$$2n_{11} - 1 = n_{110} + 2n_{111}, \tag{18}$$

$$2n_{00} - 1 = n_{001} + 2n_{000}. \tag{19}$$

(e) Sites 1 and $N-1$ are occupied and sites 2 and N are vacant [see Fig. 2(e)].

It follows that

$$2q - 1 = n_{01} + 2n_{11}, \tag{20}$$

$$2(N-q)-1=n_{01}+2n_{00}, \tag{21}$$

$$2n_{11}=n_{110}+2n_{111}, \tag{22}$$

$$2n_{00}=n_{001}+2n_{000}, \tag{23}$$

where n_{01} is odd and n_{110} and n_{001} are even.

(f) Sites 1 and N are occupied and sites 2 and $N-1$ are vacant [see Fig. 2(f)].

n_{01}, n_{110}, and n_{001} are all even. It follows that

$$2q-2=n_{01}+2n_{11}, \tag{24}$$

$$2(N-q)=n_{01}+2n_{00}, \tag{25}$$

$$2n_{11}=n_{110}+2n_{111}, \tag{26}$$

$$2n_{00}=n_{001}+2n_{000}. \tag{27}$$

(g) Sites 1 and N are vacant and sites 2 and $N-1$ are occupied [see Fig. 2(g)]

n_{01}, n_{110}, and n_{001} are all even. Then

$$2q=n_{01}+2n_{11}, \tag{28}$$

$$2(N-q)-2=n_{01}+2n_{00}, \tag{29}$$

$$2n_{11}=n_{110}+2n_{111}, \tag{30}$$

$$2n_{00}=n_{001}+2n_{000}. \tag{31}$$

(h) Sites 1, 2, and $N-1$ are occupied and site N is vacant [see Fig. 2(h)]

n_{01} and n_{110} are odd; n_{001} is even, so that

$$2q-1=n_{01}+2n_{11}, \tag{32}$$

$$2(N-q)-1=n_{01}+2n_{00}, \tag{33}$$

$$2n_{11}-1=n_{110}+2n_{111}, \tag{34}$$

$$2n_{00}=n_{001}+2n_{000}. \tag{35}$$

(i) Sites 1, 2, and N are occupied; site $N-1$ is vacant [see Fig. 2(i)]

n_{01} and n_{001} are even and n_{110} is odd. We can then write

$$2q-2=n_{01}+2n_{11}, \tag{36}$$

$$2(N-q)=n_{01}+2n_{00}, \tag{37}$$

$$2n_{11}-1=n_{110}+2n_{111}, \tag{38}$$

$$2n_{00}=n_{001}+2n_{000}. \tag{39}$$

(j) Sites 1, 2, $N-1$, and N are all occupied [see Fig. 2(j)]

n_{01}, n_{110}, and n_{001} are all even and

$$2q-2=n_{01}+2n_{11}, \tag{40}$$

$$2(N-q)=n_{01}+2n_{00}, \tag{41}$$

$$2n_{11}-2=n_{110}+2n_{11}, \tag{42}$$

$$2n_{00}=n_{001}+2n_{000}. \tag{43}$$

It should be noted that the selection n_{01}, even; n_{110}, odd; n_{001}, odd; cannot exist because if n_{01} is even, cells 1 and N are either both vacant or they are both occupied. If they are both vacant then n_{110} cannot be even because that would require at least one particle in site 1 or N. If n_{01} is even and sites 1 and N are occupied, n_{001} cannot be even because that would require that either site 1 or N be vacant.

The results of the foregoing section are summarized in Table I.

TABLE I. The eight (2^3) possibilities for the even or odd character of n_{01}, n_{110}, and n_{001} are shown together with the corresponding arrangement in Fig. 2.

n_{01}	n_{110}	n_{001}	Arrangements
even	even	even	(a), (f), (g), (j)
even	even	odd	(c)
even	odd	even	(i)
odd	even	even	(e)
even	odd	odd	not possible
odd	even	odd	(b)
odd	odd	even	(h)
odd	odd	odd	(d)

III. DETERMINATION OF $A[\{n_{ij}\},\{n_{ijk}\},q,N]$

To determine the analytic forms of $A[\{n_{ij}\},\{n_{ijk}\},q,N]$ we consider the 2^3 cases arising when the three quantities n_{01}, n_{110}, and n_{001} individually assume odd and even values (see Table I).

A. n_{01}, n_{110}, and n_{001} are all odd [see Case (d) and Fig. 2(d) discussed in Sec. II]

We define a "particle unit" to be either a single particle or group of contiguous occupied sites (see Fig. 4). Each "particle unit" is separated from other "particle units" by one or more vacancies. In this case, sites 1 and 2 are occupied while sites $N-1$ and N are vacant (or sites $N-1$ and N are occupied while sites 1 and 2 are vacant). Thus, there is associated with each "particle unit" two mixed nearest neighbors except for the "particle unit" at one end of the space which has only one mixed nearest neighbor associated with it. Consequently, there are $(n_{01}+1)/2$ "particle units."

Similarly, there are "vacancy units" each of which consists of either a single vacancy or group of contiguous vacancies (see Fig. 4). Each "vacancy unit" is separated from other "vacancy units" by one or more contiguous particles. There are associated with each "vacancy unit" (all but the two vacancy units on the ends

FIG. 4. There are five "particle units" (see dashed boxes) on this space which depicts Case (d) in Fig. 2. In Fig. 4 five "vacancy units" are also present.

of the array serve to separate the "particle units") mixed nearest neighbor pairs, one at each end, except for the "vacancy unit" incorporating sites $N-1$ and N (or sites 1 and 2) which is terminated at one end only by a mixed nearest neighbor pair. Thus, there are $(n_{01}+1)/2$ "particle units." Of such separations n_{01} are next nearest neighbors of the 101 type and $(n_{01}-1)/2 - n_{101}$ are separations involving two or more vacancies.

A group of vacancies constituting the separation between any two "particle units" may be exchanged with other groups of vacancies separating other pairs of "particle units," subject only to the constraint that n_{101} must be preserved. This implies that the group of contiguous vacancies may be permuted among themselves under the restriction that the number of separations of "particle units" which consist of a single vacancy, is prescribed. In this way

$$\binom{(n_{01}+1)/2}{n_{101}}$$

independent arrangements can be constructed without violating the specification of the number of the various kinds of nearest and next nearest pairs.

Each "particle unit" which consists of two or more particles will be terminated at each end by a next nearest neighbor pair of the 110 type, except for the "particle unit" containing the particles in sites 1 and 2 (or $N-1$ and N) which is terminated at one end by one next nearest neighbor pair of the 110 type and by the end of the lattice space at the other end. Thus there are $(n_{110}+1)/2$ "particle units" consisting of two or more particles. Then n_{111} next nearest neighbor pairs of the 111 type can be distributed among the "particle units" which contain two or more particles in

$$\binom{(n_{110}+1)/2+n_{111}-1}{n_{111}} = \binom{(n_{11}-1)/2+n_{111}}{n_{111}} \tag{44}$$

independent ways. Here we consider the "particle units" which contain two or more particles to represent numbered boxes among which the n_{111} nearest neighbor pairs of the 111 type may be distributed.[6]

Without affecting the two possible permutation processes outlined above, we can consider the permutation of the $(n_{01}+1)/2$ "vacancy units" (see Fig. 4) which are separated from each other by the $(n_{01}+1)/2-1$ "particle units." Such separations between the "vacancy units" can be classified into two types: the next nearest neighbor pairs of the 010 type involving a single particle between two "vacancy units" and separations consisting of two or more contiguous particles. The particle or groups of contiguous paticles between any two "vacancy units" can be exchanged with the particle or group of particles separating other "vacancy units" provided that for any particular arrangement n_{010} is conserved. This may be done in

$$\binom{(n_{01}-1)/2}{n_{010}}$$

independent ways.

$(n_{001}+1)/2$ of the "vacancy units" contain two or more vacancies. The n_{000} next nearest neighbor pairs of the 000 type can be distributed among the three "vacancy units" in

$$\binom{(n_{001}+1)/2+n_{000}-1}{n_{000}}$$

independent ways without affecting the number of 101 type next nearest neighbor pairs.

Thus, for the situation in which n_{01}, n_{110}, and n_{001} are odd, the composite degeneracy may be written

$$A[\{n_{ij}\},\{n_{ijk}\},q,N] = 2\binom{(n_{110}-1)/2+n_{111}}{n_{111}}\binom{(n_{001}-1)/2+n_{000}}{n_{000}} \times\binom{(n_{01}-1)/2}{n_{101}}\binom{(n_{01}-1)/2}{n_{010}}, \tag{45}$$

where the coefficient 2 arises because of the lack of reflective symmetry of the Case (d) arrangements.

B. n_{01} and n_{110} odd and n_{001} even [see Case (h) and Fig. 2(h)]

In this situation there are associated with each "particle unit" two mixed nearest neighbor pairs except for the "particle unit" which incorporates the two particles contained in sites 1 and 2. Consequently, there are $(n_{01}+1)/2$ "particle units."

Because each "vacancy unit" except that which contains site N if it is vacant (or site 1 if it is vacant) has a mixed nearest pair at each end, while the vacant site on the end has a single mixed nearest neighbor pair associated with it, there are $(n_{01}+1)/2$ "vacancy units." Of these "vacancy units" $(n_{01}+1)/2-1=(n_{01}-1)/2$ separate the "particle units" and n_{101} of these separations consist of a single vacancy and $(n_{01}-1)/2-n_{01}$ do not. Thus, there are

$$\binom{(n_{01}-1)/2}{n_{101}}$$

independent ways that the n_{101} next nearest neighbors of the 101 type may be permuted between the "particle units" without altering the number of 101 type pairs.

Each "particle unit" consisting of two or more particles except for the "particle unit" containing the particles in sites 1 and 2 (or $N-1$ and N if they are occupied) will have two next nearest neighbor pairs of the 110 type associated with it. Thus there are $(n_{110}+1)/2$ "particle units" which consist of two or more particles. The n_{111} next nearest neighbor pairs of the 111 type can be distributed among those "particle units" which contain two or more particles in

$$\binom{(n_{110}+1)/2+n_{111}-1}{n_{111}} = \binom{(n_{110}-1)/2+n_{111}}{n_{111}} \tag{46}$$

independent ways.

Conversely, the $(n_{01}+1)/2$ "vacancy units" are separated from each other by $(n_{01}+1)/2-1=(n_{01}-1)/2$ "particle units." As previously, such separations between the "vacancy units" can consist of a single particle or sequences of contiguous particles; each separa-

tion consisting of a single particle constitutes a next nearest neighbor pair of the 010 type. Such separations can be distributed among the $(n_{01}-1)/2$ separations in

$$\binom{(n_{01}-1)/2}{n_{010}}$$

independent ways.

$(n_{001}/2)$ of the "vacancy units" contain two or more vacancies because each of such "vacancy units" are terminated at each end by two next nearest neighbors of the 001 type. The n_{000} next nearest neighbor pairs of the 001 type can be distributed among the "vacancy units" constaining two or more vacancies in

$$\binom{n_{001}/2+n_{000}-1}{n_{000}}$$

independent ways without affecting the number of next nearest neighbor pairs of the 101 type. Thus when n_{01} and n_{110} are odd and n_{001} is even the composite degeneracy may be written

$$A[\{n_{ij}\},\{n_{ijk}\},q,N] = 2\binom{(n_{110}-1)/2+n_{111}}{n_{111}}\binom{n_{001}/2+n_{000}-1}{n_{000}} \times \binom{(n_{01}-1)/2}{n_{101}}\binom{(n_{01}-1)/2}{n_{010}}, \tag{47}$$

where the factor 2 again arises because with equal probability either sites 1 and 2 can be occupied *or* sites $N-1$ and N can be occupied in situations where n_{01} and n_{110} are odd and n_{001} is even.

C. n_{01} odd and, n_{110} and n_{001} are even [see Case (e) and Fig. 2(e)]

In this situation, sites 1 and $N-1$ are occupied and sites 2 and N are vacant (*or* sites 2 and N are occupied and sites 1 and $N-1$ are vacant). By reasoning similar to that contained in the two previous cases it can be shown that

$$A[\{n_{ij}\},\{n_{ijk}\},q,N] = 2\binom{n_{110}/2+n_{111}-1}{n_{111}}\binom{n_{001}/2+n_{000}-1}{n_{000}} \times \binom{(n_{01}-1)/2}{n_{101}}\binom{(n_{01}-1)/2}{n_{010}}. \tag{48}$$

D. n_{01} and n_{001} are odd and n_{110} is even [see Case (b) and Fig. 2(b)]

Here, site 1 is occupied and sites 2, $N-1$, and N are vacant (or site N is occupied and sites 1, 2, and $N-1$ are vacant) and the degeneracy is given by

$$A[\{n_{ij}\},\{n_{ijk}\},q,N] = \binom{2}{n_{110}/2+n_{111}-1}\binom{(n_{001}-1)/2+n_{000}}{n_{101}} \times \binom{(n_{01}-1)/2}{n_{101}}\binom{(n_{01}-1)/2}{n_{010}}. \tag{49}$$

E. n_{01} and n_{110} are even, and n_{001} is odd [see Case (c) and Fig. 2(c)]

This situation arises when site 2 is occupied and sites 1, $N-1$, and N are vacant (or site $N-1$ is occupied and sites 1, 2, and N are vacant).

$$A[\{n_{ij}\},\{n_{ijk}\},q,N] = 2\binom{n_{110}/2+n_{111}-1}{n_{111}}\binom{(n_{001}-1)/2+n_{000}}{n_{000}} \times \binom{n_{01}/2-1}{n_{101}}\binom{n_{01}/2}{n_{010}}. \tag{50}$$

F. n_{01} and n_{001} are even and n_{110} is odd [see Case (i) and Fig. 2(i)]

When sites 1, 2, and N are occupied and site $N-1$ is vacant (or when sites 1, $N-1$, and N are occupied and site 2 is vacant) we obtain

$$A[\{n_{ij}\},\{n_{ijk}\},q,N] = 2\binom{(n_{110}-1)/2+n_{111}}{n_{111}}\binom{n_{001}/2+n_{000}-1}{n_{000}} \times \binom{n_{01}/2}{n_{101}}\binom{n_{01}/2-1}{n_{010}}. \tag{51}$$

G. n_{01}, n_{110}, and n_{001} are all even [see Cases (a), (f), (g), and (j) and Figs. 2(a), 2(f), 2(g), and 2(j)]

This situation is complicated by the fact that n_{01}, n_{110}, n_{001} all even can arise in four distinct ways, all of which involve a symmetry of occupation/vacancy of sites 1, 2, $N-1$, and N.

When n_{01}, n_{110}, and n_{001} are all even and sites 1, 2, $N-1$, and N are all vacant, there are $(n_{01}/2)$ "particle units" because each "particle unit" has two mixed nearest neighbor pairs associated with it. These "particle units" are separated by $(n_{01}/2)-1$ "vacancy units" because the two "vacancy units" incorporating sites 1 and 2 and sites $N-1$ and N do not separate "particle units." Any group of contiguous vacancies which serves to separate two "particle units" may be permuted with other groups of vacancies, subject to the constraint that n_{01} be conserved. By such permutations,

$$\binom{(n_{01}/2)-1}{n_{101}}$$

independent arrangements can be created.

Each "particle unit" which consists of two or more particles is terminated at each end by a next nearest neighbor pair of the 110 type. Thus, there are $(n_{110}/2)$ "particle units" consisting of two or more particles. The particles of which such a "particle unit" is composed can be permuted with the particles contained in other "particle units" which also consist of two or more particles. It follows that the n_{111} next nearest neighbor pairs of the 111 type can be distributed among those "particle units" which contain two or more particles in

$$\binom{(n_{110}/2)+n_{111}-1)}{n_{111}}$$

independent ways without changing the number of 110 type next nearest neighbor pairs. Here again we con-

sider the $(n_{110}/2)$ "particle units" as numbered boxes into which we can distribute the n_{111} next nearest neighbor pairs of the 111 type.[6]

There are $(n_{01}/2)$ "particle units" which separate the $(n_{01}/2)+1$ "vacancy units." Such "particle units" can be divided into two categories: those consisting of a single particle (which constitute a next nearest neighbor pair of the 010 type) and those consisting of two or more contiguous particles. As long as n_{010} is conserved, the single particle or group of contiguous particles, between any "vacancy units," can be interchanged with other single particles or group of contiguous particles in

$$\binom{n_{01}/2}{n_{010}}$$

independent ways.

Because each "vacancy unit," (except those at the ends of the space which incorporate sites 1 and 2 and sites $N-1$ and N) which consists of two or more vacancies is terminated at each end by a next nearest neighbor pair of the 001 type, there are $(n_{001}+2)/2$ "vacancy units" which consist of two or more contiguous vacancies. The n_{000} next nearest neighbors of the 000 type can be distributed among such "vacancy units" in

$$\binom{(n_{001}+2)/2+n_{001}-1}{n_{000}}=\binom{n_{001}/2+n_{000}}{n_{000}}$$

indepndent ways.

Thus for situations in which n_{01}, n_{110}, and n_{001} are all even *and* sites 1, 2, $N-1$, and N [Case (a)] are all vacant.

$$\begin{aligned}A[\{n_{ij}\},\{n_{ijk}\},q,N]&=A[n_{01},n_{111},n_{101},n_{110},n_{000},n_{001},n_{010}]\\&=\binom{n_{110}/2+n_{111}-1}{n_{111}}\binom{n_{001}/2+n_{000}}{n_{000}}\binom{n_{01}/2-1}{n_{101}}\binom{n_{01}/2}{n_{010}}.\end{aligned}\tag{52}$$

By reasoning similar to that presented above it can be shown that when n_{01}, n_{110}, and n_{001} are all even *and* sites 1 and N are occupied and sites 2 and $N-1$ are vacant [Case (f)]

$$\begin{aligned}A[\{n_{ij}\},\{n_{ijk}\},q,N]&=A[n_{01},n_{111},n_{101},n_{110},n_{000},n_{001},n_{010}]\\&=\binom{n_{110}/2+n_{111}-1}{n_{111}}\binom{n_{001}/2+n_{000}-1}{n_{000}}\binom{n_{01}/2\quad n_{01}/2-1}{n_{101}\qquad n_{010}}.\end{aligned}\tag{53}$$

Similarly, when n_{001}, n_{110}, and n_{001} are all even and sites 2 and $N-1$ are occupied and sites 1 and N are vacant [Case (g)] we may write

$$\begin{aligned}A[\{n_{ij}\},\{n_{ijk}\},q,N]&=A[n_{01},n_{111},n_{110},n_{101},n_{000},n_{001},n_{010}]\\&=\binom{n_{110}/2+n_{111}-1}{n_{111}}\binom{n_{001}/2+n_{000}-1}{n_{000}}\\&\quad\times\binom{(n_{01}/2)-1}{n_{101}}\binom{(n_{01}/2}{n_{010}}.\end{aligned}\tag{54}$$

And when n_{01}, n_{110}, and n_{001} are all even *and* sites 1, 2, $N-1$, and N are all occupied [Case (j)] then

$$\begin{aligned}A[\{n_{ij}\},\{n_{ijk}\},q,N]&=A[n_{01},n_{111},n_{110},n_{101},n_{000},n_{001},n_{010}]\\&=\binom{n_{110}/2+n_{111}-1}{n_{111}}\binom{n_{001}/2+n_{000}-1}{n_{000}}\\&\quad\times\binom{n_{01}/2}{n_{101}}\binom{n_{01}/2-1}{n_{010}}.\end{aligned}\tag{55}$$

To conclude this section we note that if no reflective symmetry exists for the occupation of sites 1, 2, $N-1$, and N all the foregoing expressions for the degeneracy become

$$\begin{aligned}A[\{n_{ij}\},\{n_{ijk}\},q,N]&=A[n_{11},n_{00},n_{111},n_{101},n_{000},n_{010},q,N]\\&=2\binom{n_{11}-1}{n_{111}}\binom{n_{00}-1}{n_{000}}\binom{q-1-n_{111}}{n_{101}}\binom{N-q-1-n_{00}}{n_{010}}.\end{aligned}\tag{56}$$

This result is obtained, for example, by substituting into Eq. (45) values for n_{01}, n_{110}, and n_{001} from Eqs. (16)—(19). The same result is obtained also by substituting values for n_{01}, n_{110}, and n_{001} from Eqs. (32)—(35) into Eq. (47); from Eqs. (20)—(23) into Eq. (48); from Eqs. (12)—(15) into Eq. (50) and Eqs. (36)—(39) into Eq. (51). The factor of two results from the lack of reflective symmetry.

If the occupation/vacancy of sites 1, 2, $N-1$, and N exhibits reflective symmetry then Eqs. (52)—(55) yield

$$\begin{aligned}A[\{n_{ij}\},\{n_{ijk}\},q,N]&=A[n_{11},n_{00},n_{111},n_{101},n_{000},n_{010},q,N]\\&=\binom{n_{11}-1}{n_{111}}\binom{n_{00}-1}{n_{000}}\binom{q-1-n_{11}}{n_{101}}\binom{N-q-1-n_{00}}{n_{010}}\end{aligned}\tag{57}$$

when values for n_{01}, n_{110}, and n_{001} are substituted from Eqs. (4)—(7), Eqs. (24)—(27), Eqs. (28)—(31), and Eqs. (40)—(43), respectively.

Both Eqs. (56) and (57) may be interpreted in the following way: $q-1-n_{11}$ is the number of separations between particles which consist of one or more vacancies, that is, it is the number of separations between the "particle units." This follows from the fact that $q-1$ is the total number of separations between the q particles; if we substract n_{11}, the number of separations consisting of no vacancies (occupied nearest neighbors), then $q-1-n_{11}$ is the number of separations consisting of at least one vacancy. Such separations can be divided into two groups: One group which consists of separations by single vacancies (which thus form the n_{101} next nearest neighbor pairs of the 101 type), and those separations which consist of two or more contiguous vacancies. Thus,

$$\binom{q-1-n_{11}}{n_{101}}$$

arrangements can be created by permuting the single-vacancy separations with the separations which consist of two or more contiguous vacancies. In this way, the number of next nearest neighbor pairs of the 101 type is conserved.

Similarly,

$$\binom{N-q-1-n_{00}}{n_{010}}$$

represents the number of arrangements created when the separations between the "vacancy units" which consist of a single particle are permuted with the separations consisting of two or more contiguous particles.

$(n_{11}-1)$ can be thought of as representing the total number of separations between the n_{11} occupied nearest neighbor pairs. Such separations consist of no vacancies, one vacancy, two vacancies, etc. If we substract n_{111}, the number of separations involving no vacancies then $n_{11}-n_{111}-1$ is the number of such separations consisting of one or more vacancies between groups of one or more nearest neighbor pairs. Thus $(n_{11}-n_{111})$ is the number of "particle units" which consist of two or more contiguous particles. Such "particle units" can be considered to be numbered boxes into which the n_{111} next nearest neighbor pairs of the 111 type can be distributed in

$$\binom{n_{11}-n_{111}+n_{111}-1}{n_{111}}=\binom{n_{11}-1}{n_{111}}. \tag{58}$$

The factor

$$\binom{n_{00}-1}{n_{000}}$$

can be derived by reasoning similar to that presented above in which vacant sites are considered instead of occupied sites.

IV. DETERMINATION OF $A[n_{11}, n_{101}, n_{111}, q, N]$

The expressions for the composite degeneracies contained in Eqs. (56) and (57) can be utilized to develop addition relationships which describe composite degeneracies of a lower order, i.e., degeneracies in which fewer constraints are placed on the number of nearest and/or next nearest neighbor pairs. We next present an illustration of the considerations involved in determining composite degeneracies of lower order.

Expressions have been developed previously[7] which yield exactly, the degeneracy of arrangements characterized by the number of occupied nearest and next nearest neighbor pairs (as given by the number of next nearest pairs of the 101 and the 111 types), i.e., $A[n_{11},n_{101},n_{111},q,N]$. To determine $A[n_{11},n_{101},n_{111},q,N]$ from Eqs. (56) and (57) we sum over all permissible values of n_{00}, n_{010}, and n_{000},

$$\sum_{n_{00}}\sum_{n_{010}}\sum_{n_{000}} A[\{n_{ij}\},\{n_{ijk}\},q,N]=A[n_{11},n_{101},n_{111},q,N]. \tag{59}$$

Before the sums indicated in Eq. (59) can be evaluated, however, we must ascertain the range of permissible values for each of the summation variables. We first determine six essential inequalities.

Each "particle unit" (except those which incorporate a particle which occupies site 1 and/or N) is terminated at each end by a mixed nearest neighbor pair. Thus there are two mixed nearest neighbor pairs associated with each "particle unit" unless the "particle unit" is at one or the other or both ends of the array, in which case the "particle unit" is terminated at only one of its ends by a mixed nearest neighbor pair. Because there are $q-n_{11}$ "particle units," we may write

$$2(q-n_{11}-1)\leq n_{01}\leq 2(q-n_{11}). \tag{60}$$

Similarly, there are $N-q-n_{00}$ "vacancy units." These "vacancy units," except at most two (when the sites 1 and/or N are incorporated in a vacancy unit) are terminated at each end by a mixed nearest neighbor pair. Thus

$$2(N-q-n_{00}-1)\leq n_{01}\leq 2(N-q-n_{00}). \tag{61}$$

We have shown previously that $(n_{11}-n_{111})$ represents the number of "particle units" consisting of two or more contiguous particles. Because each "particle unit" which consists of two or more particles is terminated at each end by a next nearest neighbor pair of the 110 type (except for the "particle units" which incorporate sites 1 and 2 and/or $N-1$ and N), we may write

$$2(n_{11}-n_{111}-1)\leq n_{110}\leq 2(n_{11}-n_{111}). \tag{62}$$

Similarly, because $n_{00}-n_{000}$ represents the number of "vacancy units" which consist of two or more contiguous vacancies, and because each "vacancy unit," consisting of two or more contiguous vacancies, is terminated at each end by a next nearest neighbor pair of the 001 type (except in those situations where the "vacancy unit" incorporates sites 1 and 2 and/or sites $N-1$ and N) we obtain

$$2(n_{00}-n_{000}-1)\leq n_{001}\leq 2(n_{00}-n_{000}). \tag{63}$$

Because $q-n_{11}$ is the total number of "particle units," i.e., the number of "particle units" consisting of one or more contiguous particles and because $n_{11}-n_{111}$ is the number of "particle units" which consist of two or more contiguous particles, the quantity $q-2n_{11}+n_{111}$ represents the number of "particle units" which consist of exactly one particle. Except in those situations when these "particle units," consisting of a single particle, occur at either end of a lattice, there will be a next nearest neighbor pair of the 010 type associated with each single particle, "particle unit." Thus

$$(q-2n_{11}+n_{111}-2)\leq n_{010}\leq (q-2n_{11}+n_{111}). \tag{64}$$

Similarly, $N-q-n_{00}$ is the number of "vacancy units" which consist of one or more vacancies; and $n_{00}-n_{000}$ is the number of "vacancy units" which consist of two or more contiguous vacancies. Thus $N-q-n_{00}+n_{000}$ is the number of "vacancy units" which consist of exactly a single vacancy. Associated with each of such "vacancy units" there is a next nearest neighbor pair of the 101 type which (except for those single vacancy "vacancy units") exist at either or both of the sites 1 and N. We may then write

$$(N-q-2n_{00}+n_{000}-2)\leq n_{101}\leq (N-q-2n_{00}+n_{000}). \tag{65}$$

We are now in a position to determine the ranges of the summation variables n_{00}, n_{010}, and n_{000} necessary to perform the sums indicated in Eq. (59).

The limits on n_{000} can be seen from Eq. (65), i.e.,

$$q-N+2n_{00}+n_{101}\leq n_{000}\leq q-N+2n_{00}+n_{101}+)2. \tag{66}$$

The lower limit corresponds to the situation in which

no single, isolated vacancy can exist in either site 1 and/or N. There are six possible configurations involving the occupation of sites 1, 2, $N-1$, and N in which this is possible [see Fig. 5(a)].

$n_{000}=q-N+2n_{00}+n_{101}+1$ implies that either site 1 or N, but not both, contains an isolated, single vacancy [see Fig. 5(b)]. There are three possible distinguishable configurations of this type.

When single, isolated vacancies exist in sites 1 and N, $n_{000}=q-N+2n_{00}+n_{101}+2$ [see Fig. 5(c)]. (Only one possible arrangement of this type exists.)

The range for n_{010} is obtained directly from Eq. (64). The lower limit is associated with the situation in which "particle units," consisting of single, isolated particles, reside in both sites 1 *and* site N [see Fig. 6(a)]. This can occur in one way only.

$n_{010}=q-2n_{11}+n_{111}-1$ implies that a single, isolated particle resides on either sites 1 or N but not both [see Fig. 6(b)]. Three arrangements are possible.

When neither site 1 or site N is occupied by a single, isolated particle $n_{01}=q-2n_{11}+n_{111}$ [see Fig. 6(c)]. Six arrangements of this type can occur.

The range of n_{00} can be ascertained through an examination of Eqs. (61) and (2), i.e.,

$$N-2q+n_{11}-1\leq n_{00}\leq N-2q+n_{11}+1. \tag{67}$$

$n_{00}=N-2q+n_{11}-1$ signifies that both sites 1 and N are vacant. It also denotes the fact that n_{01} is even. This can be seen by subtracting Eq. (4) from Eq. (5) [Case (a)] or by subtracting Eq. (12) from Eq. (13) [Case (c)] or by subtracting Eq. (28) from Eq. (29) [Case (g)].

When $n_{00}=N-2q+n_{11}$ either site 1 *or* site N (but not both sites) is occupied. The subtraction of Eq. (8) from (9) [Case (b)] or Eq. (16) from Eq. (17) (Case (d)] or Eq. (20) from (21) [Case (e)] or Eq. (32) from Eq. (33) [Case (h)] shows that n_{01} is odd.

The upper limit on n_{00} is associated with the configuration in which sites 1 and N are both occupied. By subtracting Eq. (24) from Eq. (25) [Case (f)] or Eq. (36) from Eq. (37) [Case (i)] or Eq. (40) from Eq. (41) [Case (j)] it is seen that n_{01} is even.

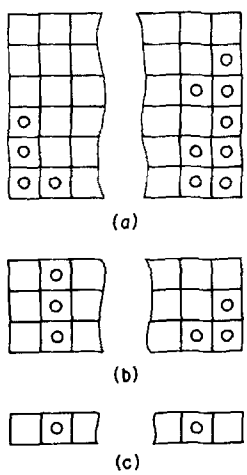


FIG. 5. (a) $n_{000}=q-N+2n_{00}+n_{101}$ implies that no "vacancy unit" consisting of a single vacancy can exist in site 1 *or* in site N. Six arrangements are possible. (b) $n_{000}=q-N+2n_{00}+n_{101}+1$ indicates that either in site 1 *or* in site N, but not both, there exists a single isolated vacancy. This can occur in three ways. (c) When "vacancy units," which consist of a single vacancy, exist in *both* sites 1 and N, then $n_{000}=q-N+2n_{00}+n_{101}+2$. Such a configuration can occur in one way only.

It follows then that Eq. (59) becomes

$$A[n_{11},n_{101},n_{111},q,N]$$

$$\sum_{n_{00}=N-2q+n_{11}-1}^{N-2q+n_{11}+1}\ \sum_{n_{110}=q-2n_{11}+n_{111}-2}^{q-2n_{11}+n_{111}}\ \sum_{n_{000}=q-N+2n_{00}+n_{101}}^{q-N+2n_{00}+n_{101}+2}$$

$$A[n_{11},n_{101},n_{111},n_{00},n_{010},n_{000},q,N]$$

$$=\binom{n_{11}-1}{n_{111}}\binom{q-n_{11}-1}{n_{101}}\Bigg\{\Bigg[\binom{q-n_{11}}{n_{11}-n_{111}}\binom{N-2q+n_{11}-2}{q-n_{11}-n_{101}}$$

$$+2\binom{q-n_{11}-1}{n_{11}-n_{111}}\binom{N-2q+n_{11}-2}{q-n_{11}-n_{101}-1}$$

$$+\binom{q-n_{11}}{n_{11}-n_{111}}\binom{N-2q+n_{11}-2}{q-n_{11}-n_{101}-2}\Bigg]$$

$$+2\Bigg[\binom{q-n_{11}-1}{n_{11}-n_{111}}\binom{N-2q+n_{11}-1}{q-n_{11}-n_{101}-1}+\binom{q-n_{11}-1}{n_{11}-n_{111}}$$

$$\times\binom{N-2q+n_{11}-1}{q-n_{11}-n_{101}-2}+\binom{q-n_{11}-1}{n_{11}-n_{111}-1}\binom{N-2q+n_{11}-1}{q-n_{11}-n_{101}-1}$$

$$+\binom{q-n_{11}-1}{n_{11}-n_{111}-1}\binom{N-2q+n_{11}-1}{q-n_{11}-n_{101}-2}\Bigg]$$

$$+\Bigg[\binom{q-n_{11}}{n_{11}-n_{111}+2}\binom{N-2q+n_{11}}{q-n_{11}-n_{101}-2}$$

$$+2\binom{q-n_{11}}{n_{11}-n_{111}+1}\binom{N-2q+n_{11}}{q-n_{11}-n_{101}-2}$$

$$+\binom{q-n_{11}}{n_{11}-n_{11}}\binom{N-2q+n_{11}}{q-n_{11}-n_{101}-2}\Bigg]\Bigg\}. \tag{68}$$

In Eq. (68) the terms in the first bracket arise when $n_{01}=N-2q+n_{11}-1$, in the second bracket when $n_{01}=N-2q+n_{11}$, and the third bracket contains terms from $n_{01}=N=2q+n_{11}+1$. The terms in each of the brackets may be summed[8] to yield

$$A[n_{11},n_{101},n_{111},q,N]$$

$$=\binom{n_{11}-1}{n_{111}}\binom{q-n_{11}-1}{n_{101}}\binom{q-n_{11}}{n_{11}-n_{111}}\binom{N-2q+n_{11}}{q-n_{11}-n_{101}}$$

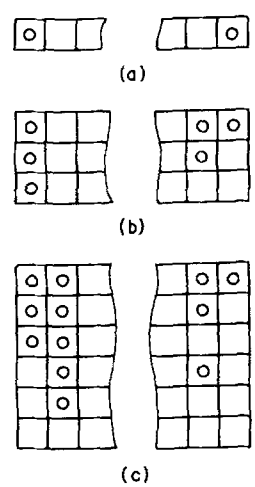


FIG. 6. (a) When "particle units" consisting of a single particle reside on both sites 1 and N, $n_{010}=q-2n_{11}+n_{111}-2$. Such a configuration can occur in only one way. (b) There are three ways in which a single particle "particle vacancy" can occupy either site 1 or site N (but not both). In such a case $n_{010}=q-2n_{11}+n_{111}-1$. (c) $n_{010}=q-2n_{11}+n_{111}$ denotes the six arrangements possible in which neither site 1 or N is occupied by a single, isolated particle.

$$+2\binom{N-2q+n_{11}}{q-2n_{11}-n_{101}}+\binom{N-2q+n_{11}}{q-n_{11}-n_{101}-2}$$
$$=\binom{n_{11}-1}{n_{111}}\binom{q-n_{11}-1}{n_{101}}\binom{q-n_{11}}{n_{11}-n_{111}}\binom{N-2q+n_{11}+2}{q-n_{11}-n_{101}}, \tag{69}$$

a result reported previously.[7]

Further reductions in the order of the composite degeneracies are possible using the summation ranges contained in Eqs. (60)—(65) As pointed out previously[7] when Eq. (69) is summed over $0 \leq n_{111} \leq n_{11}-1$ and $0 \leq n_{101} \leq q-n_{11}-1$ the result is

$$A[n_{11},q,N]=\binom{N-q+1}{q-n_{11}}\binom{q-1}{n_{11}} \tag{70}$$

and when Eq. (70) is summed over $0 \leq n_{11} \leq q-1$,

$$A[q,N]=\binom{N}{q}. \tag{71}$$

V. CONCLUSION

We have shown that for the ten general classes of arrangements considered, the specification of N, q, n_{11}, n_{00}, n_{111}, n_{000}, n_{101}, and n_{010} determines exactly, within a factor of two, the composite next nearest neighbor degeneracy for indistinguishable particles distributed on a one-dimensional lattice space.

ACKNOWLEDGMENTS

The author would like to express his appreciation to Professor D. A. King and the British Scientific Research Council whose assistance was invaluable in the research leading to the present paper.

[1]R. B. McQuistan, J. Math. Phys. **13**, 1317 (1972).
[2]R. B. McQuistan, Fibonacci Q. 14, 353 (1976).
[3]R. B. McQuistan, J. Math. Phys. **14**, 1258 (1973).
[4]R. B. McQuistan, J. Math. Phys. **15**, 1192 (1974).
[5]R. B. McQuistan, submitted to J. Math. Phys.
[6]J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (Wiley, New York, 1959), p. 438.
[7]R. B. McQuistan, J. Math. Phys. **15**, 1845 (1974).
[8]J. Riordan, *Combinational Identities* (Wiley, New York, (1968).

Exact nearest neighbor statistics for λ-bell particles on a one-dimensional lattice

R. B. McQuistan

Department of Physics and Laboratory for Surface Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201
(Received 12 October 1976)

Relationships are developed which describe exactly the degeneracies of those states specified by the number of either the occupied, mixed, or vacant nearest neighbor pairs which exist when indistinguishable λ-bell particles (occupying λ contiguous sites) are distributed on a one-dimensional lattice space.

I. INTRODUCTION

Any statistical mechanical treatment of cooperative phenomena which is predicated on the nearest neighbor approximation requires a knowledge of the multiplicity of those arrangements containing a prescribed number of nearest neighbor pairs. The purpose of the present paper is to determine the degeneracies of arrangements of λ-bell particles (particles that occupy λ contiguous sites) distributed on a one-dimensional lattice space when the numbers of nearest neighbor pairs are specified.

When q indistinguishable λ-bell particles are distributed on a lattice space of N equivalent sites, three kinds of nearest neighbor pairs can be created:

(1) both adjacent sites are occupied by portions of different λ-bell particles (1-1): occupied nearest neighbor pair;

(2) one site is occupied and an adjacent site is vacant (0-1): mixed nearest neighbor pair;

(3) both adjacent sites are vacant (0-0): vacant nearest neighbor pair.

Here we define an occupied nearest neighbor pair to be two adjacent sites occupied by portions of different λ-bell particles. A mixed nearest neighbor pair consists of two adjacent sites; one occupied, one vacant. A vacant nearest neighbor pair consists of two adjacent vacant sites. If n_{11}, n_{01}, and n_{00} are the numbers of such pairs, respectively, in any particular arrangement, they are related in the following approximate manner[1-3]:

$$2q = 2n_{11} + n_{01}, \tag{1}$$

$$2(N - \lambda q) = 2n_{00} + n_{01}. \tag{2}$$

These relationships can be seen by reference to Figs. 1A and B. If, as shown in Fig. 1A, one disregards the termination of the array at both ends and draws a line from the site occupied by the ends of a λ-bell particle to its adjacent site which is vacant or occupied by another particle, there will be $2q$ lines. These lines can be accounted for by observing that between each mixed nearest neighbor there is a horizontal line and between each occupied nearest neighbor there are two horizontal lines. Equation (1) then follows.

Next draw a line from each empty site to each of its adjacent nearest neighbors [see Fig. 1(b)]; there will be $2(N - q\lambda)$ lines on each array. Between each mixed nearest neighbor pair of sites there will be one line and between each vacant nearest neighbor pair two lines will appear and Eq. (2) arises.

Thus for a given q and N, the specification of any two of the quantities n_{11}, n_{01}, or n_{00} determines, within an error of two, the remaining number.

II. OCCUPIED NEAREST NEIGHBOR PAIR DEGENERACY

In this section we determine $A[n_{11}, q, N]$, the number of independent ways of arranging q indistinguishable λ-bell particles in a one-dimensional lattice space of N equivalent sites in such a way as to create exactly n_{11} occupied nearest neighbor pairs.

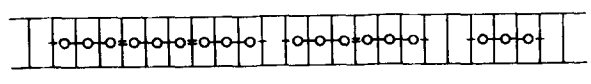
In a previous publication[4] we have shown that $A[q, N]$, the occupational degeneracy for the system under discussion, is given by the binomial coefficient

$$A[q, N] = \binom{N - q(\lambda - 1)}{q}. \tag{3}$$

From the set of such arrangements we select the subset which contains only arrays on which appear exactly n_{11} occupied nearest neighbor pairs. If we define a "unit" on such an array (see Fig. 2) to consist of one or more contiguous λ-bell particles together with a single vacancy (if one is needed) just to the right of the particles to separate a "unit" from the remainder of the particles on the array, then we observe that there are always $q - n_{11}$ "units."

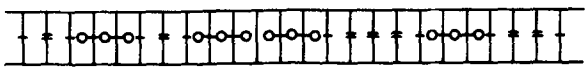
That there are always $q - n_{11}$ "units" can be seen from the following reasoning: If there are q, λ-bell particles, there are $q - 1$ separations between them. Of these $q - 1$ separations, n_{11} are occupied nearest neighbor pairs formed by the ends of adjacent λ-bell particles and $q - 1 - n_{11}$ separations are *not* between such pairs. Because these $q - 1 - n_{11}$ separations are between the "units," there must be $q - n_{11}$ "units."

The rest of the vacancies, of which there are $N - q - (q - n_{11} - 1) = N - q(\lambda + 1) + n_{11} + 1$, can be permuted



A

FIG. 1A. This figure is concerned with the derivation of Eq. (1) when $\lambda = 3$, $q = 6$, $N = 23$. A short horizontal line is drawn from the compartment occupied by the end of one λ-bell particle to its nearest neighbor site.



B

FIG. 1B. This figure is utilized in deriving Eq. (2) when $\lambda = 3$, $q = 4$, $N = 23$. A short horizontal line is drawn from a vacant site to each of its nearest neighbor sites.

with the "units" to form additional independent arrangements. It follows then that the number of individual objects which can be permuted to form independent arrangements is the sum of the number of "units" and the number of those vacancies which are not required to isolate the "units" from the remainder of the arrangement, i.e., $q - n_{11} + [N - (\lambda+1)q + n_{11} + 1] = N - \lambda q + 1$.

The number of independent ways of arranging the $N - \lambda q + 1$ entities [of which $q - n_{11}$ are indistinguishable "units" and $N - q(\lambda+1) + n_{11} + 1$ are permutable vacancies] is the binomial coefficient

$$\binom{N-\lambda q+1}{q-n_{11}} = \binom{N-\lambda q+1}{N-(\lambda+1)q+n_{11}+1}. \tag{4}$$

To derive the result contained in Eq. (4) we have assumed that the "units" are identical. Clearly the "units" cannot be indistinguishable because some contain more λ-bell particles than others. To remove the assumption that all the "units" are identical and to determine $A[n_{11}, q, N]$ we must enumerate the ways that the λ-bell particles can be arranged to form the $q - n_{11}$ "units."

There are a total of $q - 1$ indistinguishable separations between the q λ-bell particles (see Fig. 3); however, n_{11} of these separations constitute indistinguishable occupied nearest neighbor pairs which can be distributed in the $q - 1$ separations in

$$\binom{q-1}{n_{11}} = \binom{q-1}{q-1-n_{11}} \tag{5}$$

independent ways. By so doing we enumerate all the possible ways the "units" can be constituted while still preserving the number of "units" and the number of occupied nearest neighbor pairs.

Thus $A[n_{11}, q, N]$ can be determined to be the product of the number of ways the "units" can be permuted with the permutable vacancies and the number of ways in which the "units" may be constructed from the prescribed number of nearest neighbor pairs, i.e.,

$$A[n_{11}, q, N] = \binom{q-1}{n_{11}}\binom{N-\lambda q+1}{q-n_{11}}. \tag{6}$$

If $A[n_{11}, q, N]$ is summed over all values of n_{11}, the result should be the number of ways of arranging q indistinguishable λ-bell particles on a one-dimensional lattice of N equivalent sites [see Eq. (3)]. From the Vandermonde theorem[5] it is seen that such is the result.

III. MIXED NEAREST NEIGHBOR PAIR DEGENERACY

In this section we treat those nearest neighbor pairs which consist of two adjacent sites, one occupied by a portion of a λ-bell particle, the other vacant. Ising,[6] in his examination of one-dimensional ferromagnetic phase transitions, has provided an approximate answer to the degeneracy of arrangements containing a prescribed number of mixed nearest neighbor pairs when simple particles ($\lambda = 1$) are distributed on a one-dimensional lattice. An exact treatment for $\lambda = 1$ has been previously reported.[2]

Because of the symmetry conditions involved when treating mixed nearest neighbor pair degeneracy, we must deal separately with two different situations: when n_{01} is odd and when it is even.

A. n_{01} odd

If n_{01} is odd, it follows that of the two end compartments (one at each end of the space) one and only one end compartment of the lattice is occupied (see Fig. 4). Assuming that it is the compartment farthest to the left-hand end of the array which is occupied, we construct permutable "units," each of which consists of one or more contiguous λ-bell particles and if necessary a vacancy immediately to the right to isolate the "unit" from the rest of the particles on the array. Note that the particle or group of contiguous particles that are farthest to the left-hand end of the lattice do not constitute a permutable "unit" because it must remain pinned in the end position to satisfy the constraint that the left-hand compartment is occupied.

We assume initially that these "units" are indistinguishable regardless of the number of λ-bell particles they contain or whether or not they are terminated by a vacancy. (The "unit" on the right does not need a vacancy to isolate it.) The site on the extreme right must remain vacant.

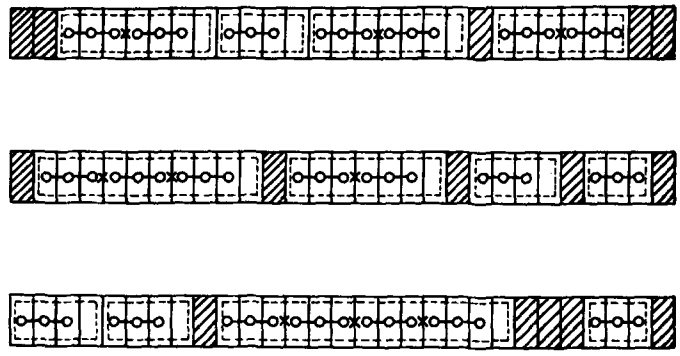


FIG. 2. $\lambda = 3$, $q = 7$, $N = 29$, $n_{11} = 3$. In this figure seven indistinguishable particles are arranged on a one-dimensional lattice space of 29 equivalent sites to yield three occupied nearest neighbor pairs (indicated by x's). There are $q - n_{11} = 4$ "units" and this figure shows the three possible ways in which the "units" may be constituted from the seven particles to form three occupied nearest neighbor pairs. Thus the top figure shows an arrangement in which each occupied nearest neighbor pair is in a different "unit"; the middle figure shows two nearest neighbor pairs in one "unit" and one pair in another "unit"; and in the bottom figure all occupied nearest neighbor pairs are in the same "unit," while the other three "units" do not contain an occupied nearest neighbor pair. There are $N - q(\lambda+1) + n_{11} + 1 = 5$ permutable vacancies (indicated by cross hatching).

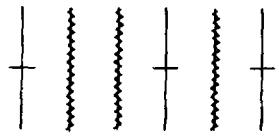


FIG. 3. The seven λ-bell particles (λ = 3) illustrated at the top of Fig. 2 have six separations between them. Three separations are associated with nearest neighbor pairs (shown by short horizontal lines) and three separations are not (jagged lined).

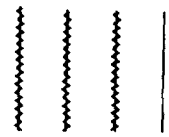


FIG. 5. For the arrangement shown in Fig. 4 there are four separations between the five particles (λ = 3); of these separations $[(n_{01}-1)/2]=3$ are composed of two mixed nearest neighbor pairs (jagged lines) and one is an occupied nearest neighbor pair. These separations can thus be arranged in $\binom{4}{1}=4$ ways.

Because only one mixed nearest neighbor pair is associated with the particle or group of contiguous particles on the left-hand end of the lattice and because every other "unit" on the lattice has two mixed nearest neighbor pairs, terminating it at each end, there is a total of $[(n_{01}-1)/2]$ "units." These "units" can be permuted with some of the indistinguishable vacancies to form unique arrangements.

There are $[N-\lambda q]$ vacancies but not all of them can be permuted with the "units" to form unique arrangements. The number of vacancies which is necessary to form the "units" is $[(n_{01}-1)/2]$ because each "unit," except the one on the right, has a vacancy associated with it but the vacancy on the right-hand end is not permutable and the particle or group of contiguous particles on the left also immobilizes a vacancy. Thus there are $N-\lambda q-[(n_{01}-1)/2]-1$ indistinguishable, permutable vacancies. It follows that there are a total of $N-\lambda q-[(n_{01}-1)/2]-1+[(n_{01}-1)/2]=N-\lambda q-1$ objects which can be arranged in

$$\binom{N-\lambda q-1}{\left[\frac{n_{01}-1}{2}\right]}=\binom{N-\lambda q-1}{N-\lambda q-1-\left[\frac{n_{01}-1}{2}\right]} \tag{7}$$

independent ways.

Obviously, the "units" are not indistinguishable entities, but can be constituted in various ways from different groupings of the λ-bell particles while still preserving the prescribed number of mixed nearest neighbor pairs on each array. To determine the number of different ways in which q indistinguishable λ-bell particles can be arranged to form the $[(n_{01}-1)/2]$ "units," see Fig. 5. We note that of the $q-1$ separations between the q λ-bell particles, $[(n_{01}-1)/2]$ of them are mixed nearest neighbor pairs. Such kinds of separations can be distributed among the $q-1$ separations between the q particles in

$$\binom{q-1}{\left[\frac{n_{01}-1}{2}\right]}=\binom{q-1}{q-1-\left[\frac{n_{01}-1}{2}\right]} \tag{8}$$

unique ways.

Equation (8) can be considered to yield the enumeration of the ways q indistinguishable λ-bell particles can be arranged to form $[(n_{01}-1)/2]$ "units" or conversely it can be considered to be the number of ways the $[(n_{01}-1)/2]$ "units" can be constituted from the q λ-bell particles.

Consequently, if we require the compartment on the left-hand side of the lattice to be occupied while the compartment on the right is vacant, then from Eqs. (7) and (8) we see that there are

$$\binom{N-\lambda q-1}{\left[\frac{n_{01}-1}{2}\right]}\binom{q-1}{\left[\frac{n_{01}-1}{2}\right]}$$

independent arrangements possible.

With equal probability the compartment on the right could have been occupied and the compartment on the left vacant, so that if n_{01} is odd we obtain

$$A[n_{01},q,N]=2\binom{N-\lambda q-1}{\left[\frac{n_{01}-1}{2}\right]}\binom{q-1}{\left[\frac{n_{01}-1}{2}\right]}_{n_{01}\ \text{odd}} \tag{9}$$

B. n_{01} even ($\neq 0$)

If n_{01} is even ($\neq 0$), then one of the two following situations must exist in any particular array:

(a) The compartments which terminate the lattice space at each end are both vacant (see Fig. 6); or

(b) Both end compartments are occupied (see Fig. 7).

If there are n_{01} mixed nearest neighbor pairs and both end compartments are empty, then there are always $[n_{01}/2]$ "units," each of which consists of one or more contiguous λ-bell particles together with a vacancy (if one is needed) immediately to, say, the left, to separate the "unit" from other particles on the array. The reason that there are $[n_{01}/2]$ "units" is that each "unit" is terminated at each end by a mixed nearest neighbor pair. As previously mentioned, we initially assume these "units" to be identical, indistinguishable entities which can be permuted with some of the indistinguishable vacancies to form independent arrangements.

Not all of the $N-\lambda q$ vacancies are permutable with the "units," i.e., some of them are required to form

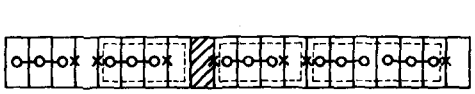


FIG. 4. Five particles (λ = 3) are arranged on $N=20$ with $n_{01}=7$. There are $[(n_{01}-1)/2]=3$ permutable "units" and $N-\lambda q-[(n_{01}-1)/2]-1=1$ permutable vacancies (see cross hatching).

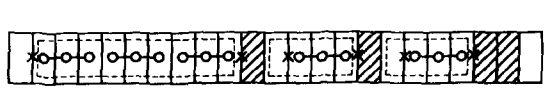


FIG. 6. In this figure, $q=5$, $N=23$, and both end compartments are empty. $n_{01}=6$. There are $[(n_{01}-1)/2]=3$ "units" and $N-\lambda q-[(n_{01}+2)/2]=4$ permutable vacancies (cross hatched).

FIG. 7. In this figure, $q=6$, $N=24$, $n_{01}=6$ and both end compartments are occupied. There are $[n_{01}/2]-1=2$ permutable "units" enclosed by a dashed-line box and $N-\lambda q-1$ $[(n_{01}/2)-1]=3$ permutable vacancies (cross hatched). The "units" at each end of the array are not permutable because the compartments at the ends of the array must be occupied.

the "units" (except for the "unit" on the extreme left-hand side of the array) and a vacancy resides at each end of the lattice. Thus $[(n_{01}-2)/2]+2$ vacancies are not available for permutation with the "units" but $N-\lambda q-[(n_{01}+2)/2]$ vacancies can be permuted with the "units" to form countable arrangements. There are, therefore, $[n_{01}/2]+N-\lambda q-[(n_{01}+2)/2]=N-\lambda q-1$ objects which can be permuted to form

$$\binom{N-\lambda q-1}{\left[\frac{n_{01}}{2}\right]}=\binom{N-\lambda q-1}{N-\lambda q-1-\frac{n_{01}}{2}} \tag{10}$$

independent arrangements.

To this point in the argument we have assumed the "units" to be identical. Clearly this is not the case; units may be composed of one or more λ-bell particles. In order to remove this constraint we assert that there are

$$\binom{q-1}{\frac{n_{01}}{2}-1}$$

unique ways of arranging the q λ-bell particles within the $[n_{01}/2]$ "units." This can be established by the following reasoning: There are $q-1$ lines which symbolize the separation of the q λ-bell particles (see Fig. 8). Of this set $[(n_{01}-2)/2]$ constitute separation between "units," i.e., separations involving two adjacent nearest neighbor pairs; and $q-1[(n_{01}-2)/2]$ lines separate adjacent λ-bell particles. Thus these $q-1$ separations can be arranged in

$$\binom{q-1}{\left[\frac{n_{01}-2}{2}\right]}$$

unique ways. This represents the number of ways the q λ-bell particles can be arranged to form the $[n_{01}/2]$ "units" (with at least one particle within each "unit") when both end compartments are vacant.

It follows that the number of possible arrangements which can be created when both end compartments are vacant is just the number of ways the "units" can be arranged with the permutable vacancies, multiplied by the number of ways the "units" can be constructed from the λ-bell particles on the lattice.

If situation (b) obtains, then there are $[n_{01}/2-1]$ permutable units (see Fig. 7). There are $N-\lambda q-1-[n_{01}/2-1]$ indistinguishable, permutable vacancies because a vacancy must be tied up with each of the "units" and an additional vacancy is required to isolate the λ-bell particle or groups of particles which constitute a "unit" at one end of the array. The λ-bell particle or groups of particles at the other end is isolated by the vacancy associated with the "unit" next to it. Thus there are a total of $N-\lambda q-1-[N_{01}/2-1]+[n_{01}/2-1]=N-\lambda q-1$ objects which can be arranged in

$$\binom{N-\lambda q-1}{\left[\frac{n_{01}}{2}-1\right]}$$

unique ways.

In this case, the "units" can be created in

$$\binom{q-1}{\frac{n_{01}}{2}}$$

ways because the $q-1$ separations between the q λ-bell particles consist of $[n_{01}/2]$ separations between the "units" and particles or groups of particles on the ends of the array and $q-1-[n_{01}/2]$ separations between adjacent λ-bell particles. Therefore, the $[n_{01}/2]$ "units" as well as the groups of particles on the ends can be created from the q λ-bell particles in

$$\binom{q-1}{\frac{n_{01}}{2}}$$

independent ways.

Consequently, when both end compartments are occupied, the q indistinguishable particles can be arranged in

$$\binom{N-\lambda q-1}{\frac{n_{01}}{2}-1}\binom{q-1}{\frac{n_{01}}{2}}$$

unique ways to form exactly n_{01} mixed nearest neighbor pairs.

It follows that $A[n_{01},q,N]$, the total number of unique arrangements which can be created when q indistinguishable λ-bell particles are arranged on a one-dimensional lattice of N equivalent compartments to form n_{01} (even) mixed nearest neighbor pairs is the sum of the number of arrangements created when both end compartments are vacant and when they are both occupied, is given by

$$A[n_{01},q,N]=\binom{N-\lambda q-1}{\left[\frac{n_{01}}{2}\right]}\binom{q-1}{\frac{n_{01}}{2}-1}+\binom{N-\lambda q-1}{\frac{n_{01}}{2}-1}\binom{q-1}{\frac{n_{01}}{2}}$$

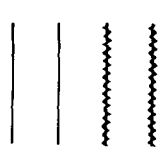


FIG. 8. In Fig. 6 there are four separations between the five particles. $[(n_{01}-2)/2]=2$ of these separations involve a pair of mixed nearest neighbors (jagged lines) and two do not. These four separations may be arranged in $\binom{4}{2}=6$ ways.

$$= 2\left[\frac{N-(\lambda-1)q-n_{01}}{n_{01}}\right]\binom{N-\lambda q-1}{\frac{n_{01}}{2}-1}\binom{q-1}{\frac{n_{01}}{2}-1}_{n_{01}\text{ even }(\neq 0)}. \tag{11}$$

The only way an arrangement of particles can contain no mixed nearest neighbor pairs is when the lattice is completely empty, $q=0$, or when it is completely filled, $q=N/\lambda$.

The normalization for $A[n_{01}, q, N]$ may be determined as follows.[4] Assuming that $q \neq 0$ or N/λ,

$$\begin{aligned}\sum_{n_{01}=1} A[n_{01}, q, N] &= \left\{2+2\left[\frac{N-q(\lambda-1)-2}{2}\right]+\binom{N-q\lambda-1}{1}\binom{q-1}{1}\right. \\ &\quad +2\left[\frac{N-q(\lambda-1)-4}{4}\right]\binom{N-q\lambda-1}{1}\binom{q-1}{1} \\ &\quad \left.+2\binom{N-q\lambda-1}{2}\binom{q-1}{2}\right. \\ &\quad +2\left[\frac{N-(\lambda-1)q-6}{6}\right]\binom{N-q\lambda-1}{2}\binom{q-1}{1}+\cdots \\ &= \left[\frac{N-q(\lambda-1)}{q}\right]\binom{N-\lambda q}{j}\binom{q-1}{q-1-j} = \binom{N-q(\lambda-1)}{q},\end{aligned} \tag{12}$$

where $A[n_{01}, q, N]$ is given alternately by Eqs. (9) and (11) as the sum proceeds over all values of n_{01}.

IV. VACANT NEAREST NEIGHBOR PAIR DEGENERACY

The number of arrangements containing a prescribed number of vacant nearest pairs can be determined as follows (see Fig. 9): All arrangements of q indistinguishable λ-bell particles on a one-dimensional lattice which contain n_{00} vacant nearest neighbor pairs always exhibit $N-\lambda q-n_{00}$ "units" which we now define to consist of one or more contiguous vacant sites; and a λ-bell particle (if one is needed) to isolate the "unit" from the rest of the lattice. We initially consider these "units" to be identical entities, indistinguishable from one another. There are $N-\lambda q-n_{00}$ of such "units" because there are $N-\lambda q$ vacancies with $N-\lambda q-1$ separations n_{00} of which are between vacant nearest neighbor pairs and $N-\lambda q-n_{00}-1$ are separations not between vacant nearest neighbor pairs, i.e., are separations between "units." It follows that if there are $N-\lambda q-n_{00}-1$ separations between "units" then there are $N-\lambda q-n_{00}$ "units."

There are q indistintuishable λ-bell particles which can be permuted with the "units" to form unique arrangements; some of the λ-bell particles must be used to separate a "unit" from the remainder of the array. There are $N-\lambda q-n_{00}-1$ λ-bell particles required to isolate the "units" (a "unit" on one of the ends does not need a λ-bell particle to isolate it). Thus there are $q-(N-\lambda q-n_{00}-1)=(\lambda+1)q+n_{00}+1-N$ indistinguishable permutable λ-bell particles remaining to permute with the "units." The $q+1$ object can be permuted in

$$\binom{q+1}{N-\lambda q-n_{00}} = \binom{q+1}{(\lambda+1)q+n_{00}+1-N} \tag{13}$$

independent ways.

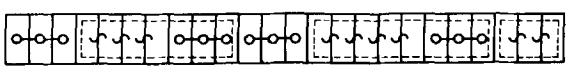


FIG. 9. In this figure, $N=24$, $q=4$, $n_{00}=9$ (denoted by ~). There are $N-\lambda q-n_{00}=3$ "units" and $(\lambda+1)qn_{00}+1-N=2$ permutable particles.

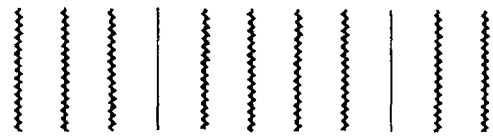


FIG. 10. In Fig. 9 there are $N-\lambda q=12$ vacancies with 11 separations between them. $n_{00}=9$ of these separations (jagged lines) involve vacant nearest neighbor pairs and two do not. These may be arranged in $\binom{11}{2}=55$ ways.

The "units" are initially considered to be indistinguishable; however the vacancies can be moved from one "unit" to another subject only to the constraint that the number of "units" and the number of vacancies are conserved.

From Fig. 10 we see that there are $N-\lambda q$ vacancies with $N-\lambda q-1$ separations between them. n_{00} of these separations are between vacant nearest neighbor pairs, and $N-\lambda q-1-n_{00}$ separate pairs of sites which are not vacant, i.e., which separate "units." Thus the total number of separations between vacant sites is composed of two permutable groups, the "units" and the free vacancies, which can be arranged in

$$\binom{N-\lambda q-1}{n_{00}} = \binom{N-\lambda q-1}{N-\lambda q-1-n_{00}} \tag{14}$$

unique ways.

$A[n_{00}, q, N]$, the degeneracy of vacant nearest neighbor pairs, is the product of the number of ways the "units" can be permuted with the particles and the number of ways the "units" can be constructed from the vacancies. Thus

$$A[n_{00}, q, N] = \binom{q+1}{N-\lambda q-n_{00}}\binom{N-\lambda q-1}{n_{00}}. \tag{15}$$

When $A[n_{00}, q, N]$ is summed over all values of n_{00} the result, by the Vandermonde theorem,[5] is

$$\binom{N-(\lambda-1)q}{q},$$

in accordance[4] with Eq. (3).

V. SUMMARY

Utilizing combinatorial arguments, the exact degeneracies are determined for arrangements of λ-bell

particles on a one-dimensional space, when the number of either occupied, mixed or vacant nearest neighbor pairs is specified.

ACKNOWLEDGMENTS

The support of the British Scientific Research Council and the suggestions of Professor D.A. King greatly facilitated the work reported in the present paper. For these assistances the author wishes to express his appreciation.

[1]K. Haung, *Statistical Mechanics* (Wiley, New York, 1963).
[2]R. B. McQuistan, J. Math. Phys. **13**, 1317 (1972).
[3]R. B. McQuistan, J. Math. Phys. **15**, 1192 (1974).
[4]R. B. McQuistan, Nuovo Cimento **58**, 86 (1968).
[5]J. Riordan, *Combinational Identities* (Wiley, New York, 1968).
[6]E. Ising, Z. Phys. **31**, 253 (1925).

Killing inequalities for relativistically rotating fluids. II*

R. O. Hansen†

Mathematical Institute, University of Oxford, Oxford OX1 3LB, England

Jeffrey Winicour

Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260
(Received 28 July 1976)

Theorems that the angular momentum density and the angular velocity of locally nonrotating observers are postive are established for general relativistic fluids which rotate differentially with posittve angular velocity. The results apply to the pseudostationary case in which the system possesses an ergoregion. An application to the stability of ergoregions is discussed.

I. INTRODUCTION

In a previous paper[1] (hereafter referred to as Paper I) an inequality was derived which guarantees that the angular momentum density of a stationary, axisymmetric, rigidly rotating, asymptotically flat, perfect fluid system in general relativity is everywhere positive (except on the axis, where it vanishes). In the Newtonian limit, this result is quite obvious and follows from *local* arguments. In the general relativistic case, the result depends upon subtle *global* properties. We here consider generalizations of this result to pseudostationary cases where ergoregions may be present and where differential rotation is allowed. In the latter case, extra boundary conditions must be imposed to establish the result. However, we establish an analogous result for the sign of the angular velcoity of locally nonrotating observers which holds under quite general conditions. Although ergoregions and differential rotation are expected to lead to instabilities,[2,3] the relevant time scales may in practice be long enough to warrant consideration of the motion as pseudostationary.

We employ the notation of Paper I and the field equations derived there. A brief summary of the formalism is given in Sec. II. In Sec. III we discuss rigidly rotating systems with ergoregions. In Sec. IV, we establish several inequalities for differentially rotating systems with ergoregions including the results mentioned above. In Sec. V, we discuss some of the limitations and applications of our results.

Throughout the paper, we assume that the fluid occupies a single simply connected solid region, as in the case of an ordinary star. Although this excludes toroidal configurations, many of our results can be extended to such cases. We also assume asymptotic flatness and that there are no horizons. These last conditions play an important role in all our results.

II. SUMMARY OF THE FIELD EQUATIONS

In Paper I, the field equations with arbitrary sources for systems admitting two commuting Killing vector fields were written as differential equations on the two-dimensional manifold S of orbits of the Killing fields, with induced metric h_{ab}. We here summarize these results for the case of stationary, axisymmetric perfect fluids whose 4-velocities are tangent to orbits of the Killing fields.

Let ξ_0^a and ξ_1^a be the timelike and rotational Killing vector fields, respectively; let $\lambda_{00}=\xi_0^m\xi_{0m}$, $\lambda_{01}=\xi_0^m\xi_{1m}$, and $\lambda_{11}=\xi_1^m\xi_{1m}$. We introduce greek indices to label the symmetric pairs (00), (01), and (11) which appear in these definitions, and we raise and lower these indices with a constant field $G_{\alpha\beta}$ and its inverse, such that

$$G^{\mu\nu}\lambda_\mu\lambda_\nu=-\tau^2=2(\lambda_{00}\lambda_{11}-\lambda_{01}^2). \tag{2.1}$$

The 4-velocity u^a of the fluid is written

$$u^a=(-\psi)^{-1/2}(\xi_0^a+\Omega\xi_1^a), \tag{2.2}$$

where ψ is given by

$$\psi=S^\mu\lambda_\mu=\lambda_{00}+2\Omega\lambda_{01}+\Omega^2\lambda_{11}. \tag{2.3}$$

We also introduce N^α with components (0, 1, Ω) and define η by

$$\eta=N^\mu\lambda_\mu=\lambda_{01}+\Omega\lambda_{11}. \tag{2.4}$$

The field equations take the form

$$D^m(\tau^{-1}D_m\lambda_\alpha)=\tau^{-3}\lambda_\alpha(D^m\lambda^\mu)(D_m\lambda_\mu)+8\pi\tau^{-1}[(\mu+3p)\lambda_\alpha+(\mu+p)\tau^2\psi^{-1}S_\alpha], \tag{2.5}$$

$$R=\tfrac{1}{2}\tau^{-2}(D^m\lambda^\mu)(D_m\lambda_\mu)+8\pi(\mu+p), \tag{2.6}$$

and

$$\tfrac{1}{2}(\mu+p)\psi^{-1}(D_a\psi-\lambda^\mu D_aS_\mu)=-D_ap, \tag{2.7}$$

where μ and p are the density and pressure of the fluid, D_a is the covariant derivative on S, and R is the scalar curvature of S. The following linear combination of Eq. (2.5) is useful,

$$A^{[\alpha}B^{\beta]}D^m[\tau^{-1}\lambda_\beta D_m\lambda_\alpha]=8\pi(\mu+p)\tau\psi^{-1}A^{[\alpha}B^{\beta]}S_\alpha\lambda_\beta. \tag{2.8}$$

When A^α and B^α are constant, this leads to

$$D^m[\tau^{-1}(B^\beta\lambda_\beta)^2D_m(A^\alpha\lambda_\alpha/B^\gamma\lambda_\gamma)]=16\pi(\mu+p)\tau\psi^{-1}A^{[\alpha}B^{\beta]}S_\alpha\lambda_\beta. \tag{2.9}$$

Many of the results of this paper are based upon choices A^α and B^α which make the right-hand side of Eq. (2.9) positive or negative so that Hofp's theorem[4] regarding extrema may be applied.

The angular momentum of the system is given by

$$J=-\int(\mu+p)\psi^{-1}\eta\xi_0^m dS_m, \tag{2.10}$$

where the integration is taken over a spacelike surface

containing the trajectories of ξ_1^a. The angular momentum density is defined to be the integrand in Eq. (2.10).

III. RIGID SYSTEMS WITH ERGOREGIONS

The argument used in Paper I to show that the angular momentum density of a rigidly rotating system is positive depends at only one step on the condition that the system does not have an ergoregion (i.e., $\lambda_{00} < 0$): The proof that $\eta > 0$ on the velocity of light curve ($\psi = 0$) requires this condition. Thus it suffices to show that $\eta > 0$ when $\psi = 0$.

We first establish that the ergoregion is disjoint from the axis. The axis regularity conditions of Paper I imply that on the axis

$$\tau = \lambda_{01} = \lambda_{11} = n^m D_m \lambda_\alpha = 0, \tag{3.1a}$$

$$a \equiv n^m D_m \tau > 0, \tag{3.1b}$$

$$b \equiv n^m D_m (n^p D_p \lambda_{11}) > 0, \tag{3.1c}$$

where n^a is the unit normal to the axis and **a** and **b** are smooth positive functions along the axis. By differentiating Eq. (2.1) we then obtain

$$b\lambda_{00} = -a^2 \tag{3.2}$$

on the axis. Thus $\lambda_{00} < 0$ on the axis. It follows immediately that the velocity of light curve is disjoint from the axis, for ψ and λ_{00} agree there. (Note we speak of a velocity of light curve rather than a velocity of light surface or cylinder since our argument refers to the two-dimensional manifold S.)

We now show that the velocity of light curve has a single connected component. Since $\psi \neq 0$ in the fluid (from causality requirements) and since $\psi = 0$ is disjoint from the axis, in any region bounded by $\psi = 0$ we have

$$D^m[\tau^{-1}\lambda_{11}^2 D_m(\psi/\lambda_{11})] = 0, \tag{3.3}$$

by Eq. (2.9) with $A^\alpha = S^\alpha$ and $B^\alpha = (0, 0, 1)$. The Killing scalar λ_{11} is everywhere positive in the region, so the Hopf theorem applies to Eq. (3.3). Thus if $\psi = 0$ forms the boundary of some region, ψ/λ_{11} has no maxima or minima in that region and must vanish everywhere. We conclude that $\psi = 0$ cannot bound a region of S. A similar argument applied to the regions bounded by the curves $\psi/\lambda_{11} = -\epsilon$ for some sufficiently small positive constant ϵ rules out bounded curves on which $\psi = 0$. Thus the velocity of light curves must extend to infinity. By asymtotic flatness, we have $\psi \sim -1 + \frac{1}{2}\Omega^2\tau^2$ at large distances from the fluid. Since there is only one positive root of $\frac{1}{2}\Omega^2\tau^2 - 1 = 0$, we conclude that the velocity of light curve has a single connected component which separates S into two regions, $\psi < 0$ and $\psi > 0$.

We observe next that η has a single sign on each component of the velocity of light curve. This follows from the equation

$$\eta^2 = \tfrac{1}{2}\tau^2 + \lambda_{11}\psi. \tag{3.4}$$

Since $\psi = 0$ does not intersect the axis, $\eta^2 > 0$ on the velocity of light curve. In particular, $\eta > 0$ at infinity. Therefore, η is everywhere positive on the velocity of light curve.

The Hopf argument of Paper I now shows that $\eta/\psi < 0$ holds everywhere except on the axis, where $\eta/\psi = 0$. Thus, off the axis, the angular momentum density of a rigidly rotating fluid is everywhere positive, even in the presence of ergoregions.

IV. DIFFERENTIAL ROTATION

We now develop some properties of the two-dimensional manifold S describing a differentially rotating, pseudostationary fluid with a uniform sense of angular velocity ($\Omega > 0$). The most important of these are property (f), which guarantees that the locally nonrotating frames rotate with the same sense as Ω, and property (g), which under certain conditions implies that the sign of the angular momentum density agrees with the sign of Ω.

Property (a): The eroregion is disjoint from the axis. The proof for the rigidly rotating case given in the last section extends without modification.

Property (b): Any connected component of the ergoregion is simply connected and must interest a finite portion of fluid. The proof follows from Eq. (2.9) with $A^\alpha = (1, 0, 0)$ and $B^\alpha = (0, 0, 1)$,

$$D^m[\tau^{-1}\lambda_{11}^2 D_m(\lambda_{00}/\lambda_{11})] = 8\pi(\mu + p)\tau\psi^{-1}(\Omega^2\lambda_{11} - \lambda_{00}).$$

A Hopf argument based upon this equation implies that the region $\lambda_{00} \geq 0$ can neither be surrounded by a closed curve lying in the vacuum region nor surround a closed curve. But by Property (a), we have $\lambda_{00} < 0$ on the axis and at infinity $\lambda_{00} = -1$. Hence the ergoregion must have a finite intersection with the fluid.

Property (c): $\lambda_{01} \neq 0$ in the ergoregion. This follows directly from $\frac{1}{2}\tau^2 = \lambda_{01}^2 - \lambda_{00}\lambda_{11}$.

Property (d): Inside matter, $\lambda_{00} \leq 0$ implies that $\nu \equiv \lambda_{00} + \Omega\lambda_{01} < 0$.

Property (e): Inside matter, $\lambda_{00} \geq 0$ implies $\lambda_{01} < 0$. The last two properties follow directly from the condition $\psi < 0$ inside matter.

Property (f): $\Omega > 0$ implies $\lambda_{01} < 0$, excluding the axis. We first establish this property for region $\mathcal{A}$, defined by $\lambda_{00} < 0$. The boundary of $\mathcal{A}$ consists of the axis for which $\lambda_{01}/\lambda_{00} = 0$, infinity for which $\lambda_{01}/\lambda_{00} = 0$, and the curve $\lambda_{00} = 0$ for which $\lambda_{01}/\lambda_{00} \to \infty$ on approach from the interior of $\mathcal{A}$ according to Properties (b), (c), and (e). Now consider Eq. (2.9) with $A^\alpha = (0, 1, 0)$ and $B^\alpha = (1, 0, 0)$,

$$D^m[\tau^{-1}\lambda_{00}^2 D_m(\lambda_{01}/\lambda_{00})] = -8\pi(\mu + p)\tau\Omega\nu\psi^{-1}.$$

Property (d) implies that the right-hand side of this equation is negative in region $\mathcal{A}$. Hence the Hopf theorem implies that $\lambda_{01}/\lambda_{00}$ cannot have an internal minimum unless $\lambda_{01}/\lambda_{00}$ is constant throughout $\mathcal{A}$. The boundary conditions thus require either $\lambda_{01}/\lambda_{00} > 0$ or $\lambda_{01}/\lambda_{00} = 0$ throughout $\mathcal{A}$. The possibility $\lambda_{01}/\lambda_{00} = 0$ can be ruled out by direct inspection of the above equation. In the portion of the ergoregion inside matter, Property (e) implies $\lambda_{01} < 0$. To complete the proof, consider Eq. (2.9) with $A^\alpha = (0, 1, 0)$ and $B^\alpha = (0, 0, 1)$ in the portion of the ergoregion outside matter

$$D^m[\tau^{-1}\lambda_{11}^2 D_m(\lambda_{01}/\lambda_{11})] = 0.$$

A Hopf argument based upon this equation implies $\lambda_{01} < 0$ in this region.

Property (f) establishes that the angular velocity $\omega = -\lambda_{01}/\lambda_{11}$ of a locally nonrotating observer is positive everywhere except on the axis. We now consider the sign of the angular momentum density. For fluids with differential rotation, S^α and N^α are no longer constant; thus the argument leading to Eq. (4.3) of Paper I, from which the positivity of the angular momentum density was derived, is no longer valid. In fact, we have

$$D_a S^\alpha = 2N^\alpha D_a \Omega \tag{4.1}$$

and

$$D_a N^\alpha = P^\alpha D_a \Omega, \tag{4.2}$$

where $P^\alpha \lambda_\alpha = \lambda_{11}$. We shall require the equation of hydrostatic support, Eq. (2.7). Using Eq. (4.1), this gives

$$(\eta/\psi) D_a \Omega = \tfrac{1}{2}\psi^{-1} D_a \psi + (\mu + p)^{-1} D_a p. \tag{4.3}$$

We now assume that the fluid is governed by an equation of state $\mu = \mu(p)$; then the right-hand side of Eq. (4.3) is a gradient. It follows that

$$D_{[a}(\eta/\psi) D_{b]} \Omega = 0 \tag{4.4}$$

so that Ω is functionally related to η/ψ. To simplify comparison with the Newtonian limit (see Sec. V), we express this functional dependence in the form

$$\Omega = F(-\eta/\Omega\psi) \tag{4.5}$$

with

$$D_m \Omega = -F' D_m(\eta/\Omega\psi), \tag{4.6}$$

where a prime denotes the differentiation of F with respect to its argument.

An equation to which the Hopf theorem may be applied is obtained from Eq. (2.9) with $A^\alpha = N^\alpha$ and $B^\alpha = S^\alpha$,

$$S^{[\alpha} N^{\beta]} D^m [\tau^{-1} \lambda_\alpha D_m \lambda_\beta] = 0. \tag{4.7}$$

By pushing S^α and N^α through the differentiations and using Eqs. (4.1), (4.2), and (4.6), we can rewrite Eq. (4.7) in the form

$$D^m[X D_m(\eta/\Omega\psi)] + Y^m D_m(\eta/\Omega\psi) = 0, \tag{4.8}$$

where

$$X = \tau^{-1}[\Omega\psi^2 - (2\eta^2 + \Omega^{-1}\lambda_{01}\psi)F'] \tag{4.9}$$

and

$$Y^m = \tau^{-1}(\psi D^m \lambda_{11} - \lambda_{11} D^m \psi + 2\lambda_{11}\eta D^m \Omega) F'. \tag{4.10}$$

Property (g): For a differentially rotating fluid ($\Omega > 0$) satisfying an equation of state $\mu = \mu(p)$, if X is a positive C^1 scalar field and Y^m is a C^0 vector field (excluding the axis) then $\eta/\Omega\psi$ must attain its extrema on the boundary of the fluid. The proof follows immediately from the Hopf theorem, which is applicable in the interior of the fluid according to the hypothesis.

Property (h): For a fluid satisfying the hypothesis of Property (g) and for which the density is nowhere zero (but decreases at large distance in accord with asymptotic flatness), the angular momentum density is nonnegative. This follows as a corollary to Property (g) since the asymptotic conditions imply $\eta \geq 0$ and $\psi < 0$ at infinity.

V. DISCUSSION

Property (g), of the last section, establishes conditions on $F(-\eta/\Omega\psi)$ under which the angular momentum density must be negative on the surface of the fluid if it is negative anywhere in the interior. In order to test the physical reasonableness of these conditions it is instructive to examine the Newtonian limit. For differentially rotating fluids, F is a state function which must be specified to determine a stationary configuration. In the Newtonian limit, Eq. (4.5) reduces to the standard form[5] $\Omega = F(\rho^2)$, where ρ is the distance from the rotation axis. The smoothness conditions on F imposed by the hypothesis of Property (g) are thus completely justified in this limit. Furthermore, the positivity condition on X reduces to the extraneous condition $\Omega > 0$. In fact, in this limit, the content of Property (g) reduces to the obvious statement that the moment arm of points in the fluid about the axis attains its extrema on the boundary of the fluid. Only in case of extremely nonrigid and extremely relativistic rotation could the conditions on F necessary for Property (g) be violated.

We have not found any way to exploit Property (g) to show that the angular momentum denstiy must be positive for a finite fluid. The difficulties in attempting to formulate such a proof are associated with the breakdown of the concept of the velocity of light curve, which plays a key role in the proof for the rigid case in Paper I. This breakdown arises because in the differentially rotating case there is apparently no natural way to define Ω in the vacuum region outside the fluid.

In establishing Property (h) we were able to circumvent this difficulty by using boundary conditions at infinity. However, it is not clear how cogent this result is in practice. Along these lines, it is interesting to note that if the condition of a fluid equation of state $\mu = \mu(p)$ were removed it would then be possible to construct examples of systems with $\Omega > 0$ for which the angular momentum density assumes negative values. For instance, consider a rapidly rotating compact body surrounded in the equatorial plane by a much less massive thin ring. The angular velocity of a locally nonrotating observer has the asymptotic behavior $\omega \sim \rho^{-3}$. Thus the ring can be positioned in a state of small stress with angular velocity $0 < \Omega < \omega$, so that the angular momentum density is negative inside the ring.

We conclude with an application of Property (f) to systems with ergoregions.[6] In the ergoregion, there exist particle orbits of negative total energy. If by some process such as decay of an unstable particle[7] or pair production a particle is created in a negative energy state in the ergoregion, it will be trapped there, and so contribute to the total mass and angular momentum of the system. The contribution to the mass is by hypothesis negative; we shall show, via the inequality $\lambda_{01} < 0$, that the same is true of the contribution to the total angular momentum.

Let ζ^a be the unit future-directed, timelike linear combination of ξ_0^a and ξ_1^a which is everywhere orthogonal to ξ_1^a, i.e., ζ^a is given by

$$\zeta^a = \sqrt{2}\,\tau^{-1}\lambda_{11}^{1/2}(\xi_0^a - \lambda_{01}\lambda_{11}^{-1}\xi_1^a). \tag{5.1}$$

(ξ^a is the locally nonrotating observer field.) The velocity v^a of a particle can be written in the form

$$v^a = \gamma^{-1}(\xi^a + \sigma\xi_1^a + \mu^a), \tag{5.2}$$

where $\gamma > 0$, since v^a is future-directed, and where $\mu^m\xi_m = \mu^m\xi_{1m} = 0$. The particle energy (per unit mass) is given by

$$-e = v^m\xi_{0m} = \gamma^{-1}(-\sqrt{1/2}\,\tau\lambda_{11}^{-1/2} + \sigma\lambda_{01}), \tag{5.3}$$

since the particle is assume to have negative total energy, $(-e) > 0$. Solving Eq. (5.3) for σ, we obtain

$$\sigma = \lambda_{01}^{-1}(-e\gamma + \sqrt{1/2}\,\tau\lambda_{11}^{1/2}). \tag{5.4}$$

The angular momentum of the particle is

$$j = v^m\xi_{1m} = \gamma^{-1}\sigma\lambda_{11}$$
$$= \lambda_{11}\lambda_{01}^{-1}[-e + \sqrt{1/2}\gamma^{-1}\tau\lambda_{11}^{1/2}]. \tag{5.5}$$

Now, since $\lambda_{01} < 0$ in the ergoregion, Eq. (5.5) implies that $j < 0$. Particles created in states of negative total energy must therefore tend to decrease the angular momentum of the system.

ACKNOWLEDGMENT

We thank A. Ashtekar for helpful discussions and for pointing out the result concerning particles in negative energy states discussed in Sec. V.

*This research was supported in part by Grant No. B/RG/5134.4 from the Science Research Council and Grant No. MPS74 18020 from the National Science Foundation.

†Present address: Department of Mathematics, University of California, Berkeley, CA 94720, USA.

[1]R.O. Hansen and J. Winicour, J. Math. Phys. **16**, 804 (1975).

[2]J. Friedman and B.Schutz, "Gravitational Radiation and Instability in Rotating Stars," University of Cardiff preprint (1975).

[3]J.B. Hartle and D.H. Sharp, Astrophys. J. **147**, 317 (1967).

[4]See, e.g., K. Yano and S. Bochner, *Curvature and Betti Number* (Princeton U.P., Princeton, N.J., 1953), p. 26.

[5]See, e.g., R. Stoeckly, Astrophys. J. **142**, 208 (1965).

[6]A. Ashtekar, private communication.

[7]R. Penrose and R.M. Floyd, Nature **229**, 177 (1971).

Inequivalent sets of commuting missing label operators for SU(4)⊃SU(2)×SU(2)

C. Quesne*

Physique Théorique et Mathématique, Université Libre de Bruxelles, Brussels, Belgium
(Received 7 December 1976)

We exhibit one possible choice for the four functionally independent label operators available. We prove that they can be separated into two inequivalent, i.e., functionally independent, sets of commuting label operators, namely the set of operators Ω and Φ first considered by Moshinsky and Nagel, and the set of operators $C^{(202)}$ and $C^{(022)}$.

1. INTRODUCTION

When reducing the irreducible representations (IR's) of a semisimple group into IR's of a semisimple subgroup, it often happens that the latter does not provide enough labels to specify the basis states completely. As is well known, one way of solving this problem is to use as basis the common eigenstates of a complete set of commuting Hermitian operators. Besides the Casimir operators of group and subgroup, such a set must include some missing label operators that remain to be found. They should be subgroup scalars belonging to the enveloping algebra of the group. The determination of all possible labeling operators is a fundamental problem when studying any noncanonical chain of groups.

Recently this problem has been solved at least in principle. A general procedure for constructing an integrity basis of the subgroup scalars in the enveloping algebra of a group has been proposed and worked out in all cases where there is one missing label.[1,2] The chain SU(4)⊃SU(2)×SU(2), for which there are two missing labels, has also been investigated along these lines.[2-4] Explicit forms of possible basic scalars have been given in Ref. 4 for this case.

Since the completion of this work, Peccia and Sharp have shown that for any semisimple group and subgroup, the number of functionally independent label operators available is just twice the number of missing labels.[5] Therefore, it has seemed interesting to us to reexamine the results of Ref. 4 in the light of this new theorem.

In Sec. 2 we exhibit one possible choice for the four functionally independent label operators available among the scalars of the integrity basis found in Ref. 4. We then prove in Sec. 3 that from these operators we can form two inequivalent, i.e., functionally independent, sets of commuting label operators.

2. FUNCTIONALLY INDEPENDENT LABEL OPERATORS

It is shown in Ref. 4 that an integrity basis for the SU(2)×SU(2) invariants belonging to the enveloping algebra of $\mathcal{SU}(4)$ includes seven independent operators besides the five Casimir operators G_2, G_3, G_4, $\mathbf{S}^2$ and $\mathbf{T}^2$ of SU(4) and SU(2)×SU(2). For those additional operators we can choose

$$\begin{aligned} C^{(111)} &= S_i T_\alpha Q_{i\alpha}, \\ C^{(202)} &= S_i S_j Q_{i\alpha} Q_{j\alpha}, \\ C^{(022)} &= T_\alpha T_\beta Q_{i\alpha} Q_{i\beta}, \\ C^{(112)} &= \epsilon_{ijk} \epsilon_{\alpha\beta\gamma} S_i T_\alpha Q_{j\beta} Q_{k\gamma}, \end{aligned} \tag{2.1}$$

and

$$\begin{aligned} C^{(113)} &= S_i T_\alpha Q_{i\beta} Q_{j\alpha} Q_{j\beta}, \\ C^{(204)} &= S_i S_j Q_{i\alpha} Q_{j\beta} Q_{k\alpha} Q_{k\beta}, \\ C^{(024)} &= T_\alpha T_\beta Q_{i\alpha} Q_{j\beta} Q_{i\gamma} Q_{j\gamma}. \end{aligned} \tag{2.2}$$

On the other hand, from the result of Peccia and Sharp,[5] we can infer that among these twelve invariants there are only nine functionally independent operators, namely the five Casimir operators plus four additional operators. We now proposed to prove that the latter can be chosen as $C^{(111)}$, $C^{(202)}$, $C^{(022)}$, and $C^{(112)}$.

As the Casimir operators can be expressed in terms of the operators (2.1) and the following five operators[4]:

$$\begin{aligned} C^{(200)} &= S_i S_i, \\ C^{(020)} &= T_\alpha T_\alpha, \\ C^{(002)} &= Q_{i\alpha} Q_{i\alpha}, \\ C^{(003)} &= \epsilon_{ijk} \epsilon_{\alpha\beta\gamma} Q_{i\alpha} Q_{j\beta} Q_{k\gamma}, \\ C^{(004)} &= Q_{i\alpha} Q_{i\beta} Q_{j\alpha} Q_{j\beta}, \end{aligned} \tag{2.3}$$

it is sufficient to show that the operators (2.1) and (2.3) are functionally independent. For that purpose we require the operators S_i, T_α, $Q_{i\alpha}$ by ordinary variables s_i, t_α, $q_{i\alpha}$, transforming in the same way under SU(2)×SU(2), and converting therefore the operators $C^{(ABC)}$ into ordinary polynomials.[4] It is then enough[6] to prove the functional independence of these polynomials.

To check this point, let us show that the Jacobian of these nine polynomials with respect to a set of nine independent variables, for instance s_1, s_2, s_3, t_1, t_2, t_3, q_{11}, q_{22}, and q_{33}, is not identically zero,

$$\frac{\partial(C^{(200)}C^{(020)}C^{(002)}C^{(111)}C^{(003)}C^{(202)}C^{(022)}C^{(112)}C^{(004)})}{\partial(s_1 s_2 s_3 t_1 t_2 t_3 q_{11} q_{22} q_{33})} \neq 0. \tag{2.4}$$

We get straightforward that the term of this Jacobian, which is of degree zero in the variables $q_{i\alpha}$ with $i \neq \alpha$, is equal to

$$1536[(s_2^2t_3^2 - s_3^2t_2^2)s_1t_1q_{11} + (s_3^2t_1^2 - s_1^2t_3^2)s_2t_2q_{22}$$
$$+ (s_1^2t_2^2 - s_2^2t_1^2)s_3t_3q_{33}](q_{11}^2 - q_{22}^2)^2(q_{22}^2 - q_{33}^2)^2(q_{33}^2 - q_{11}^2)^2. \quad (2.5)$$

This establishes the truth of Eq. (2.4).

3. INEQUIVALENT SETS OF COMMUTING LABEL OPERATORS

To solve the state labeling problem for SU(4)⊃SU(2)×SU(2), it remains now to construct a set of two commuting label operators from the four label operators available, namely $C^{(111)}$, $C^{(202)}$, $C^{(022)}$, and $C^{(112)}$. It is obvious that there are many such sets because if we have got two commuting label operators, O_1 and O_2, any two independent functions of them will do as well. Functional dependent sets of commuting label operators are however uninteresting, and will therefore be considered in the following as equivalent. We proceed now to show that the four label operators available can be separated into two inequivalent sets of commuting label operators.

For that purpose, let us look for all the independent solutions of the equation

$$[O_1, O_2] = 0, \quad (3.1)$$

where

$$O_i = \alpha_i C^{(111)} + \beta_i C^{(202)} + \gamma_i C^{(022)} + \delta_i C^{(112)}, \quad i = 1,2, \quad (3.2)$$

and α_i, β_i, γ_i, and δ_i are some numerical coefficients to be determined. The commutators of the operators $C^{(ABC)}$ can be easily calculated from those of the generators of SU(4),[4] and are given by the following relations:

$$\begin{aligned}
&[C^{(111)}, C^{(202)}] = -2iC^{(213)} + \text{l.o.t.},\\
&[C^{(111)}, C^{(022)}] = -2iC^{(123)} + \text{l.o.t.},\\
&[C^{(111)}, C^{(112)}] = [C^{(111)}, C^{(202)}] + [C^{(111)}, C^{(022)}], \quad (3.3)\\
&[C^{(202)}, C^{(022)}] = 0,\\
&[C^{(202)}, C^{(112)}] = -4iC^{(214)} + \text{l.o.t.},\\
&[C^{(022)}, C^{(112)}] = -4iC^{(124)} + \text{l.o.t.},
\end{aligned}$$

where $C^{(213)}$, $C^{(123)}$, $C^{(214)}$, and $C^{(124)}$ are SU(2)×SU(2) invariants whose explicit expressions are given in Ref. 4, and l.o.t. denotes lower order terms, i.e., invariants of degree lower than that of the first term. There are therefore four linearly independent commutators, namely those of $C^{(111)}$ and $C^{(112)}$ with $C^{(202)}$ and $C^{(022)}$. Equating their coefficients in Eq. (3.1) to zero, we get four equations in α_i, β_i, γ_i, and δ_i,

$$\begin{aligned}
&\alpha_1(\beta_2 + \delta_2) = \alpha_2(\beta_1 + \delta_1),\\
&\alpha_1(\gamma_2 + \delta_2) = \alpha_2(\gamma_1 + \delta_1),\\
&\beta_1\delta_2 = \beta_2\delta_1,\\
&\gamma_1\delta_2 = \gamma_2\delta_1.
\end{aligned} \quad (3.4)$$

It is straightforward to show that this system of equations has exactly two independent solutions, leading to the following two inequivalent sets of commuting label operators:

$$O_1^{(1)} = C^{(111)}, \quad O_2^{(1)} = C^{(202)} + C^{(022)} - C^{(112)},$$

and (3.5)

$$O_1^{(2)} = C^{(202)}, \quad O_2^{(2)} = C^{(022)}.$$

The operators (3.5) coincide with the operators Ω and Φ, which were first considered by Moshinsky and Nagel,[8] and diagonalized simultaneously in Ref. 4.

*Maitre de recherches F.N.R.S.

[1]B.R. Judd, W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. 15, 1787 (1974).
[2]R.T. Sharp, J. Math. Phys. 16, 2050 (1975).
[3]W. Miller, Jr., unpublished.
[4]C. Quesne, J. Math. Phys. 17, 1452 (1976).
[5]A. Peccia and R.T. Sharp, J. Math. Phys. 17, 1313 (1976).
[6]In fact the operator invariants corresponding to polynomial invariants should be obtained from them by replacing s_i, t_α, $q_{i\alpha}$ by S_i, T_α, $Q_{i\alpha}$, and then symmetrizing as to order.[5,7] The lack of symmetrization of the operators we consider here does not affect the proof as they differ from the corresponding symmetrized operators only by invariants of lower degree.
[7]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 17, 986 (1976).
[8]M. Moshinsky and J.G. Nagel, Phys. Lett. 5, 173 (1963).

Evolution equations possessing infinitely many symmetries

Peter J. Olver

Department of Mathematics, University of Chicago, Chicago, Illinois 60637
(Received 23 November 1976)

A general method for finding evolution equations having infinitely many symmetries or flows which preserve them is described. This is applied to the Korteweg–de Vries, modified Korteweg–de Vries, Burgers', and sine–Gordon equations.

The intense research activity of the past ten years surrounding the Korteweg–de Vries (KdV) equation was initially motivated by the discovery of an infinite series of conservation laws.[1] Noether's theorem shows that for a partial differential equation in Lagrangian form (which the KdV equation can be put into) there is an intimate connection between one-parameter symmetry groups of the equation[2] and conservation laws.[3] This was mysterious since the KdV equation possessed only a four-parameter symmetry group. However, it was noticed that the higher order analogs of the KdV equation discovered by Gardner[4] could be reinterpreted as "higher order symmetries" of the equation,[5] shedding some light on the mystery. Thus the more immediate object of interest becomes the symmetry groups, or, in more traditional terminology, the evolution equations whose flows preserve the KdV equation. The advantage of this point of view is that the symmetry groups can be systematically found, as in Theorem 1, in contrast to the ad hoc methods used to discover conservation laws. A recursion formula due to Lenard[4] for the higher order KdV equations is generalized here to provide a method for the construction of infinite series of higher order symmetries of more general evolution equations. In particular, we derive in Example 4 an infinite series of flows, all of which preserve Burgers' equation. (These symmetries however do not give rise to conservation laws since Burgers' equation cannot be placed in Lagrangian form. The precise relationship between symmetries and conservation laws shall be discussed in a future paper.) The methods employed here are differential algebraic in nature, in the same spirit as the recent work of Gelfand and Dikiĭ.[6] The calculations presented in this note will be formal; rigorous mathematical statements and proofs shall appear elsewhere in a more complete exposition.

Let $R\{u\}$ denote the algebra of polynomials in the variable u and its derivatives with respect to a single independent variable x. Given a polynomial $P\in R\{u\}$, let $\{P\}$ denote the differential ideal generated by P; it consists of all polynomials of the form $\sum Q_k D^k P$, where $D=d/dx$, the total derivative. For convenience we abbreviate

$$u_i=\frac{d^i u}{dx^i}, \quad \partial_i=\frac{\partial}{\partial u_i}.$$

Let T be the vector space of all formal polynomial partial differential operators acting on $R\{u\}$; in other words, T consists of all operators of the form

$$\mathcal{D}=\sum_I P_I\partial_I,$$

where $P_I\in R\{u\}$. Here the sum is over all multi-indices $I=(i_0, i_1, \dots, i_l)$ with $l=0, 1, 2, \cdots$, and $\partial_I=\partial_0^{i_0}\partial_1^{i_1}\cdots\partial_l^{i_l}$. T contains the total derivative operator

$$\frac{d}{dx}=D=\sum_{i=0}^{\infty}u_{i+1}\partial_i. \tag{1}$$

Since T acts on $R\{u\}$, we make T into an algebra by defining the product of $\mathcal{D}, \mathcal{D}'\in T$ to be

$$\mathcal{D}\cdot\mathcal{D}'(P)=\mathcal{D}[\mathcal{D}'(P)], \quad P\in R\{u\}.$$

Using Leibnitz' formula, if $\mathcal{D}=\sum P_I\partial_I$ and $\mathcal{D}'=\sum Q_J\partial_J$, then

$$\mathcal{D}\cdot\mathcal{D}'=\sum_{I,J}P_I\sum_{0\le M\le I}\binom{I}{M}\partial_M Q_J\cdot\partial_{I+J-M}. \tag{2}$$

Moreover, T is a Lie algebra with bracket

$$[\mathcal{D},\mathcal{D}']=\mathcal{D}\mathcal{D}'-\mathcal{D}'\mathcal{D}.$$

Now let

$$\mathcal{V}: R\{u\}\to T$$

be the map defined by

$$\mathcal{V}(P)=\sum_{j=0}^{\infty}D^j P\cdot\partial_j. \tag{3}$$

Note that in Ovsjannikov's terminology,[2] $\mathcal{V}(P)$ is the ∞-prolongation of the vector field $P\cdot\partial/\partial u$, in the case P is a polynomial in u alone.

Given a differential polynomial $K\in R\{u\}$, consider the evolution equation

$$u_t=K(u). \tag{4}$$

If we make the assumption that (4) is locally uniquely solvable for arbitrary smooth initial data $u(x,0)=f(x)$, then there is an induced flow on $C^\infty(\mathbb{R})$,

$$\hat{K}_t[f(x)]=u(x,t), \quad f\in C^\infty(\mathbb{R}),$$

where $u(x,t)$ is the solution of (4) with initial data f. If $P\in R\{u\}$ is another polynomial, then we say that the flow generated by P preserves the flow generated by K if

$$\hat{P}_s[\hat{K}_t(f)]=\hat{K}_t[\hat{P}_s(f)]$$

for all $f\in C^\infty(\mathbb{R})$ and all $s,t\in\mathbb{R}$ where the equation is defined.

Theorem 1. Let $P,K\in R\{u\}$. Then the flow generated by P preserves the flow generated by K if and only if

$$\mathcal{V}(P)K-D_t K\in\{u_t-K\}. \tag{5}$$

Condition (5) refers to the partial differential algebra $\hat{R}\{u\}$ consisting of polynomials in u and its partial derivatives with respect to both x and t; $D_t=d/dt$ is the

total derivative with respect to t. To verify condition (5) it suffices to replace the variables $\partial^{j+1}u/\partial^j x\partial t$ in the left-hand side by D^jK and equate the resulting expression to 0. In the special case that P is a polynomial in u alone this result is well known; it is just the infinitesimal criterion of invariance of (4) under the one-parameter group with infinitesimal generator $P\cdot\partial/\partial u$.[2] Thus if P, K satisfy criterion (5) we shall say that P is an *infinitesimal higher-order symmetry* of K. Note that $P=K$ trivially satisfies (5).

Next, define the map

$$\mathcal{A}: R\{u\} \to T$$

by

$$\mathcal{A}(P)=\sum_{j=0}^{\infty}\partial_j P\cdot D^j. \tag{6}$$

Note that

$$\mathcal{V}(P)K=\mathcal{A}(K)P \tag{7}$$

for P, $K\in R\{u\}$. Let T_0 be the subspace of T generated by the operator D, i.e., the elements of T_0 are operators of the form $\sum_{i=0}^{\infty}P_iD^i$ with $P_i\in R\{u\}$. Note that T_0 preserves ideals in $R\{u\}$.

Theorem 2: Suppose $\mathcal{D}\in T_0$ satisfies

$$[\mathcal{A}(K)-D_t, \mathcal{D}]P\in\{u_t-K\} \tag{8}$$

for all $P\in R\{u\}$, then K possesses an infinite series of infinitesimal symmetries

$$K^{(j)}=\mathcal{D}^jK, \quad j=0, 1, 2, \cdots. \tag{9}$$

Proof: By induction on j and Eq. (7)

$$[\mathcal{A}(K)-D_t]K^{(j-1)}\in\{u_t-K\}.$$

Using $\equiv$ to denote congruence modulo the ideal $\{u_t-K\}$, condition (8) implies that

$$\begin{aligned}[\mathcal{A}(K)-D_t]K^{(j)}&=[\mathcal{A}(K)-D_t]\mathcal{D}K^{(j-1)}\\&\equiv\mathcal{D}[\mathcal{A}(K)-D_t]K^{(j-1)}\\&\equiv 0,\end{aligned}$$

thereby proving the result.

An operator $\mathcal{D}\in T_0$ that satisfies condition (8) will be called a *recursion operator* for K. Practically, to verify the condition that $\mathcal{D}$ be a recursion operator for K it suffices to (a) compute the Lie bracket in $R\{u\}$, (b) substitute D^jK whenever the variable $\partial^{j+1}u/\partial x^j\partial t$ appears, and (c) equate the resulting coefficients of D^k for each power k to 0. Thus (8) gives a useful criterion for determining when an evolution equation possesses an infinite number of symmetries.

Actually, to apply Theorem 2 to any interesting equations, it is necessary to enlarge the class T_0 to include more general recursion operators. In particular, we wish to allow recursion operators that involve the inverse total derivative D^{-1}.[7] The problem is that D^{-1} is not well defined on all of $R\{u\}$, so more care in the assumptions is needed to ensure that the conclusions of Theorem 2 still hold for these more general operators. A precise statement of the generalization of this theorem will be deferred to the later rigorous exposition. In this note we shall be content to use condition (8) formally to find a few specific recursion operators.

Example 3: Consider the KdV equation

$$u_t=K(u)=u_{xxx}+uu_x. \tag{10}$$

Here we reprove the result of Lenard that the operator

$$\mathcal{D}=D^2+\tfrac{2}{3}u+\tfrac{1}{3}u_xD^{-1} \tag{11}$$

is (formally) a recursion operator for K. Now

$$\mathcal{A}(K)=A=D^3+uD+u_x.$$

Hence

$$\begin{aligned}A\cdot\mathcal{D}=D^5&+\tfrac{5}{3}uD^3+\tfrac{10}{3}u_xD^2+(3u_{xx}+\tfrac{2}{3}u^2)D\\&+\tfrac{5}{3}(u_{xxx}+uu_x)+\tfrac{1}{3}(u_{xxxx}+uu_{xx}+u_x^2)D^{-1}\end{aligned}$$

and

$$\begin{aligned}\mathcal{D}\cdot A=D^5&+\tfrac{5}{3}uD^3+\tfrac{10}{3}u_xD^2\\&+(3u_{xx}+\tfrac{2}{3}u^2)D+(u_{xxx}+uu_x).\end{aligned}$$

Therefore,

$$[A,\mathcal{D}]=\tfrac{2}{3}(u_{xxx}+uu_x)+\tfrac{1}{3}(u_{xxxx}+uu_{xx}+u_x^2)D^{-1}.$$

Furthermore

$$[D_t,\mathcal{D}]=\tfrac{2}{3}u_t+\tfrac{1}{3}u_{xt}D^{-1},$$

so that condition (8) is verified. The infinite series of symmetries

$$K^{(j)}=\mathcal{D}^jK,$$

when put into evolution form

$$u_t=K^{(j)}(u),$$

are just the higher-order analogs of the KdV equation.[4]

Example 4: Consider Burgers' equation

$$u_t=B(u)=u_{xx}+uu_x. \tag{12}$$

We show that B possesses the recursion operator

$$\mathcal{D}=D+\tfrac{1}{2}u+\tfrac{1}{2}u_xD^{-1}. \tag{13}$$

Here

$$\mathcal{A}(B)=A=D^2+uD+u_x.$$

Hence

$$\begin{aligned}A\cdot\mathcal{D}=\ D^3&+\tfrac{3}{2}uD^2+(\tfrac{5}{2}u_x+\tfrac{1}{2}u^2)D+\tfrac{3}{2}(u_{xx}+uu_x)\\&+\tfrac{1}{2}(u_{xxx}+uu_{xx}+u_x^2)D^{-1}\end{aligned}$$

and

$$\mathcal{D}\cdot A=D^3+\tfrac{3}{2}uD^2+(\tfrac{5}{2}u_x+\tfrac{1}{2}u^2)D+(u_{xx}+uu_x).$$

Therefore,

$$[A,\mathcal{D}]=\tfrac{1}{2}(u_{xx}+uu_x)+\tfrac{1}{2}(u_{xxx}+uu_{xx}+u_x^2)D^{-1}.$$

Furthermore

$$[D_t,\mathcal{D}]=\tfrac{1}{2}u_t+\tfrac{1}{2}u_{xt}D^{-1},$$

which proves condition (8) formally. Therefore, we have an infinite sequence of flows

$$u_t=B^{(j)}(u)=\mathcal{D}^jB(u)$$

all of which preserve the flow given by Burgers' equa-

tion. The first few of these flows are

$$
\begin{aligned}
u_t &= B^{(0)}(u) = u_{xx} + uu_x,\\
u_t &= B^{(1)}(u) = u_{xxx} + \tfrac{3}{2}uu_{xx} + \tfrac{3}{2}u_x^2 + \tfrac{3}{4}u^2u_x,\\
u_t &= B^{(2)}(u) = u_{xxxx} + 2uu_{xxx} + 5u_xu_{xx}\\
&\quad + \tfrac{3}{2}u^2u_{xx} + 3uu_x^2 + \tfrac{1}{2}u^3u_x,\\
u_t &= B^{(3)}(u) = u_{xxxxx} + \tfrac{5}{2}uu_{xxxx} + \tfrac{15}{2}u_xu_{xxx} + 5u_{xx}^2\\
&\quad + \tfrac{5}{2}u^2u_{xxx} + \tfrac{25}{2}uu_xu_{xx} + \tfrac{15}{4}u_x^3\\
&\quad + \tfrac{5}{4}u^3u_{xx} + \tfrac{15}{4}u^2u_x^2 + \tfrac{5}{16}u^4u_x.
\end{aligned}
\tag{14}
$$

Example 5: Finally, we consider the modified KdV equation

$$u_t = \tilde{K}(u) = u_{xxx} + u^2u_x, \tag{15}$$

which is known to also possess infinitely many conservations laws. In fact, the original proof of the existence of infinitely many conservation laws of the KdV equation $u_t = K(u)$ stemmed from the remarkable transformation of Miura[8] relating the two equations. Explicitly, if

$$v = u^2 + \mu u_x, \quad \text{where } \mu = \sqrt{-6},$$

then

$$(\mu D + 2u)[u_t - \tilde{K}(u)] = v_t - K(v).$$

Let us assume for the moment that $\tilde{K}$ possesses a recursion operator $\tilde{D}$ and that furthermore the higher-order analogs $u_t = \tilde{K}^{(j)} = \tilde{D}_j\tilde{K}$ are related to the higher-order analogs of the KdV equation by the same formula,

$$(\mu D + 2u)[u_t - \tilde{K}^{(j)}(u)] = v_t - K^{(j)}(v).$$

We conclude that the recursion operator $\tilde{D}$ must be related to the recursion operator D of K by the formal operator equation

$$(\mu D + 2u)\cdot\tilde{D} = D\cdot(\mu D + 2u).$$

A straightforward calculation shows that for this to hold,

$$\tilde{D} = D^2 + \tfrac{2}{3}u^2 + \tfrac{2}{3}u_xD^{-1}\cdot u. \tag{16}$$

The last term in (16) is the operator which takes a polynomial $P \in R\{u\}$, multiplies it by u, then applies D^{-1}, and finally multiplies the result by $\frac{2}{3}u_x$. We shall check that $\tilde{D}$ is indeed a recursion operator for $\tilde{K}$. We have

$$\mathcal{A}(\tilde{K}) = \tilde{A} = D^3 + u^2D + 2uu_x.$$

Note that

$$D^{-1}\cdot u = uD^{-1} - D^{-1}\cdot u_xD^{-1}.$$

Hence

$$
\begin{aligned}
u_xD^{-1}\cdot u\tilde{A} &= uu_xD^2 + u^3u_x - u_xD^{-1}\cdot(u_xD^2 + u^2u_x)\\
&= uu_xD^2 - u_x^2D + (u^3u_x + u_xu_{xx})\\
&\quad - u_xD^{-1}\cdot(u_{xxx} + u^2u_x)
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{A}\cdot u_xD^{-1}\cdot u &= [u_xD^3 + 3u_{xx}D^2 + (3u_{xxx} + u^2u_x)D\\
&\quad + (u_{xxxx} + u^2u_{xx} + 2uu_x^2)]\cdot D^{-1}\cdot u\\
&= uu_xD^2 + (2u_x^2 + 3uu_{xx})D + (3uu_{xxx} + 4u_xu_{xx}\\
&\quad + u^3u_x) + (u_{xxxx} + u^2u_{xx} + 2uu_x^2)D^{-1}\cdot u.
\end{aligned}
$$

Hence

$$
\begin{aligned}
[\tilde{A}, \tfrac{2}{3}u_xD^{-1}\cdot u] &= 2(uu_{xx} + u_x^2)D + 2(uu_{xxx} + u_xu_{xx})\\
&\quad + \tfrac{2}{3}(u_{xxxx} + u^2u_{xx} + 2uu_x^2)D^{-1}\cdot u\\
&\quad + \tfrac{2}{3}u_xD^{-1}\cdot(u_{xxx} + u^2u_x).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
[\tilde{A}, \tilde{D}] &= \tfrac{4}{3}u(u_{xxx} + u^2u_x) + \tfrac{2}{3}(u_{xxxx} + u^2u_{xx}\\
&\quad + 2uu_x^2)D^{-1}\cdot u + \tfrac{2}{3}u_xD^{-1}\cdot(u_{xxx} + u^2u_x).
\end{aligned}
$$

On the other hand,

$$D_t\cdot u_xD^{-1}\cdot u = u_{xt}D^{-1}\cdot u + u_xD^{-1}\cdot u_t + u_xD^{-1}\cdot uD_t.$$

Hence

$$[D_t, \tilde{D}] = \tfrac{4}{3}uu_t + \tfrac{2}{3}u_{xt}D^{-1}\cdot u + \tfrac{2}{3}u_xD^{-1}\cdot u_t.$$

Comparing the expressions for $[\tilde{A}, \tilde{D}]$ and $[D_t, \tilde{D}]$ shows that condition (8) holds formally under the plausible assumption that $u_xD^{-1}\cdot u_t$ and $u_xD^{-1}\cdot(u_{xxx} + u^2u_x)$ define the same operator modulo the ideal $\{u_t - \tilde{K}(u)\}$. We conclude that the flows

$$u_t = \tilde{K}^{(j)}(u) = \tilde{D}^j\tilde{K}(u)$$

all preserve the modified KdV equation. The first few of these flows are

$$
\begin{aligned}
u_t &= K^{(0)}(u) = u_{xxx} + u^2u_x,\\
u_t &= K^{(1)}(u) = u_{xxxxx} + \tfrac{5}{3}u^2u_{xxx} + \tfrac{20}{3}uu_xu_{xx} + \tfrac{5}{3}u_x^3 + \tfrac{5}{6}u^4u_x,\\
u_t &= K^{(2)}(u) = u_{xxxxxxx} + \tfrac{7}{3}u^2u_{xxxxx} + 14uu_xu_{xxxx}\\
&\quad + 21u_x^2u_{xxx} + \tfrac{70}{3}uu_{xx}u_{xxx} + \tfrac{91}{3}u_xu_{xx}^2\\
&\quad + \tfrac{35}{18}u^4u_{xxx} + \tfrac{140}{9}u^3u_xu_{xx} + \tfrac{105}{9}u^3u_x^3 + \tfrac{35}{54}u^6u_x.
\end{aligned}
$$

Example 6. In this example we consider the sine-Gordon equation

$$u_{xt} = \sin u. \tag{17}$$

This equation is already known to possess infinitely many conservation laws and symmetry groups.[9] Although this equation does not belong to the class of evolution equations, we shall indicate how the methods used previously can be modified so as to rederive the symmetry groups of (17). The analog of Theorem 1 in this case is that the flow generated by a polynomial $P \in R\{u\}$ preserves the set of solutions of the sine-Gordon equation if and only if

$$A(P) \in \{u_{xt} - \sin u\}, \tag{18}$$

where A is the operator

$$A = DD_t - \cos u.$$

In this case we call P an (infinitesimal) symmetry of the sine-Gordon equation.

Suppose we can find operators D and D' satisfying

$$[AD - D'A]Q \in \{u_{xt} - \sin u\} \tag{19}$$

for any $Q \in R\{u\}$. Then condition (18) implies that DP is a symmetry of (17) whenever P is. In other words, D would be a recursion operator for the sine-Gordon equation. However, since (17) is not an evolution equation, we are left with the task of finding one symmetry of (17)

in order to prove that there are infintely many symmetries. This is simplified by the observation that any partial differenital equation not explicitly involving the independent variable x or t is invariant under the flow

$$u_t = u_x.$$

This amounts to the statement that $f(x+\lambda, t)$ is a solution whenever $f(x,t)$ is. [In our previous examples, if we apply the recursion operators to the polynomial u_x, the flow we derive is just that of the original equation. For instance,

$$(D^2 + \tfrac{2}{3}u + \tfrac{1}{3}u_x D^{-1})u_x = u_{xxx} + uu_x. \quad] \tag{20}$$

Now consider the operators

$$\mathcal{D} = D^2 + u_x^2 - u_x D^{-1} \cdot u_{xx}, \quad \mathcal{D}' = D^2 + u_x^2 + u_{xx} D^{-1} \cdot u_x. \tag{21}$$

We have

$$\begin{aligned} A\mathcal{D} = {} & D^3 D_t - \cos u D^2 + u_x^2 D D_t + u_x u_{xx} D_t + 2u_x u_{xt} D \\ & + (u_{xx}u_{xt} + u_x u_{xxt} - u_x^2 \cos u) + (u_x \cos u \\ & - u_{xxt}) D^{-1} \cdot u_{xx} - u_{xx} D^{-1} \cdot u_{xxt} - u_{xx} D^{-1} \cdot u_{xx} D_t, \end{aligned}$$

and

$$\begin{aligned} \mathcal{D}'A = {} & D^3 D_t - \cos u D^2 + u_x^2 D D_t + u_x u_{xx} D_t + 2u_x \sin u D \\ & + u_{xx} \sin u - u_{xx} D^{-1} \cdot u_x \cos u - u_{xx} D^{-1} \cdot u_{xx} D_t, \end{aligned}$$

where we have used the identity

$$D^{-1} \cdot u_x D = u_x - D^{-1} \cdot u_{xx}.$$

Comparing these expressions verifies condition (18) formally. We conclude that the flows

$$u_t = \mathcal{D}^k(u_x), \quad k = 0, 1, 2, \cdots$$

are all symmetries of the sine-Gordon equation. The first few of these flows are

$$u_t = u_x,$$

$$u_t = u_{xxx} + \tfrac{1}{2}u_x^3,$$

$$u_t = u_{xxxxx} + \tfrac{5}{2}u_x^2 u_{xxx} + \tfrac{5}{2}u_x u_{xx}^2 + \tfrac{3}{8}u_x^5,$$

$$\begin{aligned} u_t = {} & u_{xxxxxxx} + \tfrac{7}{2}u_x^2 u_{xxxxx} + 14 u_x u_{xx} u_{xxxx} + \tfrac{21}{2}u_x u_{xxx}^2 \\ & + \tfrac{35}{2}u_x^2 u_{xx} u_{xxx} + \tfrac{35}{8}u_x^4 u_{xxx} + \tfrac{35}{4}u_x^3 u_{xx}^2 + \tfrac{5}{16}u_x^7. \end{aligned}$$

[1] R. M. Miura *et al.*, J. Math. Phys. **9**, 1204 (1968).
[2] L. V. Ovsjannikov, *Group Properties of Differential Equations*, translated by G. W. Bluman (unpublished). See also P. J. Olver, thesis, Harvard University, 1976, for a rigorous exposition.
[3] N. H. Ibragimov, "Invariance and Conservation Laws of Continuum Mechanics," in *Symmetry, Similarity and Group Theoretic Methods in Mechanics*, edited by P. G. Glockner and M. C. Singh (University of Calgary Press, Calgary, 1974), pp. 63–82.
[4] Discussed in P. D. Lax, Commun. Pure Appl. Math. **28**, 141 (1975).
[5] R. L. Anderson, S. Kumei, and C. E. Wulfman, J. Math. Phys. **14**, 1527 (1973) and S. Kumei, J. Math. Phys. **16**, 2461 (1975) also considered this concept.
[6] I. M. Gelfand and L. A. Dikii, Russ. Math. Survey, **30**, 63 (1975); Func. Anal. **10**, 18 (1975).
[7] See M. D. Kruskal *et al.*, J. Math. Phys. **11**, 952 (1970), for a precise discussion of this operator.
[8] R. M. Miura, J. Math. Phys. **9**, 1202 (1968).
[9] S. Kumei, J. Math. Phys. **16**, 2461 (1975).

Symmetries of ultralocal quantum field theories*†

B. DeFacio

Ames Laboratory-ERDA, Iowa State University, Ames, Iowa 50011
and Department of Physics, University of Missouri-Columbia, Columbia, Missouri 65201‡

C. L. Hammer

Ames Laboratory-ERDA and Department of Physics, Iowa State University, Ames, Iowa 50011
(Received 2 August 1976)

The symmetries of the gradient free ultralocal model quantum field theories are studied. The internal parameter λ introduced by J. R. Klauder is replaced by an r-component vector and used to obtain an r-vector ultralocal field operator $\vec{\phi}$. Then ultralocal many-body models with translational and rotational symmetry are set up by partitioning $\vec{\lambda}$ into N, 3-d "relative internal coordinates." Hartree-type $1/N$ limits are studied in this model and found to be accurate only if contributions from $|\vec{\lambda}| > > 0$ are negligible. A brief sketch is given of how to produce more general U_N and pairing interaction ultralocal models.

1. INTRODUCTION

It is widely believed among physicists that quantum field theories describe the fundamental particles. The proof of such assertions, however, remains a long-standing problem—now approximately fifty years old. The two conceptual problems with quantum field theories are the infinite number of degrees of freedom and the nonlinear powers of the "fields" at a single space-time point.

The idea of symmetry plays a deep and profound role in both physics and mathematics. Systems with an infinite number of degrees of freedom are presently subject to so little mathematical control that their symmetries assume an especially important role. Any system with an infinite number of degrees of freedom which can be studied mathematically is worthy of attention. Since covariant models are a small subset of the set of all model field theoreis and in view of the scarcity of mathematically sound models, one need not apologize for isolating these problems by his choice of a model. The ultralocal models to be studied in this paper are one such class of models.

Ultralocal models were created by Wentzel,[1] and studied as a lattice model by Schiff.[2] After several inconclusive studies by other authors, Klauder developed an exponential Hilbert space approach,[3] which he then used in a series of definitive papers on ultralocal models,[4–8] nonrenormalizable models,[9,10] and independent value models.[11,12] Then Newman[13] used the Araki generating functional method[14] to provide a "current algebra" formulation of ultralocal models. Goldin[15] has given an incisive discussion of Newman's formulation which he relates to Klauder's approach. DeFacio and Hammer[16] have used a bilinear approach to reproduce Klauder's ultralocal models.

The simplest definition of an ultralocal quantum field theory is that it is a theory which is obtained from a covariant theory by dropping all spatial gradients. For a single time-zero scalar field $\phi(\mathbf{x})$, with $\mathbf{x} \in R^s$ and $(\mathbf{x}, t) = x \in R^{s+1}$, i.e., space is s-dimensional and space-time is $(s+1)$-dimensional, a covariant Hamiltonian H is given by

$$H = \int d\mathbf{x}\left[\tfrac{1}{2}(\partial_t\phi)^2 + \tfrac{1}{2}(\nabla\phi)^2 + \tfrac{1}{2}m_0^2(\phi)^2 + V[\phi(\mathbf{x})]\right], \quad (1.1)$$

whereas the corresponding ultralocal Hamiltonian H is given by

$$H = \int d\mathbf{x}\left[\tfrac{1}{2}(\partial_t\phi)^2 + \tfrac{1}{2}m_0^2(\phi)^2 + V[\phi(\mathbf{x})]\right]. \quad (1.2)$$

The reduction from Eq. (1.1) to Eq. (1.2) corresponds to the collapse of the light cone to a single timelike line at the spatial point $\mathbf{x}$. The most fundamental symmetry of ultralocal model field theories is statistical independence of distinct spatial points.[17] This symmetry plays a key role in formulating these models and, in particular, it allows the use of powerful probability methods.[18,19]

The problem to be discussed in this paper is the generalization of the one-component ultralocal field ϕ to $\vec{\phi}$, an r-component ultralocal field. The details of this construction are given in Secs. 2, 3, and 4. In Sec. 5 a many-body ultralocal model is presented, as an example, which exhibits translational and rotational symmetry. The Hartree approximation to the ultralocal field is also discussed in this section. In Sec. 6 ultralocal models with more general symmetries are discussed, and our conclusions are presented.

2. MORE GENERAL ULTRALOCAL MODELS

A straightforward generalization of Klauder's ultralocal models[4–8] occurs when the number of internal variables is increased:

$$\lambda \to \vec{\lambda} = (\lambda^\alpha). \quad (2.1)$$

We emphasize that $r = \dim(\vec{\lambda}) \neq s = \dim(\mathbf{x})$ as these vectors "live in" independent spaces. For these models, $A^*(\mathbf{x}, \vec{\lambda})$, $A(\mathbf{x}, \vec{\lambda})$ act as creation and destruction operators and together with the unique vacuum $|\Omega\rangle$ generate a Fock space H_F. The only nonzero commutator is

$$[A(\mathbf{x}, \vec{\lambda}), A^*(\mathbf{x}', \vec{\lambda}')] = \delta(\mathbf{x} - \mathbf{x}')\delta(\vec{\lambda} - \vec{\lambda}') \quad (2.2)$$

and

$$A(\mathbf{x}, \vec{\lambda})|\Omega\rangle = 0 \quad (2.3)$$

for all $\mathbf{x}$ and $\vec{\lambda}$. Thus, H_F is a Fock space in s space dimensions, together with r-dimensional independent internal coordinates The non-Fock operators $B(x, \vec{\lambda})$ and $B^*(\mathbf{x}, \vec{\lambda})$ are obtained from the translations[16]

$$\exp(-\Sigma_C)A(\mathbf{x}, \vec{\lambda}) \exp(\Sigma_C) \equiv B(\mathbf{x}, \vec{\lambda}) = A(\mathbf{x}, \vec{\lambda}) + C(\vec{\lambda})$$

and

$$\exp(-\Sigma_C)A^*(\mathbf{x}, \vec{\lambda}) \exp(\Sigma_C) \equiv B^*(\mathbf{x}, \vec{\lambda}) = A^*(\mathbf{x}, \vec{\lambda}) + C(\vec{\lambda}). \tag{2.4}$$

The translation generator Σ_ζ for a general function $\zeta(\mathbf{x}, \vec{\lambda})$ is given by[16]

$$\Sigma_\zeta = \sigma^*(\zeta) - \sigma(\zeta), \tag{2.5}$$

with

$$\sigma(\zeta) = \int d\mathbf{x} \int d\vec{\lambda} J_0[\zeta(\mathbf{x}, \vec{\lambda}), A(\mathbf{x}, \vec{\lambda})],$$
$$\sigma^*(\zeta) = \int d\mathbf{x} \int d\vec{\lambda} J_0[A(\mathbf{x}, \vec{\lambda}), \zeta(\mathbf{x}, \vec{\lambda})], \tag{2.6}$$

and, unless otherwise stated,

$$J_0(\zeta_1, \zeta_2) = \zeta_1^* \zeta_2.$$

As in Klauder's one component models,[4-12] the general model function is taken to be

$$C(\vec{\lambda}) = \exp[-y(\vec{\lambda})]/|f(\vec{\lambda})|^\gamma, \tag{2.7}$$

where $\lim f(\vec{\lambda}) - |\vec{\lambda}|^\gamma$ as $\vec{\lambda} \to 0$ is the only zero and $y(\vec{\lambda})$ is a polynomial of degree $2n$ $(n \geq 1)$. Also it is convenient to choose $C(\vec{\lambda}) = C(-\vec{\lambda})$.

The integrability conditions for $C(\lambda)$ are chosen as

$$\int d\vec{\lambda} C^2(\vec{\lambda}) \to \infty,$$
$$\int d\Omega_\lambda \int_0^\infty d\lambda [\lambda^{r+1}/(1+\lambda^{r+1})] C^2(\vec{\lambda}) < \infty, \tag{2.8}$$

where $\lambda = |\vec{\lambda}|$ and $d\Omega_\lambda$ is the differential solid angle about $\vec{\lambda}$. As a result of these restrictions the singularity parameter γ of Eq. (2.7) satisfies the inequalities

$$(r/2) + 1 > \gamma \geq (r/2).$$

The r-component, time-zero, configuration field is given by

$$\vec{\phi}(\mathbf{x}) = \begin{pmatrix} \phi_1(\mathbf{x}) \\ \cdot \\ \cdot \\ \cdot \\ \phi_r(\mathbf{x}) \end{pmatrix},$$

where

$$\phi_i(\mathbf{x}) = \int d\vec{\lambda} B^*(\mathbf{x}, \vec{\lambda}) \lambda_i B(\mathbf{x}, \vec{\lambda}). \tag{2.9}$$

It is a straightforward process to form higher rank tensors such as the second rank tensors

$$\phi_{ij}(\mathbf{x}) = \int d\vec{\lambda} B^*(\mathbf{x}, \vec{\lambda}) \lambda_i \lambda_j B(\mathbf{x}, \vec{\lambda}). \tag{2.10}$$

In general, an ultralocal field operator with l indices can be written as

$$\phi_{\alpha_1 \cdots \alpha_l}(\mathbf{x}) = \int d\vec{\lambda} B^*(\mathbf{x}, \vec{\lambda}) (\lambda_{\alpha_1} \cdots \lambda_{\alpha_l}) B(\mathbf{x}, \vec{\lambda}). \tag{2.11}$$

These tensor fields lead to a generalization of Klauder's renormalized fields[5]

$$\phi_r^p(\mathbf{x}) = \int d\lambda B^*(\mathbf{x}, \lambda) \lambda^p B(\mathbf{x}, \lambda),$$

where $p = 1, 2, \cdots$ and where formally

$$\phi_r^p(\mathbf{x}) = Z^{-1}[Z\phi(\mathbf{x})]^p$$

with the renormalization factor $Z^{-1} = \delta(0)$. To verify this, observe first that

$$\phi_i(\mathbf{x})\phi_j(\mathbf{y}) = \delta(\mathbf{x}-\mathbf{y})\phi_{ij}(\mathbf{x}) + :\phi_i(\mathbf{x})\phi_j(\mathbf{y}):$$

where : : symbolizes the normal ordered product. This leads to the definition of the renormalized product

$$[\phi_{ri}(\mathbf{x})\phi_{rj}(\mathbf{x})] \equiv \phi_{ij}(\mathbf{x}) = Z\phi_i(\mathbf{x})\phi_j(\mathbf{x}).$$

Evidently this can be repeated to give

$$[\phi_{r_{\alpha_1}}(\mathbf{x})\phi_{r_{\alpha_2}}(\mathbf{x}) \cdots \phi_{r_{\alpha_l}}(\mathbf{x})] = Z^{l-1}[\phi_{\alpha_1}(\mathbf{x})\phi_{\alpha_2}(\mathbf{x}) \cdots \phi_{\alpha_l}(\mathbf{x})] \equiv \phi_{\alpha_1\alpha_2\cdots\alpha_l}(\mathbf{x}). \tag{2.12}$$

The system Hamiltonian H of an ultralocal system generates the time evolution of the fields through the operator relation

$$\phi_{\alpha_1 \cdots \alpha_l}(\mathbf{x}, t) = \exp(iHt)\phi_{\alpha_1 \cdots \alpha_l}(\mathbf{x}) \exp(-iHt). \tag{2.13}$$

The Hamiltonian can be written as

$$H(B) = \int d\mathbf{x} \int d\vec{\lambda} B^*(\mathbf{x}, \vec{\lambda}) hB(\mathbf{x}, \vec{\lambda}), \tag{2.14}$$

where the c-number operator h is given by

$$h = -D(\vec{\nabla}_\lambda) + v(\vec{\lambda}), \tag{2.15}$$

with

$$v(\vec{\lambda}) = [D(\vec{\nabla}_\lambda)C(\vec{\lambda})]/C(\vec{\lambda}), \tag{2.16}$$

and unless otherwise stated the differential operator $D(\vec{\nabla}_\lambda)$ is

$$D(\vec{\nabla}_\lambda) \equiv \tfrac{1}{2}(\vec{\nabla}_\lambda)^2 = \tfrac{1}{2}\sum_{n=1}^{r} \frac{\partial^2}{\partial \lambda_n^2}. \tag{2.17}$$

Operators $\vec{Q}$ may be defined as[16]

$$\vec{Q} = \begin{pmatrix} Q_1 \\ Q_2 \\ \cdot \\ \cdot \\ \cdot \\ Q_r \end{pmatrix},$$

where

$$Q_i(\mathbf{x}, B) = \int d\vec{\lambda} B^*(\mathbf{x}, \vec{\lambda}) q_i(\mathbf{x}, \vec{\lambda}) B(\mathbf{x}, \vec{\lambda}). \tag{2.18}$$

Since field operators are not usually well defined as point operators, we also need smeared operators

$$Q(\vec{f}, B) = \int d\mathbf{x} \vec{f}(\mathbf{x}) \cdot \vec{Q}(\mathbf{x}, B), \tag{2.19}$$

where $f(\mathbf{x})$ is an appropriately well-defined real, smearing function. Note $\phi_i(\mathbf{x})$ defined by Eq. (2.9) is a special case of Eq. (2.18) with $q_i(\mathbf{x}, \vec{\lambda}) \equiv \lambda_i$. It follows that

$$\phi(\vec{f}) = \int d\mathbf{x} \vec{f}(\mathbf{x}) \cdot \vec{\phi}(\mathbf{x}). \tag{2.20}$$

Then in terms of gauge parameters $\beta_1 \cdots \beta_n$, the operator

$$U(\vec{\beta}) = \exp[iQ(\vec{\beta}, B)]$$

is unitary if Q is self-adjoint, by Stone's theorem. If Q is a symmetry operator S, then[16]

$$[H, S] = 0 = \textstyle\sum_i \beta_i \int d\mathbf{x} \int d\vec{\lambda} B^*(\mathbf{x}, \vec{\lambda})[h, s_i]B(\mathbf{x}, \vec{\lambda})$$

implies $[h, s_i] = 0$ so that H has the same symmetries as h. Thus, the study of symmetries of Klauder's ultralocal models can be reduced to the study of the "effective potential" $v(\vec{\lambda})$ of Eq. (2.16). Several examples will be presented in a following section.

3. BILINEAR OPERATORS IN TRANSLATED FOCK REPRESENTATIONS

The states generated from the vacuum state $|\Omega\rangle$, by the operator Σ_ζ,

$$|\zeta\rangle = \exp(\Sigma_\zeta)|\Omega\rangle$$
$$= N \exp[\sigma^*(\zeta)]|\Omega\rangle, \tag{3.1}$$

where

$$N = \exp[-\tfrac{1}{2} \int d\mathbf{x} \int d\vec{\lambda} J_0(\zeta, \zeta)] \tag{3.2}$$

are coherent states[16] defined for each element $\zeta(\mathbf{x}, \vec{\lambda}) \in L^2(R^s, R^r)$. These states form a total set and are eigenstates of the operator $A(\mathbf{x}, \vec{\lambda})$,

$$A(\mathbf{x}, \vec{\lambda})|\zeta\rangle = \zeta(\mathbf{x}, \vec{\lambda})|\zeta\rangle. \tag{3.3}$$

Any operator $\exp[iQ(\vec{f}, A)]$ that satisfies

$$\exp[-iQ(\vec{f}, A)]A(\mathbf{x}, \vec{\lambda}) \exp[iQ(\vec{f}, A)$$
$$= \exp(i\vec{f} \cdot \vec{q}) A(\mathbf{x}, \vec{\lambda}) \tag{3.4}$$

when acting on a coherent state gives a new coherent state

$$\exp[iQ(\vec{f}, A)]|\zeta\rangle = |\zeta'\rangle, \tag{3.5}$$

$$\zeta' = \exp(i\vec{f} \cdot \vec{q})\zeta. \tag{3.6}$$

The matrix element of two coherent states,

$$\langle\zeta|\zeta'\rangle = NN' \exp[\int d\mathbf{x} \int d\vec{\lambda} J_0(\zeta, \zeta')], \tag{3.7}$$

never vanishes.

These properties can be used to show that the states $|\vec{f}\rangle$ generated by the smeared field operator $\phi(\vec{f})$ defined in Eq. (2.20),

$$|\vec{f}\rangle = \exp[i\phi(\vec{f})]|\Omega\rangle, \tag{3.8}$$

is a coherent state with

$$\zeta(\vec{\lambda}) = [\exp(i\vec{\lambda} \cdot \vec{f}) - 1]C(\vec{\lambda}), \tag{3.9}$$

if

$$C(\vec{\lambda}) = C(-\vec{\lambda}), \tag{3.10}$$

and providing $C(\vec{\lambda})$ satisfies Eq. (2.8).

Another useful result is

$$\langle\zeta|\exp[iQ(\vec{f}, B)]|\zeta'\rangle = \langle\zeta|\zeta'\rangle e^F, \tag{3.11}$$

where

$$F = \int d\mathbf{x} \int d\vec{\lambda}(\zeta + C)^*[\exp(i\vec{f} \cdot \vec{q}) - 1](\zeta' + C). \tag{3.12}$$

These results all reduce to Klauder's one-dimensional case if the dot product $\vec{f} \cdot \vec{q}$ is replaced by its one-dimensional analog. For this reason, the generalization of Klauder's Theorem 3.1[5] is trivial. The statement of the theorem is as follows. Consider a family of approximating operators of the form

$$Q_\epsilon(\vec{f}, B) = \int d\mathbf{x} \int d\vec{\lambda} B_\epsilon^*(\mathbf{x}, \vec{\lambda}) \vec{f}(\mathbf{x}) \cdot \vec{q}(\vec{\lambda}) B_\epsilon(\mathbf{x}, \vec{\lambda}), \tag{3.13}$$

where $\epsilon > 0$,

$$B_\epsilon(\mathbf{x}, \vec{\lambda}) = A(\mathbf{x}, \vec{\lambda}) + C_\epsilon(\vec{\lambda}),$$

and the operator $\vec{q}$ is a function of $\vec{\lambda}$ alone. Here $C_\epsilon(\vec{\lambda})$ (possibly complex) is a smoothed model function which is chosen such that $\lim_{\epsilon \downarrow 0} C_\epsilon(\vec{\lambda}) = C(\vec{\lambda})$. Specifically we require

$$\int d\vec{\lambda} |C_\epsilon(\vec{\lambda})|^2 < \infty, \quad \int d\vec{\lambda} |\vec{q} C_\epsilon(\vec{\lambda})|^2 < \infty, \tag{3.14}$$

for all $\epsilon > 0$. For example, if $\vec{q}(\vec{\lambda})$ denotes multiplication or differentiation with respect to $\vec{\lambda}$ and we are concerned about the singularity of $C(\vec{\lambda})$ at $\vec{\lambda} = 0$, we might use

$$C_\epsilon(\vec{\lambda}) = \exp[-(\epsilon^2/|\vec{\lambda}|^2)]C(\vec{\lambda}), \tag{3.15}$$

which has the property $C_\epsilon(0) = 0$ and $C_\epsilon(\vec{\lambda}) \to C(\vec{\lambda})$, $\vec{\lambda} \neq 0$. Theorem 3.1 then states that under these conditions $Q_\epsilon(\vec{f}, B)$ is self-adjoint, $\exp[iQ_\epsilon(\vec{f}, B)]$ is unitary, and

$$\text{s-}\lim_{\epsilon \downarrow 0} \exp[iQ_\epsilon(\vec{f}, B)] = \exp[iQ(\vec{f}, B)], \tag{3.16}$$

provided that

$$J(\vec{f}) = \lim_{\epsilon \downarrow 0} \int d\vec{\lambda} C_\epsilon(\vec{\lambda})(\exp(i\vec{f} \cdot \vec{q}) - 1)C_\epsilon(\vec{\lambda}) \tag{3.17}$$

is a continuous function of $\vec{f}$ and that

$$\text{s-}\lim_{\epsilon \downarrow 0} [\exp(i\vec{f} \cdot \vec{q}) - 1]C_\epsilon(\vec{\lambda}) = O(\vec{f}, \vec{\lambda}) \tag{3.18}$$

is a continuous function of $\vec{f}$ and $O(\vec{f}, \vec{\lambda}) \in L^2(R^s, R^r)$.

For $\vec{q} = \vec{\lambda}$, $Q(\vec{f}, B) = \phi(\vec{f})$, Eqs. (2.8) and (3.10) imply Eqs. (3.17) and (3.18). Thus $\phi(\vec{f})$ is a well-defined, self-adjoint, local field operator. On the other hand, the operator $\dot{\vec{\phi}}(x)$ obtained from the time differentiation of Eq. (2.13), setting $t = 0$,

$$\dot{\vec{\phi}}(\mathbf{x}) = -i[\vec{\phi}(\mathbf{x}), H]$$
$$= \int d\vec{\lambda} B^*(\mathbf{x}, \vec{\lambda})(-i\vec{\nabla}_\lambda)B(\mathbf{x}, \vec{\lambda}), \tag{3.19}$$

where

$$\exp(-iHt)A(\mathbf{x},\vec{\lambda})\exp(iHt)=\exp(iht)A(\mathbf{x},\vec{\lambda})$$

is a form rather than a local operator because for $\vec{q}(\vec{\lambda})=(-i\vec{\nabla}_{\lambda})$ Eqs. (3.17) and (3.18) are not satisfied. Additionally

$$[\vec{\phi}(\mathbf{x}),\dot{\vec{\phi}}(\mathbf{y})]=i\delta(\mathbf{x}-\mathbf{y})\int d\vec{\lambda}\,B^{*}(\mathbf{x},\vec{\lambda})B(\mathbf{x},\vec{\lambda})$$

is also a form ($\vec{q}=1$) which has a leading c-number singularity $i\delta(\mathbf{x}-\mathbf{y})\int d\vec{\lambda}\,C^{2}(\vec{\lambda})$ corresponding to an infinite field strength renormalization. Thus the ultralocal models do not possess a canonical field and momentum pair obeying the traditional commutation brackets.[5]

Although the operators $Q(\vec{f},B)$ may be well defined, they are in general neither globally or locally unitarily equivalent[5,20] to the operators $Q(\vec{f},A)$ because

$$\exp(-\Sigma_{C})Q(\vec{f},A)\exp(\Sigma_{C})=Q(\vec{f},B) \tag{3.20}$$

is not a unitary transformation as long as $C(\vec{\lambda})$ is not $L^{2}(R^{r})$. The only exception occurs if

$$\vec{q}C(\vec{\lambda})=0,\quad \vec{q}^{*}C(\vec{\lambda})=0, \tag{3.21}$$

in which case

$$Q(\vec{f},A)=Q(\vec{f},B).$$

The Hamiltonian operator defined by Eq. (2.14) is a typical example with Eq. (2.16) being the appropriate specialization of Eq. (3.21).

Note that although the Fock space symmetry operators $Q(\vec{f},A)=S(\vec{f},A)$ satisfy $[S(\vec{f},A),H(A)]=0$, defined on a dense domain, only those corresponding symmetry operators $S(\vec{f},B)$, which satisfy the restrictions implied by Eqs. (3.16)–(3.18), can also satisfy $[S(\vec{f},B),H(B)]=0$ defined on a dense domain.

In this sense certain symmetries can be said to be "spontaneously broken" by the transformation, Eq. (3.20), which is an analog of the Higgs transformation.

In particular for the "number" operator

$$N(B)=\int d\mathbf{x}\int d\vec{\lambda}\,B^{*}(\mathbf{x},\vec{\lambda})B(\mathbf{x},\vec{\lambda}),$$

$$\exp(-\Sigma_{C})N(A)\exp(\Sigma_{C})=N(B), \tag{3.22}$$

$$N(\alpha,B)=\alpha N(B),\quad \alpha\equiv\text{real const},$$

Eqs. (3.17) and (3.18) become

$$J(\alpha)=(\exp(i\alpha)-1)\lim_{\epsilon\downarrow 0}\int d\vec{\lambda}\,C_{\epsilon}^{*}(\vec{\lambda})\,C_{\epsilon}(\vec{\lambda})\to\infty$$

and

$$O(\alpha,\vec{\lambda})=(\exp(i\alpha)-1)C(\vec{\lambda}),$$

which is not $L^{2}(R^{r})$. Thus gauge symmetry of the first kind is spontaneously broken in the ultralocal models.

For a similar reason $S(\vec{f},B)$ cannot be a symmetry operator if $C(\vec{\lambda})$ is an eigenfunction of $\vec{q}(\vec{\lambda})$ with nonzero eigenvalue even though $[\vec{q}(\vec{\lambda}),h(\vec{\lambda})]=0$ since the conditions expressed by Eqs. (3.17) and (3.18) cannot be satisfied.

4. PASSAGE TO THE FREE LIMIT

A free ultralocal field can be defined from creation and annihilation parts as

$$\vec{\phi}_{F}(\mathbf{x})=(2\mu)^{-1/2}[\vec{A}_{F}(\mathbf{x})+\vec{A}_{F}^{\dagger}(\mathbf{x})], \tag{4.1}$$

$$\vec{A}_{F}(\mathbf{x})=\int d\vec{\lambda}\,\vec{U}_{F}^{*}(\vec{\lambda})A(\mathbf{x},\vec{\lambda}), \tag{4.2}$$

where μ corresponds to a mass parameter and $\vec{U}_{F}(\vec{\lambda})$ represents a free, normalized, single particle, c-number state. A smeared free particle field can be constructed as before,

$$\phi_{F}(\vec{f})=\int d\mathbf{x}\vec{f}(\mathbf{x})\cdot\vec{\phi}_{F}(\mathbf{x})=-i\Sigma_{\xi}, \tag{4.3}$$

where $\xi=i(2\mu)^{-1/2}\vec{f}(\mathbf{x})$.

Free particle symmetry operators and a Hamiltonian would be defined as

$$\begin{aligned}\mathbf{S}_{F}&=\int d\mathbf{x}\vec{A}_{F}^{\dagger}(\vec{\mathbf{x}})\cdot\overrightarrow{(\mathbf{s}A_{F}(\mathbf{x}))},\\ H_{F}&=\mu\int d\mathbf{x}\vec{A}_{F}^{\dagger}(\mathbf{x})\cdot\vec{A}_{F}(\mathbf{x}).\end{aligned} \tag{4.4}$$

The state

$$|\vec{f}\rangle_{F}=\exp[i\phi_{F}(\vec{f})]\,|\Omega\rangle \tag{4.5}$$

is a coherent state which fulfills

$$\vec{A}_{F}(\mathbf{x})\,|\vec{f}\rangle=i(2\mu)^{-1/2}\vec{f}(\mathbf{x})\,|\vec{f}\rangle_{F}$$

as can be seen from Eq. (4.3).

It then follows, from the analogs to Eqs. (3.4)–(3.6) or Eqs. (3.11) and (3.12), that

$${}_{F}\langle\vec{f}\,|\exp(i\boldsymbol{\alpha}\cdot\mathbf{S}_{F})|\vec{f}'\rangle_{F}={}_{F}\langle f|f'\rangle_{F}\exp[(2\mu)^{-1}\times\int d\mathbf{x}\vec{f}\cdot\overrightarrow{(\exp(i\boldsymbol{\alpha}\cdot\mathbf{s})-1)\vec{f}'}], \tag{4.6}$$

$${}_{F}\langle\vec{f}\,|\vec{f}'\rangle_{F}=\exp[(-4\mu)^{-1}\int d\mathbf{x}|\vec{f}-\vec{f}'\,|^{2}], \tag{4.7}$$

from which we obtain

$${}_{F}\langle\vec{f}\,|\mathbf{S}_{F}\,|\vec{f}'\rangle_{F}=\frac{1}{2\mu}\int d\mathbf{x}\vec{f}\cdot\overrightarrow{(\mathbf{s}\vec{f}')}{}_{F}\langle\vec{f}\,|\vec{f}'\rangle_{F}. \tag{4.8}$$

For the special case $S_{F}=H_{F}$

$${}_{F}\langle\vec{f}\,|H_{F}\,|\vec{f}'\rangle_{F}=\tfrac{1}{2}\int d\mathbf{x}\vec{f}\cdot\vec{f}'\,{}_{F}\langle\vec{f}\,|\vec{f}'\rangle_{F}. \tag{4.9}$$

These functionals determine $\phi_{F}(\mathbf{x})$, $\mathbf{S}_{F}$, and H_{F} up to a unitary equivalence.

It follows from Eqs. (3.7)–(3.9) that, corresponding to Eq. (4.7), the inner product of two ultralocal states is

$$\langle\vec{f}\,|\vec{f}'\rangle=NN'\exp[\int d\mathbf{x}\int d\vec{\lambda}(\exp(-i\vec{\lambda}\cdot\vec{f})-1),\times(\exp(i\vec{\lambda}\cdot\vec{f}')-1)C^{2}], \tag{4.10}$$

where

$$N=\exp[-\tfrac{1}{2}\int d\mathbf{x}\int d\vec{\lambda}(\exp(-i\vec{\lambda}\cdot\vec{f})-1)\times(\exp(i\vec{\lambda}\cdot\vec{f})-1)C^{2}]. \tag{4.11}$$

Clearly to pass to the free particle limit, Eq. (4.7), requires

$$\int d\mathbf{x} \int d\vec{\lambda}(\exp(-i\vec{\lambda}\cdot\vec{f}) - 1)(\exp(i\vec{\lambda}\cdot\vec{f}') - 1)C^2 \to (1/2\mu)\int \vec{f}\cdot\vec{f}'\, d\mathbf{x}. \tag{4.12}$$

This can be accomplished formally through a sequence of model functions $C_m(\vec{\lambda})$, such that

$$\lim_{m\to\infty} \lambda^{r+1} C_m(\vec{\lambda}) = (2\mu\Omega_\lambda)^{-1} r\delta(\lambda), \tag{4.13}$$

where

$$\Omega_\lambda = \int d\Omega_\lambda.$$

Then

$$\lim_{m\to\infty} \langle \vec{f} | \vec{f}' \rangle = \exp[-(1/4\mu)\int d\mathbf{x} |\vec{f} - \vec{f}'|^2] = \langle {}_F\vec{f} | \vec{f}' \rangle_F. \tag{4.14}$$

It is in this sense that

$$\phi(\vec{f}) \to \phi_F(\vec{f}).$$

The free particle limit for the symmetry operators is considerably more complicated algebraically. Consider a symmetry operator

$$S_\epsilon(\alpha, B) = \int d\mathbf{x} \int d\vec{\lambda} B_\epsilon^*(\mathbf{x}, \vec{\lambda})\, \boldsymbol{\alpha}\cdot\mathbf{q}(\vec{\lambda})\, B_\epsilon(\mathbf{x}, \vec{\lambda}), \tag{4.15}$$

where $\boldsymbol{\alpha}$ are in general $r+s$ gauge parameters with the r indices suppressed. Assume that $\vec{\lambda}$ transforms as

$$\exp(-i\boldsymbol{\alpha}\cdot\mathbf{q})\, \vec{\lambda} \exp(i\boldsymbol{\alpha}\cdot\mathbf{q}) = (\overrightarrow{R_\alpha \lambda}), \tag{4.16}$$

under the corresponding c-number transformation. Then the inner product $\vec{\lambda}\cdot\vec{f}$, for example, transforms as

$$\exp(i\boldsymbol{\alpha}\cdot\mathbf{q})\vec{\lambda}\cdot\vec{f}\exp(-i\boldsymbol{\alpha}\cdot\mathbf{q}) = (\overrightarrow{R_\alpha^{-1}\lambda})\cdot\vec{f} = (\vec{\lambda}\cdot\overrightarrow{R_\alpha f}). \tag{4.17}$$

The matrix element $\langle f | \exp[iS_\epsilon(\alpha, B)] | f' \rangle$, can be written as

$$\langle f | \exp[iS_\epsilon(\alpha, B)] | f' \rangle = \langle f | f' \rangle \exp(F_\epsilon),$$

$$F_\epsilon = \int d\mathbf{x} \int d\vec{\lambda}\, C_\epsilon(\vec{\lambda}) \exp(-i\vec{\lambda}\cdot\vec{f}) \times (\exp(i\boldsymbol{\alpha}\cdot\mathbf{q}) - 1)\exp(i\vec{\lambda}\cdot\vec{f}')\, C_\epsilon(\vec{\lambda})$$

$$= \int d\mathbf{x} \int d\vec{\lambda} \exp(-i\vec{\lambda}\cdot\vec{f})[\exp(i\vec{\lambda}\cdot\overrightarrow{R_\alpha f'}) - \exp(i\vec{\lambda}\cdot\vec{f}')]\, C_\epsilon^2(\vec{\lambda}) + \int d\mathbf{x}\int d\vec{\lambda} \exp[-i\vec{\lambda}\cdot(\vec{f} - \overrightarrow{R_\alpha f'})] \times C_\epsilon(\vec{\lambda})(\exp(i\boldsymbol{\alpha}\cdot\mathbf{q}) - 1)\, C_\epsilon(\vec{\lambda}).$$

If we impose the restrictions implied by Eqs. (3.17) and (3.18),

$$F = \lim_{\epsilon\to 0} F_\epsilon = \int d\mathbf{x}\int d\vec{\lambda} \exp(-i\vec{\lambda}\cdot\vec{f})[\exp(i\vec{\lambda}\cdot\overrightarrow{R_\alpha f'}) - \exp(i\vec{\lambda}\cdot\vec{f}')]C^2(\vec{\lambda}) + \int d\mathbf{x}\int d\vec{\lambda} \times \exp[-i\vec{\lambda}\cdot(\vec{f} - \overrightarrow{R_\alpha f'})]C(\vec{\lambda})O(\alpha, \vec{\lambda}). \tag{4.18}$$

Once again replacing $C(\vec{\lambda})$ by a sequence of model function $C_m(\vec{\lambda})$, as in Eq. (4.13), gives

$$\lim_{m\to\infty} F_m = \frac{1}{2\mu} \int d\mathbf{x}\vec{f}\cdot[\overrightarrow{(R_\alpha - 1)f'}] + J(\alpha, \vec{f}, \vec{f}'),$$

where

$$J(\alpha, \vec{f}, \vec{f}') = \lim_{m\to\infty} \int d\mathbf{x}\int d\vec{\lambda}\, C_m(\vec{\lambda}) \exp[-i\vec{\lambda}\cdot(\overrightarrow{\vec{f} - R_\alpha f'})]O_m(\alpha, \vec{\lambda}) \tag{4.19}$$

is a continuous function of α, $\vec{f}$ and $\vec{f}'$. Thus, if R_α is written in the form $R_\alpha = \exp(i\boldsymbol{\alpha}\cdot\mathbf{s})$, to correspond to Eq. (4.6),

$$\lim_{m\to\infty} \langle \vec{f} | \exp(iS(\alpha, B)) | \vec{f}' \rangle = {}_F\langle f | \exp(i\boldsymbol{\alpha}\cdot\mathbf{S}_F) | f' \rangle_F \exp[J(\alpha, \vec{f}, \vec{f}')], \tag{4.20}$$

which is for small α

$$\lim_{m\to\infty} \langle \vec{f} | S(\alpha, B) | f' \rangle = \boldsymbol{\alpha}\cdot[{}_F\langle \vec{f} | \mathbf{S}_F | \vec{f}' \rangle_F + \lim_{m\to\infty} \int d\mathbf{x}\int d\vec{\lambda} \times \exp[-i\vec{\lambda}\cdot(\vec{f} - \vec{f}')]C_m(\vec{\lambda})\mathbf{q}(\vec{\lambda})C_m(\vec{\lambda})]. \tag{4.21}$$

The naive free particle limit, Eq. (4.6), is therefore obtained only if the last term in Eq. (4.21) is equal to zero, which is the case if $q(\vec{\lambda})C(\vec{\lambda}) = 0$ [see Eq. (3.20)]. In the parlance of solid state, many-body physics, the last term in Eq. (4.21), represents the contribution of the condensate to the matrix element, whereas the first term represents the contribution of the excitations. In the free particle limit the two contributions uncouple as should be expected.

As Klauder points out,[5] the free particle limit of the Hamiltonian is a more delicate operation because $[\phi(\mathbf{x}, t), H]$ is ill defined. The procedure to be followed parallels that of Klauder's[5] in the $r = 1$ case. For $D(\lambda) = (\vec{\nabla}_\lambda)^2$, an operator $g(\vec{\lambda})$ can be defined where

$$h(\vec{\lambda}) = g^*(\vec{\lambda})g(\vec{\lambda}), \quad g(\vec{\lambda}) = \sqrt{\tfrac{1}{2}}\, C(\vec{\lambda})\vec{\nabla}_\lambda C^{-1}(\vec{\lambda}). \tag{4.22}$$

The Hamiltonian can then be written as

$$H = \int d\mathbf{x}\int d\vec{\lambda}[g(\vec{\lambda})A(\mathbf{x}, \vec{\lambda})]^*[g(\vec{\lambda})A(\mathbf{x}, \vec{\lambda})]. \tag{4.23}$$

A regularized Hamiltonian can be defined by

$$H = \lim_{\epsilon\to 0} H_\epsilon,$$

$$H_\epsilon = \int d\mathbf{x}\int d\vec{\lambda}[g(\vec{\lambda})A(\mathbf{x}, \vec{\lambda})]^* M_\epsilon[g(\vec{\lambda})A(\mathbf{x}, \vec{\lambda})], \tag{4.24}$$

where

$$M_\epsilon = \exp[-(\epsilon^2/\lambda^2)]/\int d\vec{\lambda} C^2(\vec{\lambda}) \exp[-(\epsilon^2/\lambda^2)]. \tag{4.25}$$

The numerator of M_ϵ takes care of the singularity at $\vec{\lambda} = 0$ and the denominator takes care of the field strength renormalization. From Eqs. (3.3) and (3.9), it can be shown that

$$\langle \vec{f} | H_\epsilon | \vec{f}' \rangle = \langle \vec{f} | \vec{f}' \rangle \int d\mathbf{x} \int d\vec{\lambda} [g(\vec{\lambda}) C(\vec{\lambda})\, \vec{\lambda} \cdot \vec{\psi}]^* \times M_\epsilon [g(\vec{\lambda}) C(\vec{\lambda})\, \vec{\lambda} \cdot \vec{\psi}'], \tag{4.26}$$

where

$$\vec{\lambda} \cdot \vec{\psi} = [\exp(i\vec{\lambda} \cdot \vec{f}) - 1], \quad \vec{\lambda} \cdot \vec{\psi}' = [\exp(i\vec{\lambda} \cdot \vec{f}') - 1].$$

The limiting form for H_ϵ then follows directly from application of Eqs. (4.13), (4.22), and (4.25),

$$\lim_{m \to \infty} \lim_{\epsilon \to 0} \langle \vec{f} | H_\epsilon | \vec{f}' \rangle = \tfrac{1}{2} \int d\mathbf{x} \vec{f} \cdot \vec{f}'_F \langle \vec{f} | \vec{f}' \rangle_F = {}_F\langle \vec{f} | H_F | \vec{f}' \rangle_F. \tag{4.27}$$

5. EXAMPLE, MANY-BODY ULTRALOCAL MODEL

If the $\vec{\lambda}_i$ internal variables are considered as relative coordinates, Klauder's model can be generalized to describe a system of N equal mass particles with unit mass. Then

$$\lambda_i = \mathbf{x}_i - \mathbf{x}, \tag{5.1}$$

where the space coordinate $\mathbf{x}$ can be considered as the cm coordinate,

$$\mathbf{x} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_i, \tag{5.2}$$

and

$$\vec{\lambda} = \begin{pmatrix} \lambda_1 \\ \cdot \\ \cdot \\ \cdot \\ \lambda_r \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \cdot \\ \cdot \\ \cdot \\ \lambda_N \end{pmatrix}. \tag{5.3}$$

The constraint

$$\sum_i \lambda_i = 0 \tag{5.4}$$

implies that $d\vec{\lambda}$ is replaced by

$$d\vec{\lambda} = \delta \left(\sum_i \lambda_i \right) \prod_i^N d\lambda_i. \tag{5.5}$$

The differential operator $D(\nabla_\lambda)$ becomes

$$-D(\vec{\nabla}_\lambda) = \tfrac{1}{2} \sum_{i=1}^{N} \mathbf{p}_i'^2 = -\tfrac{1}{2} \left[\sum_i \nabla_{\lambda_i}^2 - \frac{1}{N} \left(\sum_i \nabla_{\lambda_i} \right) \cdot \left(\sum_j \nabla_{\lambda_j} \right) \right], \tag{5.6}$$

where the relative momentum operator is

$$\mathbf{p}_i' = -i\nabla_{\lambda_i} - \frac{1}{N} \sum_{j=1}^{N} (-i\nabla_{\lambda_j}), \quad \sum_i \mathbf{p}_i' = 0. \tag{5.7}$$

For model functions where

$$C(\vec{\lambda}) \equiv C(|\lambda_i - \lambda_j|) \quad (i \neq j), \tag{5.8}$$

$$\left(\sum_i \mathbf{p}_i'^2 \right) C(\vec{\lambda}) = -\left(\sum_i \nabla_{\lambda_i}^2 \right) C(\vec{\lambda}), \tag{5.9}$$

so that

$$v(|\lambda_i - \lambda_j|) = \{ \tfrac{1}{2} (\textstyle\sum \nabla_{\lambda_i}^2) C(\vec{\lambda}) \} / C(\vec{\lambda}). \tag{5.10}$$

Clearly $h(\vec{\lambda})$ is both rotationally and translationally invariant.

In particular the two-body case (for $s = 3$) reduces in the usual way to the one-dimensional ultralocal model described by Klauder if one uses

$$C(\vec{\lambda}) = C_K(r)/r, \tag{5.11}$$

where C_K is Klauder's singular, one-dimensional C-function of Eq. (2.7)

$$C_K(r) = \exp[-y(r)]/r^\gamma, \quad \tfrac{3}{2} > \gamma \geq \tfrac{1}{2}, \tag{5.12}$$

with $\mathbf{r} = \lambda_1 - \lambda_2$ and $r = |\mathbf{r}|$. Then

$$v(r) = \frac{1}{2} \left(\frac{1}{r^2} \frac{\partial}{\partial r}\, r^2 \frac{\partial C}{\partial r} \right) \Big/ C(r) = \frac{1}{2} \frac{C_K''}{C_K}. \tag{5.13}$$

The definition of a "scalar" ultralocal field operator for this model is

$$\phi(\mathbf{x}) = \tfrac{1}{2} \int d\Omega_r \int_{-\infty}^{\infty} B^*(\mathbf{x}, \mathbf{r}) r B(\mathbf{x}, \mathbf{r}) r^2\, dr, \tag{5.14}$$

where $d\Omega_r$ is the differential solid angle about $\mathbf{r}$. The odd extension in r has been used to eliminate the singularity in $C_K(r)$ at the origin.

We could have equally well defined a three-dimensional vector field

$$\boldsymbol{\phi}(\mathbf{x}) = \int d\Omega_r \int_0^{\infty} B^*(\mathbf{x}, \mathbf{r}) \mathbf{r} B(\mathbf{x}, \mathbf{r})\, r^2\, dr. \tag{5.15}$$

In this case, the singular term is eliminated because

$$\int d\Omega_r \mathbf{r} = 0. \tag{5.16}$$

An entire class of two-body ultralocal models can be studied by choosing as the model function

$$C_K(r) = r^{1/2} H^{(1)}_{(1/n)(\gamma+1/2)}(\beta r^n), \tag{5.17}$$

with n a positive integer greater than or equal to two and $H_\nu^{(1)}$ a νth-order Hankel function. This leads, due to Eq. (5.13), to the potential

$$2v(r) = [\gamma(\gamma+1)/r^2] + (\beta n r^{n-1})^2. \tag{5.18}$$

This determines the Hamiltonian, which in turn can be used in the form $[\dot{\phi}(\mathbf{x}), H] = i\ddot{\phi}(\mathbf{x})$ to determine the form equations of motion for the ultralocal field operator. These are, using the result of Eq. (3.19),

$$-\dddot{\phi}(\mathbf{x}) = \int d\mathbf{r}\, B^*(\mathbf{x}, \mathbf{r}) [\nabla v(r)] B(\mathbf{x}, \mathbf{r}),$$

which becomes by virtue of Eq. (2.12) ($\beta = 1$)

$$\ddot{\boldsymbol{\phi}}(\mathbf{x}) = [(\gamma(\gamma+1)/Z^4 |\boldsymbol{\phi}|^4) - (n-1) n^2 (Z^2 \boldsymbol{\phi} \cdot \boldsymbol{\phi})^{n-2}] \boldsymbol{\phi}(\mathbf{x}). \tag{5.19}$$

It is interesting to compare this result to the special case obtained from

$$[C_K(r)]_{HA} = r^{1/2} H^{(1)}_{(1/2)(\gamma+1/2)}(\alpha r^2), \tag{5.20}$$

which implies

$$2v_{HA}(r) = [\gamma(\gamma+1)/r^2] + 4\alpha^2 r^2.$$

For the choice

$$\alpha = (n/2)[(n-1)\langle (Z^2 \phi_{HA} \cdot \phi_{HA})^{n-2} \rangle]^{1/2}$$

the form equation of motion for the field $\phi_{HA}(\mathbf{x})$ is

$$\ddot{\phi}_{HA} = [\gamma(\gamma+1)/Z^4 |\, \phi_{HA}|^4 - (n-1)n^2 \times \langle (Z^2 \phi_{HA} \cdot \phi_{HA})^{n-2} \rangle] \phi_{HA}(\mathbf{x}) \tag{5.21}$$

which is a Hartree-like approximation[21] to Eq. (5.19). In this sense α represents a renormalization of the oscillator frequency parallel to the case discussed by Chang.[21,22]

The model functions C_K, Eqs. (5.17) and (5.20), are similar only near the origin in r. For larger r the solutions are different since both the order of the Hankel function and the power the argument are different. The fields ϕ and ϕ_{HA} are uniquely determined by these model functions. Thus that part of the field determined by the values of $r \gg 0$ cannot be accurately obtained from Hartree type methods.

The operator equations of motion, which are more complicated than those of the form equations of motion, also follow as a trivial extension of Klauder's work.[5] For example, for the model function given in Eq. (5.17), Klauder's result[5]

$$\frac{d^2}{dt^2} \phi_r^2(\mathbf{x}) = 4H(\mathbf{x}) - \sum_n w_{2n} \{ \phi_r^{2n}(\mathbf{x}) - \langle \Omega |\, \phi_r^{2n}(\mathbf{x}) \,| \Omega \rangle \} \tag{5.22}$$

generalizes to

$$\frac{d^2}{dt^2} (\phi_r \cdot \phi_r) = 4H(\mathbf{x}) - 2\beta^2 n^3 \{ (\phi_r \cdot \phi_r)^{n-1} - \langle \Omega |\, (\phi_r \cdot \phi_r)^{n-1} \,| \Omega \rangle \}, \tag{5.23}$$

where $H(\mathbf{x})$ is the Hamiltonian density.

A simple extension of the "two-body" ultralocal models to N interacting pairs can be obtained from the product function

$$C(\vec{\lambda}) = \prod_{\substack{i,j=1 \\ i \neq j}} \frac{\exp[-\gamma(|\mathbf{r}_{ij}|)]}{|\mathbf{r}_{ij}|^\gamma}, \tag{5.24}$$

where

$$\mathbf{r}_{ij} = \lambda_i - \lambda_j. \tag{5.25}$$

In this case the local field operator is a tensor

$$\vec{\Phi}(\mathbf{x}) = \int B^*(\mathbf{x}, \vec{\lambda}) \prod_{\substack{i,j \\ i \neq j}} \mathbf{r}_{ij} \; B(\mathbf{x}, \vec{\lambda})\, d\vec{\lambda}, \tag{5.26}$$

where $d\vec{\lambda}$ is given in Eq. (5.5).

In view of the remarks regarding the Hartree approximation to the two-body case, it is possible that this method may not be among the interesting class of $1/N$ approximations for the case of N interacting pairs.

6. CONCLUSIONS

We have shown how to generalize Klauder's one-component ultralocal models to arbitrary finite N-components and have discussed Hartree-like approximations. In each case the symmetry was inserted in the model function $C(\vec{\lambda})$ which determined both the dynamics (including symmetry!) and the representation of the field.

By choosing a more general symmetry for $C(\vec{\lambda})$ one could obtain more general ultralocal symmetries. One such more general class of ultralocal models could be based on the Moshinsky chain on unitary groups of order N

$$U_N \supset \begin{pmatrix} U_{N-1} & 0 \\ 0 & 1 \end{pmatrix} \supset \begin{pmatrix} U_{N-2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \supset \cdots \supset \begin{pmatrix} U_1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & & 0 \\ 0 & 0 & 1 & & 0 \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{6.1}$$

For each $r = 1, 2, \ldots, N-1$ the generators of the U_r satisfy the commutation relations of the Lie algebra of the unitary groups and their irreducible representations can be represented as partitions of integers $[n_1, n_2, \ldots, n_r]$. Because the notation for the states is both lengthy and well known, we simply refer to the elegant monograph by Moshinsky.[23] The full analysis is painstakingly carried out there.

There is another chain of physically interesting subgroups which Moshinsky[24] has also analyzed in detail. This chain includes the pairing force and the Elliott SU_3 model,[25] which has had some success in nuclear structure physics. This chain is

$$U_{2l+1} \supset R_{2l+1} \supset D^{(l)}(R_3), \tag{6.2}$$

where R_{2l+1} is the rotation group in $2l+1$ dimensions and $D^{(l)}$ is the $(2l+1) \times (2l+1)$ matrices which represent the l-dimensional irreducible representations of R_3. The physical origin of the chain in Eq. (6.2) is the matrix elements of a Yukawa interaction in a harmonic oscillator basis.

We make no claims that our gradient-less ultralocal models are physically realistic, although it would appear from this work that they could apply to systems that have a fixed center of mass. They are, however, rigorously soluble for one or many components so that we were able to study the popular $1/N$ Hartree-type methods. Since the Hartree approximations to ultralocal models were found to be accurate only if the contribution to $\phi(\mathbf{x})$ from $C(r)$, $r \gg 0$, is negligible, one might wonder about the general applicability of these methods.

ACKNOWLEDGMENTS

We wish to thank Dr. J.R. Klauder, Professor Umezawa, Professor Bing-Lin Young, Professor Justin Huang, and Professor Franklin E. Schroeck, Jr., for useful conversations. Dr. Klauder is especially thanked for some encouragement and a great deal of good advice. Since, however, we have not always fol-

lowed his good advice, he is not responsible for any of the views expressed here—those are our own. BDF thanks Iowa State University Physics Department and Ames Laboratory US ERDA for their hospitality when this work was started, the University of Missouri for a sabbatical leave at that time, and the Research Council of the University of Missouri for a fellowship to complete the work. A preliminary version of this work was presented to the 82nd Annual Meeting of the American Mathematical Society, January, 1976, in San Antonio, Texas.

*Prepared for the Energy Research and Development Administration under contract No. W7405-eng-82.
†Supported in part by the University of Missouri Research Council.
‡Permanent address.

[1]G. Wentzel, Helv. Phys. Acta **13**, 269 (1974).
[2]L.I. Schiff, Phys. Rev. **92**, 766 (1953).
[3]J.R. Klauder, J. Math. Phys. **11**, 609 (1970).
[4]J.R. Klauder, Comm. Math. Phys. **18**, 307 (1970).
[5]J.R. Klauder, Acta. Phys. Austr. Suppl. **VIII**, 227 (1971).
[6]J.R. Klauder, "Heisenberg Equations of Motion," in Proceedings of the Symposium on Basic Questions in Elementary Particle Physics, Munich, June 8-18, 1971.
[7]J.R. Klauder, Ann. Phys. (N.Y.) **79**, 111 (1973).
[8]J.R. Klauder, in *Boulder Lectures in Theoretical Physics, Vol. XIVB*, edited by W. Brittin (Univ. of Colorado Press, Boulder, 1973), pp. 329 *et seq.*
[9]J.R. Klauder, Phys. Lett. B **47**, 523 (1973).
[10]J.R. Klauder, Acta Phys. Austr. Suppl. **IX**, 341 (1973).
[11]J.R. Klauder, Acta Phys. Austr. Suppl. **XIV**, 581 (1975).
[12]J.R. Klauder, Acta Phys. Austr. **41**, 237 (1975).
[13]C.M. Newman, Commun. Math. Phys. **26**, 169 (1972).
[14]H. Araki, J. Math. Phys. **1**, 492 (1960).
[15]G.A. Goldin, in D.H. Sharp and A.S. Wightman, Eds., *Local Currents and Their Applications* (North-Holland, Amsterdam, 1974), pp. 101 *et seq.*
[16]B. DeFacio and C.L. Hammer, J. Math. Phys. **17**, 267 (1976).
[17]I.M. Gel'fand and N.Ya. Vilenkin, *Generalized Functions, Vol. 4: Applications to Harmonic Analysis* (translated by Amiel Feinstein) (Academic, New York, 1964), pp. 273 *et seq.*
[18]E. Lukacs, *Characteristic Functions* (Hafner, New York, 1970), 2nd ed.
[19]H.D.I. Abarbanal, J.R. Klauder, and J.G. Taylor, Phys. Rev. **152**, 1198 (1966).
[20]J.R. Klauder, J. McKenna, and E.J. Wood, J. Math. Phys. **7**, 822 (1966).
[21]S.J. Chang, Phys. Rev. D **12**, 1071 (1975).
[22]S.J. Chang, Phys. Rev. D **13**, 2778 (1976).
[23]M. Moshinsky, J. Math. Phys. **4**, 1128 (1963).
[24]M. Moshinsky, *Group Theory and the Many-Body Problem* (Gordon and Breach, New York, 1967).
[25]J.P. Elliott, Proc. Roy. Soc. A **245**, 128 (1958).

Axiomatic basis for spaces with noninteger dimension

Frank H. Stillinger

Bell Laboratories, Murray Hill, New Jersey 07974
(Received 9 December 1976)

Five structural axioms are proposed which generate a space S_D with "dimension" D that is not restricted to the positive integers. Four of the axioms are topological; the fifth specifies an integration measure. When D is a positive integer, S_D behaves like a conventional Euclidean vector space, but nonvector character otherwise occurs. These S_D conform to informal usage of continuously variable D in several recent physical contexts, but surprisingly the number of mutually perpendicular lines in S_D can exceed D. Integration rules for some classes of functions on S_D are derived, and a generalized Laplacian operator is introduced. Rudiments are outlined for extension of Schrödinger wave mechanics and classical statistical mechanics to noninteger D. Finally, experimental measurement of D for the real world is discussed.

I. INTRODUCTION

Continuous variation in the number of dimensions D for space emerges as a useful concept in several areas of physics. It was first introduced, apparently, to aid in understanding critical phenomena exhibited by the binary fluid of "Gaussian molecules."[1] More recently, expansions for critical exponents in terms of 4-D have been developed for a wide range of cooperative many-body systems.[2,3] In addition, quantum field theory has been studied as a function of D, which then serves as a regularizing parameter.[4-6] Finally, atomic bound states (as described by the Schrödinger equation) have been studied for continuously variable D.[7]

In each of the cited examples, extending D from the positive integers to the real line (or complex plane) has been an obvious procedure advertised by the way that D occurs in certain key quantities. Typical such quantities would be the Gaussian integral

$$\int d\mathbf{r} \exp(-\alpha r^2) = (\pi/\alpha)^{D/2}, \tag{1.1}$$

or the radial Laplace operator

$$\frac{d^2}{dr^2} + \frac{(D-1)}{r}\,\frac{d}{dr}\,, \tag{1.2}$$

wherein precisely the same form can be adopted for the extended D domain. Of course the extension is not unique, since one can always augment a given interpolation formula with extra terms which vanish at the positive integers. But regardless of which forms for extension of the key quantities are selected, one must be concerned about their logical independence as assumptions, or even about their logical compatibility.

This paper presents a mathematically concrete realization of spaces with noninteger D. In fact, the formalism shows that the specific expressions (1.1) and (1.2) as interpolations are indeed compatible. The broader aim is to provide systematic rules for computation in spaces with noninteger D. In the interests of future application to physical theory, we indicate how Schrödinger wave mechanics and Gibbsian statistical mechanics transform into the general-D regime.

The concrete realization offered here may encourage new results in the areas of physics which originally motivated it. The theory of critical phenomena seems to be a good candidate. In particular, convergence properties of critical-exponent expansions in 4-D are uncertain at present. But now that statistical mechanics takes more tangible form for noninteger D, it becomes clearer how one might formulate and attempt to prove perturbation convergence theorems for expansions in 4-D, at least for some domain of positive values for this parameter.

Even leaving aside trivial modifications [such as replacement of D by $D + 0.1 \sin(\pi D)$ in the interpolation formulas], the formalism offered here for noninteger D may not be unique. Nevertheless, it appears to combine simplicity and utility in a way not easily challenged by alternative approaches. Furthermore, the present formalism is attractive on account of the rich opportunities it displays for pure mathematics; in particular the geometry of sphere packings for noninteger D becomes a valid area for inquiry.[8]

II. TOPOLOGICAL ASSUMPTIONS

We let S_D denote the space of interest. It contains points $\mathbf{x}$, $\mathbf{y}$, $\cdots$, and has topological structure specified by the following axioms:

A1. *S_D is a metric space.*

A2. *S_D is dense in itself.*

A3. *S_D is metrically unbounded.*

The distance between points $\mathbf{x}$ and $\mathbf{y}$ implied by A1. will be written as $r(\mathbf{x}, \mathbf{y})$. It must satisfy the conventional criteria required of metrics[9]:

(a) $r(\mathbf{x}, \mathbf{y}) \geq 0$,

(b) $r(\mathbf{x}, \mathbf{y}) = r(\mathbf{y}, \mathbf{x})$,

(c) $r(\mathbf{x}, \mathbf{x}) = 0$,

(d) if $r(\mathbf{x}, \mathbf{y}) = 0$, then $\mathbf{x} = \mathbf{y}$,

(e) $r(\mathbf{x}, \mathbf{y}) + r(\mathbf{x}, \mathbf{z}) \geq r(\mathbf{y}, \mathbf{z})$ (triangle inequality).

(2.1)

The existence of a metric for S_D permits neighborhoods of given positive radius to be constructed about each point. That S_D is dense, Axiom A2., simply means that every such neighborhood about an arbitrary point $\mathbf{x} \in S_D$ contains at least one other point $\mathbf{y}$. Axioms A1. and A2. together require that S_D contain an infinite number of points.

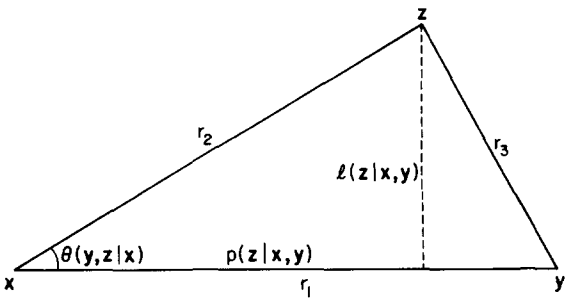


FIG. 1. Definition of geometric quantities for triangles.

Axiom A3. means that for every $\mathbf{x}\in S_D$, and any $R>0$, there exists a point $\mathbf{y}$ such that

$$r(\mathbf{x},\mathbf{y})>R, \tag{2.2}$$

i.e., the space is infinite in extent.

Real or complex-valued functions $f(\mathbf{x})$ can be defined on S_D. If we have $(i=1,2,3,\cdots)$

$$\lim_{i\to\infty} f(\mathbf{x}_i)=f(\mathbf{x}) \tag{2.3}$$

for all sequences $\{\mathbf{x}_i\}$ with the property

$$\lim_{i\to\infty} r(\mathbf{x}_i,\mathbf{x})=0, \tag{2.4}$$

then f is continuous at $\mathbf{x}$. Similar statements apply to continuity of functions of several variables.

Since any three points $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ define three distances obeying the triangle inequality, it will be convenient to adopt some familiar geometrical results for triangles (see Fig. 1). In particular, the angle $0\leq\theta(\mathbf{y},\mathbf{z}|\mathbf{x})\leq\pi$ subtended by $\mathbf{y}$ and $\mathbf{z}$ at $\mathbf{x}$ can be obtained from the "cosine law,"

$$\cos\theta(\mathbf{y},\mathbf{z}|\mathbf{x})=\frac{r_1^2+r_2^2-r_3^2}{2r_1r_2}, \tag{2.5}$$

where

$$r_1\equiv r(\mathbf{x},\mathbf{y}),\quad r_2\equiv r(\mathbf{x},\mathbf{z}),\quad r_3\equiv r(\mathbf{y},\mathbf{z}). \tag{2.6}$$

This definition leads immediately to expressions for the "projection of $\mathbf{z}$ along $(\mathbf{x},\mathbf{y})$," written $p(\mathbf{z}|\mathbf{x},\mathbf{y})$, as well as its orthogonal complement $l(\mathbf{z}|\mathbf{x},\mathbf{y})$:

$$p(\mathbf{z}|\mathbf{x},\mathbf{y})=r(\mathbf{x},\mathbf{z})\cos\theta(\mathbf{y},\mathbf{z}|\mathbf{x})=\frac{r_1^2+r_2^2-r_3^2}{2r_1}, \tag{2.7}$$

$$l(\mathbf{z}|\mathbf{x},\mathbf{y})=\frac{1}{2r_1}\left[2(r_1^2r_2^2+r_1^2r_3^2+r_2^2r_3^2)-r_1^4-r_2^4-r_3^4\right]^{1/2}, \tag{2.8}$$

$$r_2^2=p^2+l^2. \tag{2.9}$$

In ordinary Euclidean spaces, vector addition is permitted,

$$\mathbf{u}=a\mathbf{x}+b\mathbf{y}, \tag{2.10}$$

and the result is again an element of the space. We must specifically reject (2.10) for noninteger D, since any vector space must have a finite integer, or infinite, number of basis vectors,[10] and that number inevitably becomes the space dimension. Hence S_D normally will not be a vector space.

Again for Euclidean vector spaces, a triangle formed from three points as shown in Fig. 1 has sides (and altitude l) which are themselves embedded in the space. This obviously follows from the fact that any point on a line can be expressed as a linear combination of the endpoints, in the form of Eq. (2.10). But for our noninteger-D spaces, the available axioms A1., A2., and A3. are insufficient to ensure that *any* points of S_D lie precisely between two triangle vertices, let alone an entire side.

To rectify matters, it will be necessary to include a fourth axiom:

A4. *For any two points* $\mathbf{y},\mathbf{z}\in S_D$, *and any* $\epsilon>0$, *there exists an* $\mathbf{x}\in S_D$ *such that*:

(a) $r(\mathbf{x},\mathbf{y})+r(\mathbf{x},\mathbf{z})=r(\mathbf{y},\mathbf{z})$;

(b) $|r(\mathbf{x},\mathbf{y})-r(\mathbf{x},\mathbf{z})|<\epsilon r(\mathbf{y},\mathbf{z})$.

Part (a) permits equality always to be achieved in the triangle inequality; an equivalent phrasing would be that $\theta(\mathbf{y},\mathbf{z}|\mathbf{x})=\pi$. Either way, it places an $\mathbf{x}$ directly between $\mathbf{y}$ and $\mathbf{z}$. Part (b) permits $\mathbf{x}$ to be near the midpoint. The full implication of A4. is that any two points in S_D are connected by a continuous line embedded in that space.

III. INTEGRATION MEASURE

The topological structure imposed on S_D must now be supplemented with a statement of volume element size, so that a linear integration operation becomes possible. Considering the fact that, thus far, only points and distances exist for S_D, we are obliged to introduce weights,

$$W_n(\mathbf{x}_1\cdots\mathbf{x}_n|r_1\cdots r_n) \tag{3.1}$$

for a fixed set of points $\mathbf{x}_1\cdots\mathbf{x}_n$, and distances $r_1\cdots r_n$ measured from them. If thin "spherical" shells (with inner and outer radii r_1 and r_1+dr_1, r_2 and r_2+dr_2, $\cdots$) are erected respectively about $\mathbf{x}_1\cdots\mathbf{x}_n$, then $W_n\,dr_1\cdots dr_n$ gives the content of the mutual intersection of those shells. Once having the W_n in hand, it becomes possible to integrate functions $h(r_{01}\cdots r_{0n})$ of the distances $r_{0j}\equiv r(\mathbf{x}_0,\mathbf{x}_j)$ over all $\mathbf{x}_0\in S_D$ by the simple expedient of using the r_{0j} as separate conventional integration variables,

$$\begin{aligned}\int d\mathbf{x}_0\,h(r_{01}\cdots r_{0n})&=\int_0^\infty dr_{01}\cdots\int_0^\infty dr_{0n}\,W_n(\mathbf{x}_1\cdots\mathbf{x}_n|r_{01}\cdots r_{0n})\\&\quad\times h(r_{01}\cdots r_{0n}).\end{aligned} \tag{3.2}$$

Repeated application of this general procedure would permit evaluation of multiple integrals, over several $\mathbf{x}_j$'s in a finite point set, of functions of distances in that set.

In principle, explicit formulas could be provided for the W_n as functions of the $\frac{1}{2}n(n+1)$ distances r_{ij} $(0\leq i\leq j\leq n)$. In practice, it is more efficient to define those functions implicitly by demanding that multiply-rooted Gaussian integrals have preassigned values. Consequently, we now state the fifth axiom for S_D:

A5. *For any positive integer n,*

$$\int d\mathbf{x}_0 \exp\left(-\sum_{j=1}^{n} \alpha_j r_{0j}^2\right) = \left(\frac{\pi}{\tau}\right)^{D/2} \exp\left(-\frac{1}{\tau} \sum_{j<k=1}^{n} \alpha_j \alpha_k r_{jk}^2\right), \quad \tau = \sum_{j=1}^{n} \alpha_j. \tag{3.3}$$

This is the only point at which the dimension parameter D enters the present axiomatic formalism. It should be noted that (3.3) coincides with standard integral results when D is a positive integer. When $n=1$, Eq. (3.3) agrees with Eq. (1.1).

Axiom A5. confers overall uniformity on S_D. The result produced by integrating any rooted Gaussian depends only on distances between root points (which can be anywhere in S_D), and not in any way on absolute position in S_D. In this sense there are no distinguished points in S_D.

By combining Eqs. (3.2) and (3.3), along with the variable change

$$t_j = r_{0j}^2, \tag{3.4}$$

one discovers the identities

$$\int_0^\infty dt_1 \cdots \int_0^\infty dt_n \left[\frac{W_n(\cdots | t_1^{1/2} \cdots t_n^{1/2})}{2^n (t_1 \cdots t_n)^{1/2}}\right] \times \exp\left(-\sum_{j=1}^{n} \alpha_j t_j\right) = \left(\frac{\pi}{\tau}\right)^{D/2} \exp\left(-\frac{1}{\tau} \sum_{j<k=1}^{n} \alpha_j \alpha_k r_{jk}^2\right). \tag{3.5}$$

Essentially, this provides the result of an n-fold Laplace transform on W_n. The weight itself can be computed from the appropriate transform inversion formula[11]

$$W_n(\cdots | t_1^{1/2} \cdots t_n^{1/2}) = \frac{(t_1 t_2 \cdots t_n)^{1/2}}{(\pi i)^n} \int_{c_1-i\infty}^{c_1+i\infty} d\alpha_1 \cdots \int_{c_n-i\infty}^{c_n+i\infty} d\alpha_n \left(\frac{\pi}{\tau}\right)^{D/2} \times \exp\left(\sum_{j=1}^{n} t_j \alpha_j - \frac{1}{\tau} \sum_{j<k=1}^{n} r_{jk}^2 \alpha_j \alpha_k\right). \tag{3.6}$$

The simplest of the weights, W_1, allows integrals of radially symmetric functions to be computed:

$$\int d\mathbf{x}_0 f[r(\mathbf{x}_0, \mathbf{x}_1)] = \int_0^\infty dr\, W_1(r) f(r). \tag{3.7}$$

The inverse Laplace transform needed to find W_1 from Eq. (3.6) is a standard form.[12] The result is found to be

$$W_1(r) = \sigma(D) r^{D-1}, \quad \sigma(D) = \frac{2\pi^{D/2}}{\Gamma(D/2)}. \tag{3.8}$$

When D is a positive integer this agrees precisely with the known spherical volume element for mutlidimensional Euclidean spaces.[13] This is the central fact which justifies the claim that S_D is a "space of D dimensions."

The volume of the radius-R sphere in S_D follows immediately from W_1,

$$V(R, D) = \int_0^R W_1(r)\, dr = \frac{\pi^{D/2} R^D}{\Gamma(1+\frac{1}{2}D)}. \tag{3.9}$$

For any $R>0$ this has the property

$$\lim_{D\to 0} V(R, D) = 1, \tag{3.10}$$

which in turn implies that W_1 is a Dirac delta function in the same limit,

$$\lim_{D\to 0} W_1(r) = \delta(r-0). \tag{3.11}$$

Consequently, integration weight in S_D collapses to zero extension, in spite of the fact that Axiom A3. still maintains a sparse set of widely separated pairs of points. For a continuous function f,

$$\lim_{D\to 0} \int d\mathbf{x}\, f(r) = f(0). \tag{3.12}$$

Equation (3.6) may be used to derive a consistency property of the weights,

$$\int_0^\infty dr_n\, W_n(\mathbf{x}_1 \cdots \mathbf{x}_n | r_1 \cdots r_n) = W_{n-1}(\mathbf{x}_1 \cdots \mathbf{x}_{n-1} | r_1 \cdots r_{n-1}). \tag{3.13}$$

IV. DENSITY OF MUTUALLY PERPENDICULAR LINES

Inverting the Laplace transforms, as required by Eq. (3.6) to obtain W_n, becomes an increasingly arduous task as n increases. But experience shows that no insuperable difficulties arise—one needs recourse only to a small number of recurrent tabulated inverse-trans-from types.

One finds the following expression for the two-center weight (valid for all real D):

$$W_2(\mathbf{x}_1, \mathbf{x}_2 | r_{01}, r_{02}) = 2^{D-3}\sigma(D-1) r_{01} r_{02} r_{12}^{2-D} \Delta^{D-3}(r_{01}, r_{02}, r_{12}), \tag{4.1}$$

where Δ is the area of the triangle having sides r_{01}, r_{02}, and r_{12},

$$\Delta(r_{01}, r_{02}, r_{12}) = \tfrac{1}{4}[2(r_{01}^2 r_{02}^2 + r_{01}^2 r_{12}^2 + r_{02}^2 r_{12}^2) - r_{01}^4 - r_{02}^4 - r_{12}^4]^{1/2}. \tag{4.2}$$

If no triangle can be formed, Δ must be set equal to zero. By setting $D=3$, expression (4.1) reduces to a familiar weight for the nonorthogonal bipolar coordinate system,

$$W_2(D=3) = 2\pi r_{01} r_{02}/r_{12}. \tag{4.3}$$

A right triangle will be formed if $r_{01}=r_{02}=R$, $r_{12}=2^{1/2}R$, with the right angle at vertex 0. By inserting these values in W_2 we obtain a measure for the density of mutually perpendicular lines,

$$W_2(2^{1/2}R | R, R) = 2^{2-D/2}\pi^{(D-1)/2} R^{D-2}/\Gamma\left(\frac{D-1}{2}\right). \tag{4.4}$$

This result is positive for all $D>1$. It leads to the striking conclusion that the number of mutually perpendicular lines can exceed the dimension of a space, specifically when $2>D>1$.

Strictly speaking, we have not proven that triplets of points $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$ exist with connecting lines at exactly

a right angle. The result on a density being positive in the neighborhood of this configuration is a weaker statement. However the spaces S_D are dense, so the distinction for most purposes is unimportant.

The three-center weight has the following lengthy form (valid for all real D):

$$W_3(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3\,|\,r_{01},r_{02},r_{03}) = 2^{7-2D}\sigma(D-2)\,r_{01}r_{02}r_{03}[\Delta(r_{12},r_{13},r_{23})]^{3-D} \times\{-r_{12}^2r_{13}^2r_{23}^2+r_{01}^2(r_{12}^2r_{23}^2+r_{13}^2r_{23}^2-r_{23}^4) +r_{02}^2(r_{12}^2r_{13}^2+r_{13}^2r_{23}^2-r_{13}^4)+r_{03}^2(r_{12}^2r_{13}^2+r_{12}^2r_{23}^2-r_{12}^4) -r_{01}^4r_{23}^2-r_{02}^4r_{13}^2-r_{03}^4r_{12}^2+r_{01}^2r_{02}^2(r_{13}^2+r_{23}^2-r_{12}^2) +r_{01}^2r_{03}^2(r_{12}^2+r_{23}^2-r_{13}^2)+r_{02}^2r_{03}^2(r_{12}^2+r_{13}^2-r_{23}^2)\}^{(D-4)/2}. \tag{4.5}$$

The density of mutually perpendicular lines in S_D can be extracted from this formula upon setting $r_{01}=r_{02}=r_{03}=R$, and $r_{12}=r_{13}=r_{23}=2^{1/2}R$. This yields

$$W_3(2^{1/2}R\cdots|R\cdots)=2\pi^{(D-2)/2}R^{D-3}/3^{(D-3)/2}\,\Gamma\left(\frac{D-2}{2}\right), \tag{4.6}$$

indicating a positive density for all $D>2$. Analogous to the preceding case, we have found that the number of mutually perpendicular lines can exceed the dimension. The orthodox position that the maximal number of mutually perpendicular lines gives D is not valid in the present context.

Careful study of the W_n, using Eq. (3.6), shows that they always consist of nonnegative factors divided by an uncompensated term $\Gamma[\frac{1}{2}(D-n+1)]$. When sets of distances serving as arguments for W_n are chosen so that this weight does not vanish identically, then the incorporated term $1/\Gamma[\frac{1}{2}(D-n+1)]$ will cause sign alternation as a function of D when $D<n-1$. In particular, one will have

$$W_n<0 \quad (n-4j-3<D<n-4j-1),$$
$$=0 \quad (D=n-2j-1), \tag{4.7}$$

where $j=0,1,2,3,\cdots$. For all other values of D, W_n will be positive. The occurrence of negative integration weights for noninteger D indicates a complicated and unanticipated "geometric" structure for the S_D. In fact with finite noninteger D an arbitrary number of mutually perpendicular lines can be erected, though the corresponding weights $W_n(2^{1/2}R\cdots|R\cdots)$ will have indefinite signs. The possibility of continuously variable D evidently has been bought at the expense of negative integration weights, which have no precedent in ordinary geometry.

If M lines emanate from point $\mathbf{x}_0$ in S_D, projection of any $\mathbf{x}$ along each of these lines can be computed by means of Eq. (2.7); they might be denoted by

$$p_1(\mathbf{x}),\ p_2(\mathbf{x}),\ \cdots,p_M(\mathbf{x}). \tag{4.8}$$

Provided that D is not an integer, the M lines can be chosen to be arbitrarily close to perpendicular to one another, regardless of how large M might be. These lines can then be regarded as a set of orthogonal axes along which the "pseudocoordinates" $p_1\cdots p_M$ are measured. There are several fundamental questions about these pseudocoordinates that deserve eventually to be investigated, such as:

(a) If $M\geq D$, and $\mathbf{x}\neq\mathbf{y}$, are the sets of pseudocoordinates for $\mathbf{x}$ and for $\mathbf{y}$ distinct?

(b) Are straight lines in S_D always representable as linear parametric expressions in terms of pseudocoordinates?

(c) Under what circumstances can pseudocoordinates serve as integration variables?

(d) What is the content ("volume") of the region defined by $0<p_j<L$, if $M\geq D$?

(e) Can pseudocoordinates be used to carry the concept of convex regions into the noninteger-D regime?

(f) How do pseudocoordinates transform under the translation and rotation groups in S_D?

(g) How can the Pythagorean formula, which for $M=D=$integer has the form

$$[r(\mathbf{x},\mathbf{y})]^2=\sum_{j=1}^{M}[p_j(\mathbf{x})-p_j(\mathbf{y})]^2, \tag{4.9}$$

be generalized to arbitrary M and D? In particular, do $D\leq M$ and $D\geq M$ require corresponding inequalities between the members of Eq. (4.9)?

We shall not consider these open problems any further in the remainder of this exposition.

V. CONVOLUTION THEOREM

In Euclidean spaces ($D=$integer), integrals of the type

$$T\{f,h\}=\int d\mathbf{r}_3\,f(r_{13})\,h(r_{23}) \tag{5.1}$$

can be reduced to simpler quadratures by introducing Fourier transforms for the functions f and h. The general reduction scheme is usually referred to as the "convolution theorem[14]", whose extension to noninteger D we now identify.

To prepare the way for introduction of Fourier transforms in S_D, it will first be necessary to have an integration weight in terms of quantities p and l [Eqs. (2.7)–(2.9)]. This can be produced from the general doubly-rooted Gaussian integral, which we now write in the following manner:

$$\int d\mathbf{x}_0\exp(-\alpha_1r_{01}^2-\alpha_2r_{02}^2) =\int_{-\infty}^{+\infty}dp\int_0^{\infty}dl\,W(p,l)\exp[-\alpha_1p^2-\alpha_2(r_{12}-p)^2 -(\alpha_1+\alpha_2)l^2]. \tag{5.2}$$

Here the fixed points 1 and 2 are separated by r_{12}, $p\equiv p(\mathbf{x}_0\,|\,\mathbf{x}_1,\mathbf{x}_2)$ is the projection of $(\mathbf{x}_1,\mathbf{x}_0)$ on $(\mathbf{x}_1,\mathbf{x}_2)$, and l is its orthogonal complement.

Since S_D is uniform, W cannot depend on position p measured along the arbitrary axis passing through $\mathbf{x}_1$ and $\mathbf{x}_2$. This fact permits the p integral in Eq. (5.2) to be carried out explicitly. Furthermore, A5. specifies the value to be assigned to (5.2), so we have

$$\left(\frac{\pi}{\alpha_1+\alpha_2}\right)^{(D-1)/2} = \int_0^\infty dl\, W(p,l)\exp[-(\alpha_1+\alpha_2)l^2]. \tag{5.3}$$

This is equivalent to a Laplace transform, and the inversion operation leads to the result

$$W(p,l) = \sigma(D-1)\,l^{D-2}. \tag{5.4}$$

Comparing this result with Eq. (3.8), we see that $W(p,l)$ is equivalent to a radial weight (with $r=l$) for $\mathcal{S}_{D-1}$. This confirms the expectation that a constant-p subspace in $\mathcal{S}_D$ has dimension $D-1$, and we note in passing that this subspace can be proved to have all other integration properties of the type embodied in A5.

Now we are in a position to evaluate the "Fourier transform" of a Gaussian function in $\mathcal{S}_D$,

$$G(k) = \int d\mathbf{x}\exp[-\alpha r^2(\mathbf{x}) + ikp(\mathbf{x})]. \tag{5.5}$$

Here $p(x)$ is the projection along a preselected axis through the origin. Using Eq. (5.4) we have

$$\begin{aligned} G(k) &= \int_{-\infty}^{+\infty} dp\exp(-\alpha p^2+ikp)\int_0^\infty dl\,\sigma(D-1)\,l^{D-2}\exp(-\alpha l^2) \\ &= (\pi/\alpha)^{D/2}\exp(-k^2/4\alpha). \end{aligned} \tag{5.6}$$

Identifying parameter k as a distance function $k(\mathbf{x})$ in $\mathcal{S}_D$, with $p(\mathbf{x})$ the corresponding projection, we also derive

$$\begin{aligned} g(r) &= (2\pi)^{-D}\int d\mathbf{x}\,G[k(\mathbf{x})]\exp[-irp(\mathbf{x})] \\ &= (4\pi\alpha)^{-D/2}\int d\mathbf{x}\exp\{-[k^2(\mathbf{x})/4\alpha] - irp(\mathbf{x})\} \\ &= \exp(-\alpha r^2). \end{aligned} \tag{5.7}$$

This constitutes the inverse to transform Eq. (5.5).

Consider next the class of functions which consist of linear combinations of Gaussians,

$$f(r) = \sum_{j=1}^m A_j\exp(-a_j r^2). \tag{5.8}$$

Our generalized Fourier transformation is linear, so that

$$F(k) = \sum_{j=1}^m A_j(\pi/a_j)^{D/2}\exp(-k^2/4a_j) \tag{5.9}$$

is the corresponding transform function. At least within this function class, the symbolic Fourier transform pair has the following appearance:

$$F(k) = \int d\mathbf{x}\,f[r(\mathbf{x})]\exp[ikp(\mathbf{x})], \tag{5.10a}$$

$$f(r) = (2\pi)^{-D}\int d\mathbf{x}\,F[k(\mathbf{x})]\exp[-irp(\mathbf{x})]. \tag{5.10b}$$

Define the following two-center integral (fixed points 1 and 2) in $\mathcal{S}_D$:

$$T(r_{12}) = \int d\mathbf{x}_0\, f(r_{10})\,h(r_{02}), \tag{5.11}$$

where both f and h belong to the function class denoted by Eq. (5.8). Thus

$$\begin{aligned} T(r_{12}) &= \int d\mathbf{x}_0\left[\sum_{j=1}^m A_j\exp(-a_j r_{10}^2)\right]\left[\sum_{l=1}^n B_l\exp(-b_l r_{02}^2)\right] \\ &= \sum_{j=1}^m\sum_{l=1}^n A_jB_l\int d\mathbf{x}_0\exp(-a_jr_{10}^2 - b_lr_{02}^2) \\ &= \sum_{j=1}^m\sum_{l=1}^n A_jB_l\left(\frac{\pi}{a_j+b_l}\right)^{D/2}\exp\left[-\left(\frac{a_jb_l}{a_j+b_l}\right)r_{12}^2\right], \end{aligned} \tag{5.12}$$

by invoking A5. The remaining Gaussian factor may itself be written as an integral,

$$\begin{aligned} &\exp\left[-\left(\frac{a_jb_l}{a_j+b_l}\right)r_{12}^2\right] \\ &= \left(\frac{a_j+b_l}{4\pi a_jb_l}\right)^{D/2}\int d\mathbf{x}\exp\left[-\left(\frac{a_j+b_l}{4a_jb_l}\right)k^2(\mathbf{x}) \right. \\ &\quad \left. - ir_{12}p(\mathbf{x})\right]. \end{aligned} \tag{5.13}$$

Substituting and rearranging we have

$$\begin{aligned} T(r_{12}) &= (2\pi)^{-D}\int d\mathbf{x}\exp[-ir_{12}p(\mathbf{x})] \\ &\quad\times\left\{\sum_{j=1}^m\sum_{l=1}^n A_jB_l\left(\frac{\pi}{a_j}\right)^{D/2}\left(\frac{\pi}{b_l}\right)^{D/2}\right. \\ &\quad\left.\times\exp\left[-\frac{1}{4}\left(\frac{1}{a_j}+\frac{1}{b_l}\right)k^2(\mathbf{x})\right]\right\} \\ &= (2\pi)^{-D}\int d\mathbf{x}\exp[-ir_{12}p(\mathbf{x})]\,F[k(\mathbf{x})]\,H[k(\mathbf{x})]. \end{aligned} \tag{5.14}$$

This is the desired convolution theorem. Similarly to the case with integer D, it reduces the evaluation of doubly-rooted integrals to an integral of the product of Fourier transforms.

Aside from complex exponentials, integrals of the types (5.10) and (5.14) involve only functions of distance. Consequently they may be simplified. Starting with the prototype integral

$$\begin{aligned} I &= \int d\mathbf{x}\,\phi[r(\mathbf{x})]\exp[ikp(\mathbf{x})] \\ &\equiv \int_{-\infty}^{+\infty} dp\int_0^\infty dl\, W(p,l)\,\phi(p^2+l^2)\exp(ikp), \end{aligned} \tag{5.15}$$

we introduce the change of variables

$$l = r\sin\Theta, \quad p = r\cos\Theta \tag{5.16}$$

[recall Eqs. (2.5)–(2.9)]. This allows one to express I as follows

$$\begin{aligned} I &= \sigma(D-1)\int_0^\infty dr\int_0^\pi d\Theta\, r^{D-1}(\sin\Theta)^{D-2} \\ &\quad\times\exp(ikr\cos\Theta)\,\phi(r), \end{aligned} \tag{5.17}$$

upon using Eq. (5.4) for $W(p,l)$. By expanding the exponential function, the Θ integral may be carried out explicitly (after recognizing the Bessel function series),

$$I = (2\pi)^{D/2}\int_0^\infty dr\,(kr)^{(2-D)/2}J_{(D-2)/2}(kr)\,\phi(r). \tag{5.18}$$

By employing result (5.18), we simplify the D-dimensional Fourier transform pair (5.10) to

$$F(k) = (2\pi)^{D/2}\int_0^\infty dr\,(kr)^{(2-D)/2}J_{(D-2)/2}(kr)\,f(r), \tag{5.19a}$$

$$f(r) = (2\pi)^{-D/2} \int_0^\infty dk\,(rk)^{(2-D)/2} J_{(D-2)/2}(rk)\, F(k). \qquad (5.19b)$$

In analogous fashion, the convolution theorem (5.14) can be written

$$T(r_{12}) = (2\pi)^{-D/2} \int_0^\infty dk\,(r_{12}k)^{(2-D)/2} J_{(D-2)/2}(r_{12}k)\, F(k)\, H(k), \qquad (5.20)$$

for evaluation of the doubly-rooted integral (5.11).

Equations (5.18)—(5.20) are Hankel transforms,[15] with minor modifications. We see that they arise naturally in spaces with fractional dimension. Although we have derived (5.18)—(5.20) only for the limited class of functions shown in Eq. (5.8), consisting of finite sums of Gaussians, it should be clear that extension is possible to convergent sequences of such functions. The available general theory of Hankel transforms[16] is relevant in this connection.

VI. LAPLACE OPERATOR

For the moment, we revert to the special case that D is a positive integer, so that $\mathcal{S}_D$ can be treated as a conventional vector space. A form of the linear Laplace operator ∇^2 will be constructed which will serve as a convenient device for extension to noninteger D.

Introduce a "local weighting function" $w(r)$ with the following properties:

$$\lim_{r\to\infty} w(r) = 0, \qquad (6.1a)$$

$$\int d\mathbf{r}\, w(r) = 0, \qquad (6.1b)$$

$$\int d\mathbf{r}\, r^2 w(r) \equiv w_2 \neq 0. \qquad (6.1c)$$

Then for any function $f(\mathbf{r})$ defined over the vector space, consider the integral (we assume it converges),

$$L(\mathbf{r}_1, \xi) = \xi^{D+2} \int d\mathbf{r}\, w(\xi\,|\mathbf{r}-\mathbf{r}_1|)\, f(\mathbf{r}), \quad \xi > 0. \qquad (6.2)$$

When ξ is large, the integrand will differ from zero only in the immediate neighborhood of the point $\mathbf{r}_1$. Presuming that f is at least twice differentiable, it would then suffice to represent this function in L by the leading terms in its multiple Taylor expansion about $\mathbf{r}_1$

$$L(\mathbf{r}_1, \xi) = \xi^{D+2} \int d\mathbf{r}\, w(\xi\,|\mathbf{r}-\mathbf{r}_1|)[f(\mathbf{r}_1) + (\mathbf{r}-\mathbf{r}_1)\cdot\nabla f(\mathbf{r}_1) + \tfrac{1}{2}(\mathbf{r}-\mathbf{r}_1)(\mathbf{r}-\mathbf{r}_1) : \nabla\nabla f(\mathbf{r}_1) + \cdots]. \qquad (6.3)$$

In the limit $\xi \to +\infty$, the remainder beyond terms shown should be negligible, so we drop it.

The leading term in Eq. (6.3) vanishes, due to condition (6.1b). The next term (containing ∇f) also vanishes by symmetry. Therefore, we need only examine the quadratic terms, which may now be written out explicitly,

$$L(\mathbf{r}_1, \xi) = \tfrac{1}{2}\xi^{D+2} \sum_{i,j=1}^{D} \frac{\partial^2 f(\mathbf{r}_1)}{\partial x_i\, \partial x_j} \int d\mathbf{r}\, w(\xi\,|\mathbf{r}-\mathbf{r}_1|) \times (x_i - x_{i1})(x_j - x_{j1}) + \cdots. \qquad (6.4)$$

Here the separate vector components have been denoted by x_i, etc. Only those integrals with $i=j$ in (6.4) survive. Since

$$\int d\mathbf{r}\, w(\xi\,|\mathbf{r}-\mathbf{r}_1|)(x_i - x_{i1})^2 = w_2/(D\xi^{D+2}), \qquad (6.5)$$

we have

$$L(\mathbf{r}_1, \xi) = \frac{w_2}{2D} \sum_{i=1}^{D} \frac{\partial^2 f(\mathbf{r}_1)}{\partial x_i^2} + \cdots \equiv \frac{w_2}{2D} \nabla^2 f(\mathbf{r}_1) + \cdots. \qquad (6.6)$$

In the limit, one has the following identity:

$$\nabla^2 f(\mathbf{r}_1) = \frac{2D}{w_2} \lim_{\xi\to\infty} \xi^{D+2} \int d\mathbf{r}\, w(\xi\,|\mathbf{r}-\mathbf{r}_1|)\, f(\mathbf{r}). \qquad (6.7)$$

The right side of the last equation immediately suggests the form in which a linear Laplace operator (which we continue to denote by ∇^2) ought to be defined for noninteger D,

$$\nabla^2 f(\mathbf{x}_1) = \frac{2D}{w_2} \lim_{\xi\to\infty} \xi^{D+2} \int d\mathbf{x}_2\, w(\xi r_{12})\, f(\mathbf{x}_2), \qquad (6.8)$$

where Eqs. (6.1) are taken over to $\mathcal{S}_D$ in the obvious way. We now explore some implications of this definition.

One of the simplest cases to which Eq. (6.8) can be applied is that in which f depends only on radial distance r_{02} from some origin $\mathbf{x}_0$. For this case the Laplacian to be evaluated will depend only on distance r_{01},

$$\nabla^2 f(r_{01}) = \frac{2D}{w_2} \lim_{\xi\to\infty} \xi^{D+2} \int d\mathbf{x}_2\, w(\xi r_{12})\, f(r_{02}). \qquad (6.9)$$

On account of the (large) scale factor ξ that occurs in the variable for w, attention need only be focused on the region of small r_{12}. Referring to Fig. 2, we have

$$r_{02} = r_{01}\left\{1 - 2\cos\Theta\left(\frac{r_{12}}{r_{01}}\right) + \left(\frac{r_{12}}{r_{01}}\right)^2\right\}^{1/2} = r_{01}\left\{1 - \cos\Theta\left(\frac{r_{12}}{r_{01}}\right) + \left(\tfrac{1}{2} - \tfrac{1}{2}\cos^2\Theta\right)\left(\frac{r_{12}}{r_{01}}\right)^2 + O\left[\left(\frac{r_{12}}{r_{01}}\right)^3\right]\right\}. \qquad (6.10)$$

This expansion may be used in conjunction with the Taylor expansion for f to yield the following:

$$\nabla^2 f(r_{01}) = \frac{2D}{w_2} \lim_{\xi\to\infty} \xi^{D+2} \int d\mathbf{x}_2\, w(\xi r_{12})\Big[f(r_{01}) - r_{12}\cos\Theta f'(r_{01}) + r_{12}^2\left(\left(\tfrac{1}{2} - \tfrac{1}{2}\cos^2\Theta\right)\frac{f'(r_{01})}{r_{01}} + \tfrac{1}{2}\cos^2\Theta f''(r_{01})\right) + O(r_{12}^3)\Big]. \qquad (6.11)$$

Equation (6.1b) causes the $f(r_{01})$ term in this last expression to vanish; the following term (proportional to

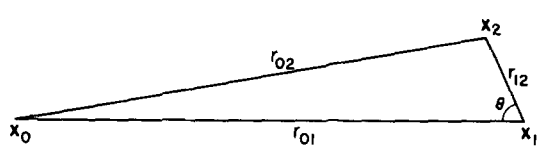


FIG. 2. Triangle used in evaluation of Eq. (6.9).

r_{12}) vanishes by symmetry. Furthermore the $O(r_{12}^3)$ remainder will vanish in the indicated limit. Consequently we are left with

$$\nabla^2 f(r_{01}) = \left[\left\langle \cos^2\Theta \right\rangle f''(r_{01}) + \left(1 - \left\langle \cos^2\Theta \right\rangle\right) \frac{f'(r_{01})}{r_{01}} \right], \tag{6.12}$$

where

$$\langle \cos^2\Theta \rangle = (\xi^{D+2}/w_2) \int d\mathbf{x}_2\, r_{12}^2 \cos^2\Theta\, w(\xi r_{12}). \tag{6.13}$$

It is natural to use r_{12} and Θ as integration variables for evaluation of this last average. The proper integration weight

$$\sigma(D-1)(r_{12})^{D-1}(\sin\Theta)^{D-2} \tag{6.14}$$

was obtained earlier in connection with Eq. (5.17). Thus we find

$$\langle \cos^2\Theta \rangle = 1 - \langle \sin^2\Theta \rangle = \frac{\int_0^\pi (\sin\Theta)^D\, d\Theta}{\int_0^\pi (\sin\Theta)^{D-2}\, d\Theta} = 1/D. \tag{6.15}$$

This converts Eq. (6.12) to the desired Laplacian formula,

$$\nabla^2 f(r) = f''(r) + [(D-1)/r] f'(r), \tag{6.16}$$

where for simplicity the distance subscripts have been suppressed. Note that this confirms the compatibility of expressions (1.1) and (1.2) in the Introduction.

It is only a bit more complicated to compute the Laplacian in $\mathcal{S}_D$ for $g(p,l)$, a function of a projection p and its orthogonal complement. From Eq. (6.8) we have

$$\nabla^2 g[p(\mathbf{x}_1), l(\mathbf{x}_1)] = \frac{2D}{w_2} \lim_{\xi \to \infty} \xi^{D+2} \int d\mathbf{x}_2\, w(\xi r_{12}) \times g[p(\mathbf{x}_2), l(\mathbf{x}_2)]. \tag{6.17}$$

Insert into the integrand the Taylor expansion of g through second order,

$$g[p(\mathbf{x}_2), l(\mathbf{x}_2)] = g[p(\mathbf{x}_1), l(\mathbf{x}_1)] + \frac{\partial g}{\partial p}\Delta p + \frac{\partial g}{\partial l}\Delta l + \frac{1}{2}\frac{\partial^2 g}{\partial p^2}(\Delta p)^2 + \frac{\partial^2 g}{\partial p \partial l}\Delta p \Delta l + \frac{1}{2}\frac{\partial^2 g}{\partial l^2}(\Delta l)^2 + \cdots, \tag{6.18}$$

where

$$\Delta p = p(\mathbf{x}_2) - p(\mathbf{x}_1), \quad \Delta l = l(\mathbf{x}_2) - l(\mathbf{x}_1). \tag{6.19}$$

The first, second, and fifth terms in (6.18) integrate to zero. The remaining three terms may be evaluated by the procedure used to derive the earlier result (6.16). One finally obtains

$$\nabla^2 g(p,l) = \left[\frac{\partial^2}{\partial p^2} + \frac{\partial^2}{\partial l^2} + \frac{D-2}{l}\frac{\partial}{\partial l} \right] g(p,l). \tag{6.20}$$

For many applications it may be convenient to transform variables in g from p and l to the polar variables r and Θ,

$$p = r\cos\Theta, \quad l = r\sin\Theta. \tag{6.21}$$

The Laplacian in Eq. (6.20) may be transformed according to the standard procedure[17] to yield

$$\nabla^2 g = \left[\frac{\partial^2}{\partial r^2} + \frac{D-1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2 \sin^{D-2}\theta}\frac{\partial}{\partial \Theta} \sin^{D-2}\Theta \frac{\partial}{\partial \Theta} \right] g. \tag{6.22}$$

VII. SCHRÖDINGER WAVE MECHANICS

Using suitable reduced units, the quantum-mechanical motion of a particle subject to potential U is described by the time-dependent Schrödinger equation

$$[-\tfrac{1}{2}\nabla^2 + U]\Psi = i\frac{\partial \Psi}{\partial t}. \tag{7.1}$$

The general solution consists of a linear superposition of terms

$$\psi \exp(-iEt), \tag{7.2}$$

where the spatial wavefunctions ψ obey the spatial wave equation

$$[-\tfrac{1}{2}\nabla^2 + U - E]\psi = 0. \tag{7.3}$$

In seeking solutions to Eq. (7.3) relevant to unbounded space, both square-integrable eigenfunctions (bound states) and scattering solutions (asymptotic plane waves) normally are sought.

In view of our generalized Laplace operator, Eq. (6.8), it is now possible to extend study of the Schrödinger wave equation to spaces with noninteger dimension. We examine several simple examples.

Let U be restricted to central form, i.e., it will depend only on radial distance r from some chosen origin in $\mathcal{S}_D$. We then search for solutions to the generalized spatial equation (7.3) which have the form $\psi(r,\theta)$. Here angle θ is measured relative to *any* axis in $\mathcal{S}_D$ passing through the origin. Appealing to Eq. (6.22), we have

$$\left[\frac{1}{r^{D-1}}\frac{\partial}{\partial r} r^{D-1}\frac{\partial}{\partial r} + \frac{1}{r^2 \sin^{D-2}\theta}\frac{\partial}{\partial \theta}\sin^{D-2}\theta\frac{\partial}{\partial \theta} + 2E - 2U(r) \right] \psi(r,\theta) = 0. \tag{7.4}$$

This equation is separable; set

$$\psi(r,\theta) = R(r)\,\Theta(\theta). \tag{7.5}$$

The resulting radial and angular differential equations are the following:

$$\left[\frac{d^2}{d\theta^2} + (D-2)\cot\theta\frac{d}{d\theta} + \Lambda(\Lambda + D - 2) \right] \Theta(\theta) = 0, \tag{7.6}$$

$$\left[\frac{d^2}{dr^2} + \frac{D-1}{r}\frac{d}{dr} + 2E - 2U(r) - \frac{\Lambda(\Lambda+D-2)}{r^2} \right] R(r) = 0. \tag{7.7}$$

The appropriate solutions to angular Eq. (7.6) are Gegenbauer polynomials in $\cos\theta$ [18]:

$$\Theta(\theta) = C_\Lambda^{(D/2-1)}(\cos\theta), \quad \Lambda = 0, 1, 2, 3, \cdots. \tag{7.8}$$

These polynomials satisfy the following orthogonality relation:

$$\int_0^\pi C_\Lambda^{(D/2-1)}(\cos\theta)\, C_{\Lambda'}^{(D/2-1)}(\cos\theta) \sin^{D-2}\theta\, d\theta = h(\Lambda)\,\delta(\Lambda,\Lambda'), \tag{7.9}$$

where

$$h(\Lambda)=\frac{2^{3-D}\pi\Gamma(\Lambda+D-2)}{\Lambda!\,(\Lambda+\frac{1}{2}D-1)[\Gamma(\frac{1}{2}D-1)]^2}. \tag{7.10}$$

The first few Gegenbauer polynomials are

$$\begin{aligned} &C_0^{(D/2-1)}(z)=1,\\ &C_1^{(D/2-1)}(z)=(D-2)z,\\ &C_2^{(D/2-1)}(z)=(\tfrac{1}{2}D-1)(Dz^2-1). \end{aligned} \tag{7.11}$$

The nature of solutions to the radial equation naturally depends on U. The simplest case is that for free-particle motion, $U\equiv 0$. The radial solutions are then found to be expressible in terms of Bessel functions,

$$R(r)=(kr)^{1-D/2}J_{D/2+\Lambda-1}(kr),\quad k=(2E)^{1/2}. \tag{7.12}$$

Free particle motion can just as well be described by a "plane wave" ψ. In $\mathcal{S}_D$ the appropriate form is

$$\psi(\mathbf{x})=\exp[ikp(\mathbf{x})], \tag{7.13}$$

where $p(\mathbf{x})$ represents the projection of point $\mathbf{x}$ along the chosen polar axis. This polar axis is the direction of propagation. The "plane wave" may be expanded as follows [recall $p(\mathbf{x})=r(\mathbf{x})\cos\theta$]:

$$\exp[ikp(\mathbf{x})]=\sum_{\Lambda=0}^{\infty}A_\Lambda C_\Lambda^{(D/2-1)}(\cos\theta)(kr)^{1-D/2}J_{D/2+\Lambda-1}(kr),$$

$$A_\Lambda=2^{D/2-1}(\tfrac{1}{2}D+\Lambda-1)\,\Gamma(\tfrac{1}{2}D-1)\,i^\Lambda. \tag{7.14}$$

By choosing

$$U(r)=\tfrac{1}{2}Kr^2 \tag{7.15}$$

we obtain an isotropic harmonic oscillator in $\mathcal{S}_D$. The corresponding discrete spectrum results from the requirement that the radial function $R_n(r)$ vanish at infinity. With this boundary condition the solutions involve generalized Laguerre polynomials,[18]

$$\begin{aligned} &R(r)=\exp(-\tfrac{1}{2}s^2)\,s^\Lambda L_n^{(D/2+\Lambda-1)}(s^2),\\ &s=K^{1/4}r,\quad n=0,1,2,3,\cdots. \end{aligned} \tag{7.16}$$

The corresponding energy eigenvalues are

$$E=K^{1/2}(\tfrac{1}{2}D+\Lambda+2n). \tag{7.17}$$

The lowest-order generalized Laguerre polynomials have the following explicit forms:

$$\begin{aligned} &L_0^{(D/2+\Lambda-1)}(z)=1,\quad L_1^{(D/2+\Lambda-1)}(z)=\tfrac{1}{2}D+\Lambda-z,\\ &L_2^{(D/2+\Lambda-1)}(z)=\tfrac{1}{2}z^2-(\tfrac{1}{2}D+\Lambda+1)z+\tfrac{1}{2}(\tfrac{1}{2}D+\Lambda)(\tfrac{1}{2}D+\Lambda+1). \end{aligned} \tag{7.18}$$

The "Coulomb" problem in D dimensions for present purposes will refer to the inverse-distance potential ($Z\geq 0$)

$$U(r)=-Z/r. \tag{7.19}$$

[An alternative convention might have been adopted, of course, with U proportional to the radial Green's function for our D-dimensional Laplacian.] The corresponding radial equation

$$\left[\frac{d^2}{dr^2}+\frac{D-1}{r}\frac{d}{dr}+2E+\frac{2Z}{r}-\frac{\Lambda(\Lambda+D-2)}{r^2}\right]R(r)=0 \tag{7.20}$$

has solutions regular at the origin which may be written in terms of the confluent hypergeometric function $M(a,b,z)$.[19] Setting

$$E=-\tfrac{1}{2}\kappa^2, \tag{7.21}$$

one finds,

$$R(r)=r^\Lambda\exp(-\kappa r)\,M(\Lambda+\tfrac{1}{2}D-\tfrac{1}{2}-Z/\kappa,\ 2\Lambda+D-1,2\kappa r). \tag{7.22}$$

These radial functions are square-integrable only for discrete values of κ, which in fact cause M to reduce to a polynomial in r. The criterion for this reduction is the following:

$$\begin{aligned} &\Lambda+\tfrac{1}{2}D-\tfrac{1}{2}-Z/\kappa=1+\Lambda-n,\\ &n=\Lambda+1,\ \Lambda+2,\ \Lambda+3,\cdots, \end{aligned} \tag{7.23}$$

which introduces the principal quantum number n. Equation (7.23) may be written in terms of E to show the spectrum of bound-state energies,

$$E=-Z^2/2(n+\tfrac{1}{2}D-\tfrac{3}{2})^2. \tag{7.24}$$

It is noteworthy that orbital degeneracy continues to exist for $D\neq 3$.[20] For each n, the eigenfunctions with $\Lambda=0,1,\ldots,n-1$ all possess the same energy.

Explicit polynomial forms for the M functions may easily be computed. Some of the simpler cases are now listed.

$$\begin{aligned} &n=1,\ \Lambda=0\colon\ M(0,D-1,z)=1,\\ &n=2,\ \Lambda=0\colon\ M(-1,D-1,z)=1-z/(D-1),\\ &n=2,\ \Lambda=1\colon\ M(0,D+1,z)=1,\\ &n=3,\ \Lambda=0\colon\ M(-2,D-1,z)=1-\frac{2z}{D-1}+\frac{z^2}{D(D-1)},\\ &n=3,\ \Lambda=1\colon\ M(-1,D+1,z)=1-\frac{z}{D+1},\\ &n=3,\ \Lambda=2\colon\ M(0,D+3,z)=1. \end{aligned} \tag{7.25}$$

When D is an integer, the set of solutions $\psi(r,\theta)$, including all possible polar axes, generates the full set of solutions to the spatial wave equation, by taking appropriate linear combinations. Presumably the same is true for noninteger D, but a proof is presently lacking. More to the point, it is not yet clear how one can identify a complete orthogonal set of solutions.

VIII. CLASSICAL PARTITION FUNCTION

Consider N structureless particles of mass m, confined to a region Ω with integer dimension D. Let vectors $\mathbf{R}_1\cdots\mathbf{R}_N$ and $\mathbf{P}_1\cdots\mathbf{P}_N$ denote the positions and momenta, respectively, and let $\Phi(\mathbf{R}_1\cdots\mathbf{R}_N)$ be the interaction potential. The classical partition function has the following form:

$$Z_N=(1/N!h^{DN})\int_\Omega d\mathbf{R}_1\cdots\int_\Omega d\mathbf{R}_N\int d\mathbf{P}_1\cdots\int d\mathbf{P}_N$$

$$\times\exp[-(\beta/2m)\sum_{j=1}^{N}P_j^2-\beta\Phi(\mathbf{R}_1\cdots\mathbf{R}_N)]. \tag{8.1}$$

Here h is Planck's constant, and $\beta=1/k_BT$ is the inverse temperature parameter. Contact between Z_N and

thermodynamic properties for the system of particles is provided by the Helmholtz free energy F,

$$\beta F = -\ln Z_N. \tag{8.2}$$

In the large system limit, with fixed temperature and density N/Ω, the free energy per particle F/N becomes independent of Ω, provided that this region is such that most particles are far from its boundary. In this limit, any convenient shape for Ω can then be employed, such as the D-dimensional "sphere" of appropriate radius.

In seeking to extend Z_N to noninteger D, procedures must be indentified for carrying out both momentum and position integrations. The former provide no difficulty, since Axiom A5. immediately affords the result

$$h^{-D}\int d\mathbf{P}_j \exp(-\beta P_j^2/2m) = \lambda_T^{-D}, \tag{8.3}$$

where λ_T is the mean thermal deBroglie wavelength,

$$\lambda_T = h/(2\pi m k_B T)^{1/2}. \tag{8.4}$$

However the position integration requires more detailed consideration.

We can use the integration weights W_n introduced by Eq. (3.2), and treat the position integrations as a multiple integral over all distances. The distances involved of course include the $N(N-1)/2$ interparticle separations r_{ij}. However we shall in fact treat Ω as a D-dimensional "sphere" (with radius L), so that the N distances r_{0i} of the particles from its center are also relevant. Without significant loss of generality, we can suppose that the potential energy Φ is a function just of the r_{ij}.

Under these circumstances, Z_N can be put into the following form:

$$\begin{aligned} Z_n = (1/N!\lambda_T^{DN}) &\int_0^L dr_{01}\, W_1(0|r_{01}) \int_0^L dr_{02} \\ &\times \int_0^{2L} dr_{12}\, W_2(0,1|r_{02},r_{12}) \int_0^L dr_{03} \int_0^{2L} dr_{13} \\ &\times \int_0^{2L} dr_{23}\, W_3(0,1,2|r_{03},r_{13},r_{23}) \cdots \int_0^L dr_{0N} \\ &\times \int_0^{2L} dr_{1N} \int_0^{2L} dr_{N-1,N}\, W_N(0\cdots N|r_{0N}\cdots r_{N-1,N}) \\ &\times \exp[-\beta\Phi(r_{12}\cdots r_{N-1,N})]. \end{aligned} \tag{8.5}$$

Strictly speaking, the upper limits $2L$ on the r_{ij} $(0<i,j)$ integrals could be extended to infinity, since the affected weights would automatically vanish over the extension.

Evaluation of Z_N in form (8.5) represents no less a formidable challenge than its integer-D predecessor in Eq. (8.1). Nevertheless some of the standard techniques in statistical mechanics can be carried over. In particular it is possible to develop the Ursell–Mayer cluster theory[21] for noninteger D. For this purpose we make the conventional simplification that Φ consists of a sum of central pair potentials,

$$\Phi = \sum_{i<j=1}^{N} \phi(r_{ij}). \tag{8.6}$$

Then the Boltzmann factor $\exp(-\beta\Phi)$ in the partition function may be developed into a sum of products of Mayer f functions,

$$\begin{aligned} \exp(-\beta\Phi) &= \prod_{i<j=1}^{N} [1+f(r_{ij})] \\ &= 1 + \sum_{i<j=1}^{N} f(r_{ij}) + \sum_{i<j<k=1}^{N} [f(r_{ij})f(r_{jk}) \\ &\quad + f(r_{ij})f(r_{ik}) + f(r_{ik})f(r_{jk}) \\ &\quad + f(r_{ij})f(r_{ik})f(r_{jk})] + \cdots, \end{aligned} \tag{8.7}$$

where

$$f(r_{ij}) = \exp[-\beta\phi(r_{ij})] - 1. \tag{8.8}$$

At this stage one can essentially follow the usual cluster-theory procedure.[21] The only novel feature is the necessity to use contraction properties (3.13) for the weights W_n in the case of integrals containing sets of distances in only a trivial way. Finally one obtains the irreducible cluster expansion for the Helmholtz free energy; in the large-system limit the result has the following form:

$$\frac{\beta F}{N} = \ln\left(\frac{\lambda_T^D N}{e\Omega}\right) - \sum_{k=1}^{\infty} \frac{\beta_k}{k+1}\left(\frac{N}{\Omega}\right)^k. \tag{8.9}$$

The β_k are sums of irreducible cluster integrals for $k+1$ particles, and may be expressed thus,

$$\begin{aligned} \beta_k = (1/k!) &\int_0^\infty dr_{12}\, W_1(1|r_{12}) \cdots \int_0^\infty dr_{1,k+1} \cdots \int_0^\infty dr_{k,k+1} \\ &W_k(1\cdots k|r_{1,k+1}\cdots r_{k,k+1})\, S_k(r_{12}\cdots r_{k,k+1}). \end{aligned} \tag{8.10}$$

Here S_k is the sum of those f-function products for the $k+1$ particles which correspond to connected graphs without articulation points.

The pressure p for the N-particle system may be obtained from F by the relation

$$p = -(\partial F/\ \Omega)_{N,\beta}. \tag{8.11}$$

Within the convergence radius of the cluster expansion (8.9) one therefore has

$$\frac{\beta p\Omega}{N} = 1 - \sum_{k=1}^{\infty} \frac{k\beta_k}{k+1}\left(\frac{N}{\Omega}\right)^k, \tag{8.12}$$

which is the usual virial expansion.

The second virial coefficient in the pressure series (8.12) has the following explicit form:

$$B_2 = -\tfrac{1}{2}\beta_1 = -\tfrac{1}{2}\sigma(D)\int_0^\infty dr\, r^{D-1} f(r). \tag{8.13}$$

If the pair potential ϕ describes rigid "spheres" with collision diameters a, then

$$\begin{aligned} f(r) &= -1 \quad (0 \le r \le a), \\ &= 0 \quad\ \ (a < r), \end{aligned} \tag{8.14}$$

so that the general-D second virial coefficient becomes

$$B_2 = \frac{\pi^{D/2} a^D}{D\Gamma(D/2)}. \tag{8.15}$$

The third virial coefficient

$$B_3 = -\tfrac{2}{3}\beta_2 \tag{8.16}$$

involves a single cluster integral whose integrand contains the triangular f product, $f(r_{12})f(r_{13})f(r_{23})$. Specifically,

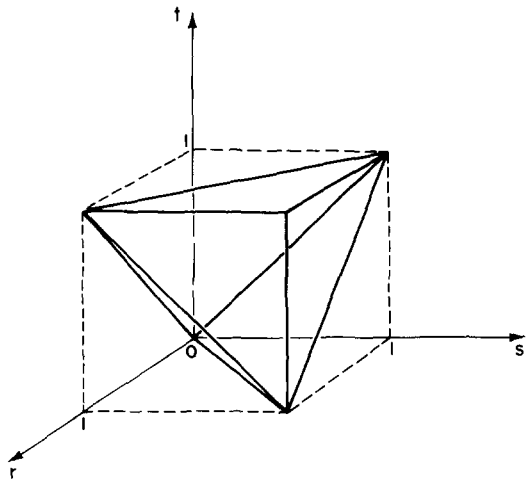


FIG. 3. Hexahedral region (solid lines) over which the generalized third virial coefficient integral (8.19) must be carried out.

$$B_3 = -\tfrac{1}{3}\int_0^\infty dr_{12}\, W_1(1\,|\,r_{12}) f(r_{12}) \int_0^\infty dr_{13} \times \int_0^\infty dr_{23}\, W_2(1,2\,|\,r_{13},r_{23}) f(r_{13}) f(r_{23}). \tag{8.17}$$

Expressions for W_1 and W_2 were derived earlier, and allow B_3 to be written as

$$B_3 = -\frac{2^{5-D}\pi^{D-1/2}}{3\Gamma(D/2)\,\Gamma((D-1)/2)} \int_0^\infty dr_{12} \int_0^\infty dr_{13} \times \int_0^\infty dr_{23}\, f(r_{12}) f(r_{13}) f(r_{23})\, r_{12} r_{13} r_{23} \times [2(r_{12}^2 r_{13}^2 + r_{12}^2 r_{23}^2 + r_{13}^2 r_{23}^2) - r_{12}^4 - r_{13}^4 - r_{23}^4]^{(D-3)/2} \times T_0(r_{12}, r_{13}, r_{23}). \tag{8.18}$$

The function T_0 is present only to ensure that r_{12}, r_{13}, and r_{23} can form a triangle, and if they can it is unity; otherwise T_0 vanishes. This criterion precisely determines the region over which the quartic factor $[\cdots]$ in Eq. (8.18) is positive.

In the case of rigid "spheres," Eq. (8.18) simplifies somewhat,

$$B_3 = \frac{2^{5-D}\pi^{D-1/2} a^{2D}}{3\Gamma(D/2)\Gamma((D-1)/2)} \times \int_0^1 dr \int_0^1 ds \int_0^1 dt\, rst\, T_0(r,s,t) F_D(r,s,t),$$

$$F_D(r,s,t) = [2(r^2s^2 + r^2t^2 + s^2t^2) - r^4 - s^4 - t^4]^{(D-3)/2}. \tag{8.19}$$

Figure 3 shows the region in r, s, t space over which the integral in Eq. (8.19) must be carried out; this hexahedral region is determined both by integration limits and by the condition that T_0 be unity. The figure is useful in transforming expression (8.19) to the following form:

$$B_3 = \frac{2^{6-D}\pi^{D-1/2} a^{2D}[I_A(D) + I_B(D)]}{3\Gamma(D/2)\Gamma((D-1)/2)},$$

$$I_A(D) = \int_{1/2}^1 ds \int_{1-s}^s dt \int_{s-t}^1 dr\, rst\, F_D(r,s,t),$$

$$I_B(D) = \int_0^{1/2} dt \int_t^{1-t} ds \int_{s-t}^{s+t} dr\, rst\, F_D(r,s,t) = \frac{\pi^{1/2}\Gamma((D-1)/2)}{4(D-2)^2\Gamma(D/2)} [F(D-2, 2-D; D-1; \tfrac{1}{2}) - 2^{1-D}],$$

where $F(a,b;c;z)$ is the hypergeometric function.[22] Unfortunately I_A does not simplify significantly unless D is an integer. However the form shown is suited for numerical evaluation, should the need arise.

An alternative route to B_3 would employ the convolution theorem discussed in Sec. V.

Using the three-center weight in Eq. (4.5), explicit (though complicated) integrals can be worked out for the fourth virial coefficient $B_4 = -\frac{3}{4}\beta_3$.

IX. DISCUSSION

The preceding exposition implicitly raises a fundamental physical question. Specifically, should we regard the dimension D of the space in which we live as a possibly noninteger quantity that is locally subject to experimental determination? No one can seriously doubt that our world is locally close to three-dimensional. But how close? Results in Sec. IV above show that it does not help much to exhibit three mutually perpendicular lines, since this provides neither a necessary nor a sufficient condition for D to equal 3.

Probably the most direct experimental approach to determination of D would be the measurement of mass content of a series of homogeneous spherical bodies. The expected result for $D=3$ of course is that this mass would be strictly proportional to the radius (or diameter) cubed. However, accumulated errors in weighing, in size and shape measurement, and in density variations (due to composition and temperature inhomogeneity, and to body stresses) would likely limit the precission in determination of the exponent D to about 1 part in 10^6. By this means one presumably would conclude that D was 3 ± 10^{-6} in our terrestrial locale.

In seeking alternative procedures with greater precision, it might be valuable to examine mathematically how spheres pack when D departs slightly from 3. With $D=3$ exactly, spheres can be fitted together in infinitely extended close packings (f. c. c., h. c. p., or hybrids of the two) with each sphere touching twelve neighbors. If D were slightly larger than 3, attempts to build a known $D=3$ packing outward from a central sphere would begin to produce gaps, eventually allowing extra sphere insertions. By contrast, the case with D slightly less than three would not permit a full complement of spheres to pack properly in the successive shells expected for $D=3$; in terms of material spheres forced into those shells, an accumulation of elastic stress would result. Proper interpretation of the physical construction of large sphere packings thus might help to place tight bounds on our ambient dimension.

In any case, experiments designed to determine D to 1 part in 10^9 or better would likely require the utmost

sophistication in concept and perserverence in execution.

In general relativity, gravitational fields are understood to be geometric perturbations (curvatures) in our spacetime,[23] rather than entities residing within a flat spacetime. The concept that physical force fields generally might be related to purely geometric distortions in space is appealing, and leads one to inquire if dimension D itself might not play an important role as a field variable. The preceding development has considered only uniform spaces S_D for which D had a fixed value. However a more general class of spaces can also be generated within which D varies continuously from point to point (integration weights W_n would exhibit the change explicitly). Under the assumption that general relativity is an incomplete description of reality, it might be appropriate to ask if regions of strong gravitational field display perturbed dimension. More generally, local space dimension may provide geometric field variables in addition to those of general relativity, that would have a place in a unified description of all the forces in nature.

Finally, mention should be made of a paper by Wilson,[24] which also offers an axiomatic description of spaces with noninteger dimension. While most of Wilson's results on integrals appear to be consistent with those deduced here, it is not at all clear that the mathematical spaces generated in the two approaches are isomorphic. In particular, Wilson permits vector addition, and requires an infinite number of vector components when D is not an integer; in the present case vector addition [Eq. (2.10)] has explicitly been excluded, and we have seen that negative integration weights inevitably occur.

ACKNOWLEDGMENTS

I am indebted to D. K. Stillinger for discussions and suggestions concerning several fundamental aspects of this work. I also wish to thank Professor M. E. Fisher for drawing my attention to Ref. 24.

[1]E. Helfand and F.H. Stillinger, J. Chem. Phys. **49**, 1232 (1968); see specifically Sec. V.
[2]K.G. Wilson and M.E. Fisher, Phys. Rev. Lett. **28**, 240 (1972).
[3]M.E. Fisher, Rev. Mod. Phys. **46**, 597 (1974).
[4]C.G. Bollini and J.J. Giambiagi, Nuovo Cimento B **12**, 20 (1972).
[5]G. 't Hooft and M. Veltman, Nuclear Phys. B **44**, 189 (1972).
[6]J.F. Ashmore, Commun. Math. Phys. **29**, 177 (1973).
[7]D.R. Herrick and F.H. Stillinger, Phys. Rev. A **11**, 42 (1975).
[8]C.A. Rogers, *Packing and Covering* (Cambridge U.P., Cambridge, 1964).
[9]J.L. Kelley, *General Topology* (Van Nostrand, New York, 1955), p. 118.
[0]M. Reed and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis* (Academic, New York, 1972), p. 44, Theorem II.5.
[11]R.V. Churchill, *Operational Mathematics* (McGraw-Hill, New York, 1958), Chap. 6.
[12]A. Erdélyi, *Tables of Integral Transforms, Vol. I* (McGraw-Hill, New York, 1954), p. 238, formula (1).
[13]M.G. Kendall, *A Course in the Geometry of n Dimensions* (Hafner, New York, 1961), p. 36.
[14]H.L. Friedman, *Ionic Solution Theory* (Interscience, New York, 1962), p. 127.
[15]Reference 12, Vol. II., Chap. VIII.
[16]E.C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford U.P., London, 1948), 2nd ed., p. 240.
[17]W. Kaplan, *Advanced Calculus* (Addison-Wesley, Reading, Mass., 1973), 2nd ed., p. 172.
[18]M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions*, NBS Applied Mathematics Series, No. 55 (U.S. Government Printing Office, Washington, 1968), Chap. 22.
[19]Reference 18, Chap. 13.
[20]D.R. Herrick, J. Math. Phys. **16**, 281 (1975).
[21]J.E. Mayer and M.G. Mayer, *Statistical Mechanics* (Wiley, New York, 1940), Chap. 13.
[22]Reference 18, Chap. 15.
[23]C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[24]K.G. Wilson, Phys. Rev. D 7, 2911 (1973).

A stochastic derivation of the Klein–Gordon equation

William J. Lehr

Department of Mathematics, University of Petroleum and Minerals, Dhahran, Saudi Arabia

James L. Park

Department of Physics, Washington State University, Pullman, Washington 99163

(Received 17 December 1976)

Several years ago Nelson succeeded in deriving the nonrelativistic Schrödinger equation within a stochastic model which included Newton's second law as the fundamental dynamical rule. Unfortunately, the relativistic extension of Nelson's work is not so straightforward as might at first be supposed. This paper examines the difficulties inherent in such a relativization and proposes supplemental axioms which resolve those difficulties. A stochastic derivation of the Klein–Gordon equation is then presented.

1. INTRODUCTION

Quantum mechanics and the theory of stochastic processes are normally perceived as being rather separate branches of physics. It is possible, however, to establish a formal relationship between them by exploiting an obvious similarity, viz., that each embodies the concept of probability as an irreducible element in its axiomatic framework. The most widely cited effort along these lines was made by Nelson,[1] who succeeded in deriving the nonrelativistic Schrödinger equation as a theorem within a stochastic model based upon Newton's second law. Nelson[2] suggested later that it might be possible to extend his techniques to the relativistic case, but no extension of this kind has so far been published.

In the present paper we develop a method for obtaining such an extension, and compare our results with the work of other authors[3,4] who have attempted alternative stochastic derivations of relativistic wave mechanics.

2. RUDIMENTS OF STOCHASTIC MECHANICS

The stochastic particle of interest is regarded as a classical punctiform mass, occupying at every instant a single point in space and traveling, therefore, along a trajectory. The probabilistic element, which is essential to provide a link to quantum mechanics, is introduced by assuming that this trajectory is continually influenced by a hidden thermostat similar, for example, to those suggested by Bohm[5] and deBroglie[6] in connection with hidden-variables theories. While the exact properties of the thermostat-particle interaction are unknown, the fluctuations of the particle position resulting from this interaction are presumed to be describable as a Markoff process.

Accordingly, the position probability density $\rho(\mathbf{x}, t)$ must obey the Smoluchowski equation

$$\rho(\mathbf{x}, t+\Delta t) = \int P(\mathbf{x}-\Delta\mathbf{x}, t \mid \Delta\mathbf{x}, \Delta t)\rho(\mathbf{x}-\Delta\mathbf{x}, t)d^3(\Delta x), \tag{1}$$

where $P(\mathbf{x}-\Delta\mathbf{x}, t \mid \Delta x, \Delta t)$ is the conditional probability density that a particle at position $\mathbf{x}-\Delta\mathbf{x}$ at time t will be displaced by $\Delta\mathbf{x}$ during the interval Δt, thus reaching position $\mathbf{x}$ at time t.

Similarly, let $F(\mathbf{x}+\Delta\mathbf{x}, t \mid \Delta\mathbf{x}, \Delta t)$ denote the probability density that a particle with position $\mathbf{x}+\Delta\mathbf{x}$ at time t has been displaced through $\Delta\mathbf{x}$ in the preceding interval Δt and thus would have been found at $\mathbf{x}$ at the earlier instant $t-\Delta t$. The analog of Smoluchowski's equation with F instead of P is therefore

$$\rho(\mathbf{x}, t-\Delta t) = \int F(\mathbf{x}+\Delta\mathbf{x}, t \mid \Delta\mathbf{x}, \Delta t)\rho(\mathbf{x}+\Delta\mathbf{x}, t)d^3\Delta x. \tag{2}$$

Since the motion of the particle is to be conceived as a stochastic process, it follows, strictly speaking, that $\mathbf{x}(t)$ is not differentiable. To meet this difficulty, Nelson suggested two possible alternatives for the time derivative of position, the mean forward derivative $D\mathbf{x}(t)$ and the mean backward derivative $D^*\mathbf{x}(t)$, defined respectively as

$$D\mathbf{x}(t) \equiv \lim_{\Delta t\to 0}\frac{\langle \mathbf{x}(t+\Delta t)-\mathbf{x}(t)\rangle}{\Delta t} \equiv \lim_{\Delta t\to 0}\int\left(\frac{\mathbf{x}(t+\Delta t)-\mathbf{x}(t)}{\Delta t}\right)P(\mathbf{x}, t \mid \Delta\mathbf{x}, \Delta t)d^3(\Delta x), \tag{3}$$

$$D^*\mathbf{x}(t) \equiv \lim_{\Delta t\to 0}\frac{\langle \mathbf{x}(t)-\mathbf{x}(t-\Delta t)\rangle^*}{\Delta t} \equiv \lim_{\Delta t\to 0}\int\left(\frac{\mathbf{x}(t)-\mathbf{x}(t-\Delta t)}{\Delta t}\right)F(\mathbf{x}, t \mid \Delta\mathbf{x}, \Delta t)d^3(\Delta x). \tag{4}$$

By hypothesis the motion $\mathbf{x}(t)$ is regarded as being separable into two parts, an ordinary functional part and a Wiener process $\omega(t)$, i.e.,

$$d\mathbf{x}(t) = \mathbf{b}[\mathbf{x}(t), t]\,dt + d\omega(t), \tag{5}$$

where

$$\langle d\mathbf{x}\rangle = \mathbf{b}\,dt, \tag{6}$$

$$\langle (d\mathbf{x})^2\rangle = \langle (d\omega)^2\rangle + O(dt^2), \tag{7}$$

or,

$$\langle (d\mathbf{x})^2\rangle \approx 2\nu\,dt, \tag{8}$$

where ν is, by definition, the diffusion constant.

The stochastic derivative of a function $f(\mathbf{x})$ is given by means of Ito's rule[7]:

$$Df(\mathbf{x}) \equiv \left(\frac{\partial}{\partial t} + \mathbf{b}\cdot\nabla + \nu\nabla^2\right)f(\mathbf{x}). \tag{9}$$

The value of ν to be selected in any particular realization of stochastic mechanics depends upon the nature of the thermostat. In order to match the predictions of nonrelativistic quantum mechanics, Nelson chose $\nu = 3\hbar/2m$.

For the backward case, there are relations analogous

to Eqs. (5)–(8), viz.,

$$\langle d\mathbf{x}\rangle^* = \mathbf{b}^*\, dt, \tag{10}$$

$$\langle (dx)^2\rangle^* \approx \nu\, dt = (3\hbar/2m)\, dt, \tag{11}$$

$$D^* f(\mathbf{x}) = \left(\frac{\partial}{\partial t} + \mathbf{b}\cdot\nabla - \nu\nabla^2\right) f(\mathbf{x}). \tag{12}$$

By using the mathematical properties of $\mathbf{x}(t)$ described in Eqs. (5)–(12), it is now possible to generate Taylor expansions of the integrands in the Smoluchowski equations (1) and (2) and thereby extract the normal Fokker–Planck equation

$$\frac{\partial\rho}{\partial t} = -\nabla\cdot(\rho\mathbf{b}) + \frac{\hbar}{2m}\nabla^2\rho \tag{13}$$

and its backward counterpart

$$\frac{\partial\rho}{\partial t} = -\nabla\cdot(\rho\mathbf{b}^*) - \frac{\hbar}{2m}\nabla^2\rho. \tag{14}$$

Adding (13) and (14) then yields the continuity equation

$$\frac{\partial\rho}{\partial t} = -\nabla\cdot(\rho\mathbf{v}), \tag{15}$$

where $\mathbf{v}$ denotes the average of $\mathbf{b}$ and $\mathbf{b}^*$. The quantities $\mathbf{b}$, $\mathbf{b}^*$, and $\mathbf{v}$ are called, respectively, forward, backward, and total drift velocity.

3. FORMULATION OF THE RELATIVISTIC CASE

To construct a specific quantal wave equation from the elements of the general stochastic mechanics just reviewed requires the adoption of some specific dynamical rule. Thus to obtain the Schrödinger equation, Nelson assumed the Newtonian rule $\mathbf{F} = m\mathbf{a}$, the acceleration $\mathbf{a}$ being defined as follows:

$$\mathbf{a}(t) = \tfrac{1}{2}(DD^* + D^*D)\mathbf{x}(t) = \lim_{\Delta t\to 0}\frac{1}{2}\left(\frac{\langle \mathbf{b}(t) - \mathbf{b}(t-\Delta t)\rangle^*}{\Delta t} + \frac{\langle \mathbf{b}^*(t+\Delta t) - \mathbf{b}^*(t)\rangle}{\Delta t}\right). \tag{16}$$

Kracklauer[8] has criticized this definition of $\mathbf{a}$ on the ground that Nelson failed to provide an adequate physical rationale for it. However, careful scrutiny[9] of the operational meaning of an acceleration measurement does indeed produce the desired rationalization and rebut Kracklauer's objection to Nelson's theory.

By combining this stochastic version of Newton's second law with the continuity equation, Nelson derived two real equations which are equivalent to the complex Schrödinger equation.

At first glance one might conjecture that the relativistic wave equations of quantum mechanics should be similarly derivable simply by substituting for the $\mathbf{F} = m\mathbf{a}$ of Nelson's theory the analogous rule in classical relativistic mechanics. There are, however, certain unexpected difficulties.

First, Hakim[10] has shown that, if the limit $\Delta t \to 0$ is used to calculate conditional probability densities like P and F, the only value for the diffusion constant ν compatible with relativistic invariance is zero. To circumvent this difficulty, we shall discretize the time variable in the stochastic description, so that in the sequences of events which define trajectories, adjacent events have a nonzero minimum temporal separation τ. Obviously the value of this time period must be sufficiently small to ensure that, in any "practical" reference frame, there can be no conflict between the predictions of the stochastic theory and actually realizable macroscopic observations. A specific value for τ will be given later when we develop the model in detail.

A second difficulty in extending Nelson's work arises in the fact that the nonrelativistic formulation assumes that the particle could traverse even an infinite distance in a finite period of time, since both $P(\mathbf{x}, t\,|\,\Delta\mathbf{x}, t)d^3(\Delta x)$ and $F(\mathbf{x}, t\,|\,\Delta\mathbf{x}, \Delta t)d^3(\Delta x)$ are nonzero for the free particle no matter how large $|\Delta\mathbf{x}|$ becomes. To eliminate these spacelike trajectories from the theory, we propose to restructure the stochastic development around a postulate inspired by a curious feature of Dirac's relativistic theory of the electron.[11] According to the latter, the operator for the magnitude of the instantaneous velocity of a free electron possesses just one eigenvalue, the speed c of light in vacuo. Thus if Dirac's theory is correct, the only possible result of such a speed measurement on a free electron would indeed be c. Dirac himself rationalized this proposition through an ambiguous discussion of the Heisenberg uncertainty principle. However, we simply adopt it as an axiom of the stochastic model, i.e., we postulate that c is the instantaneous speed of the relativistic stochastic particle between interactions with the hidden thermostat. Of course the particle cannot travel too far at the velocity of light, for such behavior would surely be incompatible with known facts. In the Dirac theory, the particle executes the physically unexplained Zitterbewegung. In our stochastic model such an oscillation will also appear, but with an evident physical cause, viz., the interactions with the thermostat. That is, the particle travels for a short period of time at the speed of light, then interacts with the thermostat, instantaneously changing direction but not speed.

To summarize, our stochastic derivation of quantum dynamics is to be founded upon Nelson's postulates supplemented by two special axioms: (i) the discretization of time in the stochastic model, and (ii) the attribution of the speed of light to the stochastic particle between interactions with the thermostat.

Assumption (i) must now be made more explicit. We begin by recalling (8) and Nelson's choice of $\nu = 3\hbar/2m$, which leads in the nonrelativistic case to the relation

$$\lim_{\Delta t\to 0}\frac{1}{2}\frac{\langle (d\mathbf{x})^2\rangle}{\Delta t} = \frac{3\hbar}{2m}. \tag{17}$$

In order to avoid the difficulty mentioned by Hakim, (17) must be changed in the relativistic description to

$$\lim_{\Delta t\to\tau}\frac{1}{2}\frac{\langle (d\mathbf{x})^2\rangle}{\Delta t} = \frac{3\hbar}{2m} \tag{18}$$

or,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\frac{(\Delta\mathbf{x}_i)^2}{\tau} = \frac{3\hbar}{2m}, \tag{19}$$

where $\Delta\mathbf{x}_i$ represents the ith sample displacement in an ensemble of n free excursions of the particle between

interactions with the thermostat. To determine the value of the free travel time τ, we note that assumption (ii) implies

$$|\Delta \mathbf{x}_i|/\tau = c, \quad \forall\, \Delta \mathbf{x}_i \tag{20}$$

which, combined with (19), yields

$$\tfrac{1}{2}c^2\tau = 3\hbar/2m \tag{21}$$

or,

$$\tau = 3\hbar/mc^2. \tag{22}$$

Interestingly, this free flight time τ is exactly $3/\pi$ times the Zitterbewegung period of Dirac's electron theory; it differs by a factor of $3/2\pi$ from deBroglie's intrinsic vibration period for a quantum particle.[12] It is important to note that τ is a fixed quantity, not an average; thus in each interval τ, the particle travels exactly the free path length

$$\lambda \equiv c\tau = 3\hbar/mc. \tag{23}$$

Hence the behavior of the relativistic stochastic particle characterized by assumptions (i) and (ii) is in fact a physical realization of the mathematical random walk concept.

Some interesting speculation can be carried out by examining the free path length of various particles. For electrons, $\lambda \approx 10^{-10}$ cm; and for protons, $\lambda \approx 10^{-14}$ cm. If we now compare these lengths to the effective range of the strong nuclear force ($\sim 10^{-13}$ cm), we see that an electron could participate in the nuclear interaction only if its behavior violated our random walk model, whereas the proton could be influenced by the nuclear force without contradicting our assumptions. Thus the proton could change direction at free path endpoints in response to the nuclear interaction as well as to the hidden thermostat. It is of course an experimental fact that the proton, and not the electron, is affected by the strong nuclear interaction.

Finally, it is necessary to specify the Lorentz frame of reference relative to which the stochastic particle executes the proposed random walk, with characteristic values for period τ and length λ given by (22) and (23). We shall take this frame to be the drift rest frame defined precisely in the next section. Physically, the velocity of the drift rest frame relative to an observer is the velocity v the observer would assign to the stochastic particle on the basis of ordinary measurement, in contradistinction to the empirically unmeasurable instantaneous velocity (of light) actually possessed by the stochastic particle in each of its free excursions. To be assured that the assumed value for τ is small enough so that its existence would not contradict known facts, it suffices to note that even for an electron ($\tau \approx 10^{-21}$ sec) that had been accelerated to $v = 0.999c$, the time-dilated interval between interactions with the hidden thermostat would still be only

$$\Delta t = \frac{\tau}{(1 - v^2/c^2)^{1/2}} \approx 10^{-19} \text{ sec}. \tag{24}$$

4. DERIVATION OF RELATIVISTIC QUANTUM WAVE EQUATIONS

It will now be demonstrated that the proposed relativistic extension of stochastic mechanics leads to two real equations which are equivalent to the complex quantum equation for the spinless relativistic particle, the Klein–Gordon equation. Moreover, since each spin component of the Dirac equation satisfies the Klein–Gordon equation, this approach could be considered as explaining the motion of half-integral-spin particles when the spin can be neglected.

In Nelson's derivation of the Schrödinger equation from nonrelativistic stochastic mechanics, conditional probability distributions P and F had to give probability estimates over all space for the future and past locations of the particle. However, in the relativistic random walk particle, given that the particle is at $\mathbf{x}$ at time t, the only possible positions for it at time $t-\tau$ and $t+\tau$ are those on the sphere of radius λ surrounding $\mathbf{x}$. Thus we only need to define the probabilities $P_R(\Omega)\,d\Omega$ and $F_R(\Omega)\,d\Omega$ that the particle will go to or came from the area of unit solid angle Ω to $\Omega + d\Omega$. In general, P_R and F_R will be dependent on the past history of the particle but it will be assumed that we may describe the particle motion by a Markoff process so that we can write $P_R(\Omega) = P_R(\mathbf{x}, t\,|\,\hat{n}(\Omega), \tau)$ and $F_R(\Omega) = F_R(\mathbf{x}, t\,|\,\hat{n}(\Omega), \tau)$, where $\hat{n}(\Omega)$ is a unit vector and $P_R(\mathbf{x}, t\,|\,\hat{n}(\Omega), \tau)$ describes the probability density that the particle located at x at time t will be found at time $t+\tau$ in the position $\mathbf{x} + \Delta\mathbf{x} = \mathbf{x} + \lambda\hat{n}(\Omega)$. A similar definition holds for $F_R(\mathbf{x}, t\,|\,\hat{n}(\Omega), \tau)$. In analogy to the nonrelativistic case, we can define a relativistic forward drift 4-velocity $b_\mu = (b_0, b_i)$, where

$$b_i = b^i \equiv \frac{\lambda}{\tau}\int n_i(\Omega) P_R(\mathbf{x}, t\,|\,\hat{n}(\Omega), \tau)\,d\Omega$$

$$= c\langle n_i\rangle, \tag{25}$$

$$b_0 = -\,b^0 \equiv -\,c(\tau/\tau) = -\,c. \tag{26}$$

Similarly, we may introduce a backward drift 4-velocity $b^*_\mu = (b^*_0, b^*_i)$, where

$$b^*_i \equiv \frac{\lambda}{\tau}\int n_i F_R(\mathbf{x}, t\,|\,\hat{n}(\Omega), \tau)\,d\Omega$$

$$= c\langle n_i\rangle^*, \tag{27}$$

$$b^*_0 \equiv -\,c(\tau/\tau) = -\,c. \tag{28}$$

As in the nonrelativistic situation, the exact location of the particle at any instant in time will be presumed not known. However, instead of specifying just the position probability density, which is not a relativistic invariant, we shall need a probability current density 4-vector j_μ, where j_0 is $-c$ multiplied by the probability density. The definition of the other components will be given later.

We now consider the problem of the "rest frame," which at first seems a perplexing situation, since the particle, continually moving with the speed of light, has no admissible rest frame in the usual sense. However, a useful definition may be given as follows: The *drift rest frame* is that Lorentz frame in which the one nonvanishing component of j_μ is $j_0 = -c\rho$. In this frame the stochastic particle will obey the Smoluchowski equation

$$\rho(\mathbf{x}, t+\tau) = \int \rho(\mathbf{x} - \lambda\hat{n}, t) P_R(\mathbf{x} - \lambda\hat{n}, t\,|\,\hat{n}(\Omega), \tau)\,d\Omega. \tag{29}$$

If we expand the left-hand side in a Taylor series about t, and the right-hand side in a Taylor series about x, we get

$$\rho(\mathbf{x}, t+\tau) = \rho(\mathbf{x}, t) + \tau \frac{\partial}{\partial t}\rho(\mathbf{x}, t) + \frac{\tau^2}{2}\frac{\partial^2}{\partial t^2}\rho(\mathbf{x}, t) \tag{30}$$

and

$$\begin{aligned}
&\rho(\mathbf{x} - \lambda\hat{n}, t)P_R(\mathbf{x} - \lambda\hat{n}, t\,|\,\hat{n}, \tau) \\
&\quad = \rho(\mathbf{x}, t)P_R(\mathbf{x}, t\,|\,\hat{n}, \tau) - \lambda P_R(\mathbf{x}, t\,|\,\hat{n}, \tau)(\hat{n}\cdot\nabla\rho(\mathbf{x}, t)) \\
&\quad\quad - \lambda[\hat{n}\cdot\nabla P_R(\mathbf{x}, t\,|\,\hat{n}, \lambda)]\rho(\mathbf{x}, t) + \frac{\lambda^2}{2}\sum_{i=1}^{3}\left[\left(n_i^2 P_R(\mathbf{x}, t\,|\,\hat{n}, \tau)\right.\right. \\
&\quad\quad \times\frac{\partial^2}{\partial x_i^2}\rho(\mathbf{x}, t)\Big) + \left(n_i^2\rho(\mathbf{x}, t)\frac{\partial^2}{\partial x_i^2}P_R(\mathbf{x}, t\,|\,\hat{n}, \tau)\right) \\
&\quad\quad \left.\left. + 2\left(n_i^2\frac{\partial}{\partial x_i}P_R(\mathbf{x}, t\,|\,\hat{n}, \tau)\frac{\partial}{\partial x_i}\rho(\mathbf{x}, t)\right)\right] + O(\lambda^3).
\end{aligned} \tag{31}$$

Using the relations

$$\int P_R(\mathbf{x}, t\,|\,\hat{n}, \tau)\, d\Omega = 1, \tag{32}$$

$$\int \lambda\hat{n}\frac{\partial}{\partial x}[P_R(\hat{x}, t\,|\,\hat{n}, \tau)]d\Omega = \lambda\frac{\partial}{\partial x}\langle\hat{n}\rangle = \tau\frac{\partial}{\partial x}\mathbf{b}, \tag{33}$$

$$\int \lambda^2\left(\sum_{i=1}^{3} n_i^2\right)P_R\, d\Omega = \int \lambda^2 P_R\, d\Omega = \lambda^2, \tag{34}$$

$$\int n_i^2 P_R\, d\Omega = \langle n_i^2\rangle = \tfrac{1}{3}\langle\hat{n}^2\rangle = \tfrac{1}{3}, \tag{35}$$

where the last equation ensures space isotropy for the hidden thermostat, we can now integrate (30) and (31) term by term to obtain (approximately)

$$\frac{\partial\rho}{\partial t} + \frac{\tau}{2}\frac{\partial^2}{\partial t^2}\rho = -\nabla\cdot(\rho\mathbf{b}) + \frac{\lambda^2}{6\tau}\nabla^2\rho. \tag{36}$$

If we substitute the values for τ and λ from (22) and (23) and make use of the 4-gradient $\partial^\mu \equiv (-\partial/\partial(ct), \nabla)$, (36) may be expressed in relativistic notation as

$$\partial^\mu(\rho b_\mu) - (\hbar/2m)\Box^2\rho = 0. \tag{37}$$

In the drift rest frame, the equivalent backward version of Smoluchowski's equation will also hold. Hence,

$$\rho(\mathbf{x}, t-\tau) = \int \rho(\mathbf{x} + \lambda\hat{n}, t)F(\mathbf{x} + \lambda\hat{n}, t\,|\,\hat{n}, \tau)\, d\Omega. \tag{38}$$

As analysis parallel to that given above for the forward diffusion situation yields the backward analog of (37),

$$\partial^\mu(\rho b_\mu^*) + (\hbar/2m)\Box^2\rho = 0. \tag{39}$$

If we define

$$j_b^\mu \equiv \rho b^\mu \tag{40}$$

and

$$j_{b*}^\mu \equiv \rho(b^*)^\mu \tag{41}$$

we can construct the total drift current

$$j^\mu \equiv \tfrac{1}{2}(j_b^\mu + j_{b*}^\mu) \tag{42}$$

and a drift velocity as

$$v^\mu \equiv j^\mu/\rho. \tag{43}$$

Then j_μ satisfies the continuity equation

$$\partial^\mu j_\mu = 0 \tag{44}$$

as can be seen by adding (37) and (39).

Equations (37) and (39) were derived on the assumption of a special Lorentz frame (the drift rest frame) and, as written, are not covariant with respect to Lorentz transformations since ρ is not a world scalar. However, we can define the term

$$(-1/c^2)j_\lambda j^\lambda \equiv |\rho|^2 \tag{45}$$

which is a world scalar, and generalize (37) and (39) to the covariant forms

$$\partial^\mu(j_b)_\mu - (\hbar/2m)\Box^2|\rho| = 0, \tag{46}$$

$$\partial^\mu(j_{b^*})_\mu + (\hbar/2m)\Box^2|\rho| = 0. \tag{47}$$

When $j_\mu = (-c\rho, 0, 0, 0)$, these reduce to the previous forms (37) and (39).

In order to construct the Klein—Gordon equation, it is necessary to adopt a definition for 4-acceleration, a_μ. Nelson's three-dimensional formula was

$$a(\mathbf{t}) = \tfrac{1}{2}\lim_{\Delta t\to 0}\left(\frac{\langle\mathbf{b}(t) - \mathbf{b}(t-\Delta t)\rangle^*}{\Delta t} + \frac{\langle\mathbf{b}^*(t+\Delta t) - \mathbf{b}^*(t)\rangle}{\Delta t}\right). \tag{48}$$

To extend this to four dimensions, we define a as follows:

$$\begin{aligned}
a_\mu &\equiv \frac{1}{2}\left(\frac{\langle b_\mu(t) - b_\mu(t-\tau)\rangle^*}{\tau} + \frac{\langle b_\mu^*(t+\tau) - b_\mu^*(\tau)\rangle}{\tau}\right) \\
&= \tfrac{1}{2}(b_\lambda^*\partial^\lambda)b_\mu + \tfrac{1}{2}(b_\lambda\partial^\lambda)b_\mu^* + \frac{1}{2}\frac{\hbar}{2m}\Box^2(b_\mu - b_\mu^*),
\end{aligned} \tag{49}$$

where Ito's rule for differentiation of forward and backward stochastic processes has been utilized as in the three-dimensional case. If the substitutions,

$$v_\mu = \tfrac{1}{2}(b_\mu + b_\mu^*), \tag{50}$$

$$\nu_\mu \equiv \tfrac{1}{2}(b_\mu - b_\mu^*) = (\hbar/2m)\partial_\mu \ln|\rho|, \tag{51}$$

are made, (49) can be written in the more tractable form

$$a_\mu = v_\lambda\partial^\lambda v_\mu - \nu_\lambda\partial^\lambda\nu_\mu - (\hbar/2m)\Box^2\nu_\mu. \tag{52}$$

For the fundamental dynamical rule, we adopt Einstein's relativistic version of Newton's law,

$$F_\mu = ma_\mu, \tag{53}$$

where F_μ is the force 4-vector. In particular, we shall consider only the Lorentz force F^μ associated with electromagnetic field $F^{\mu\nu}$, i.e.,

$$F^\mu = (e/c)F^{\mu\lambda}v_\lambda = (e/c)[\partial^\mu A^\lambda - \partial^\lambda A^\mu]v_\lambda, \tag{54}$$

where A^μ is the electromagnetic potential. The Lorentz gauge will be used so that

$$\partial_\mu A^\mu = 0. \tag{55}$$

The generalized momentum will be assumed to be derivable from the 4-gradient of the world scalar S, Hamilton's principal function of ordinary relativistic classical mechanics. Thus

$$\partial^\mu S = mv^\mu + (e/c)A^\mu. \tag{56}$$

If (52), (54), and (56) are substituted into (53), the result is

$$\left[\frac{1}{m}\left(\partial^\lambda S-\frac{e}{c}A^\lambda\right)\partial_\lambda\left(\partial^\mu S-\frac{e}{c}A^\mu\right)-m\nu^\lambda\partial_\lambda\nu^\mu-\frac{\hbar}{2}\Box^2\nu^\mu\right]$$
$$=\frac{e}{c}\left[\partial^\mu A^\lambda-\partial^\lambda A^\mu\right]\left(\partial_\lambda S-\frac{e}{c}A_\lambda\right). \tag{57}$$

If use is made of (55), this reduces to

$$\frac{1}{m}(\partial^\lambda S)\partial_\lambda(\partial^\mu S)-\frac{1}{m}\partial^\mu\left(\frac{e}{c}A^\lambda\partial_\lambda S\right)+\frac{1}{2m}\partial^\mu\left[\left(\frac{e}{c}\right)^2A_\lambda A^\lambda\right]$$
$$=m\nu^\lambda\partial_\lambda v^\mu-(\hbar/2)\Box^2\nu^\mu. \tag{58}$$

Since both $\partial^\mu S$ and $\nu^\mu=(\hbar/2m)\partial^\mu\ln|\rho|$ are gradients of scalars, (58) may be written as

$$\partial^\mu\left[\frac{1}{2m}(\partial_\lambda S)(\partial^\lambda S)-\frac{e}{mc}A_\lambda\partial^\lambda S+\frac{1}{2m}\left(\frac{e}{c}\right)^2A_\lambda A^\lambda\right.$$
$$\left.-\frac{m}{2}\nu_\lambda\nu^\lambda-\frac{\hbar^2}{4m}\Box^2\ln|\rho|\right]=0, \tag{59}$$

or

$$\frac{1}{m}\left(\partial_\lambda S-\frac{e}{c}A_\lambda\right)^2+m(\nu_\lambda)^2-\frac{\hbar^2}{2m}\Box^2\ln|\rho|=M, \tag{60}$$

where M is a constant. To determine the value of M, consider the classical limit case ($\hbar$ negligible) with vanishing vector potential A^μ. Then (60) becomes

$$(1/m)(\partial_\lambda S)^2=M \tag{61}$$

or

$$mv_\lambda v^\lambda=M. \tag{62}$$

Since in the drift rest frame

$$v_\lambda=(-c,0,0,0) \tag{63}$$

it follows that

$$M=-mc^2. \tag{64}$$

Thus the two basic equations of relativistic stochastic mechanics are (44) and (60). Both of these equations are of manifestly covariant form; all of the terms in them are either world scalars or, with the exception of j^μ and A^μ, the 4-gradients of world scalars. Moreover, since both j^μ and A^μ, when contracted with covariant 4-vector operator ∂^μ, form the world scalar zero, they must be covariant 4-vectors. Therefore the stochastic model satisfies all the necessary requirements of invariance with respect to Lorentz transformations.

To compare the stochastic relativistic model with orthodox quantum formalism, we recall the Klein–Gordon equation

$$\left(\partial^\mu-\frac{e}{c}iA^\mu\right)^2\psi-\left(\frac{m^2c^2}{\hbar^2}\right)\psi=0. \tag{65}$$

Separating (65) into its real and imaginary parts, we obtain

$$\left(\partial^\mu S-\frac{e}{c}A^\mu\right)^2-\frac{\hbar^2}{R}\Box^2R=-m^2c^2 \tag{66}$$

and

$$\partial^\mu\left[R^2\left(\partial_\mu S'-\frac{e}{c}A_\mu\right)\right]=0, \tag{67}$$

where R and S' are real functions which determine ψ through

$$\psi=R\exp(iS'/\hbar). \tag{68}$$

The term $\Box^2R/R^2$ obeys the vector identities

$$\frac{\Box^2R}{R}=\frac{\Box^2R^2}{2R^2}-\tfrac{1}{2}(\partial^\mu\ln R^2)^2 \tag{69}$$

and

$$\Box^2(\ln R^2)=(1/R^2)\Box^2R^2-(\partial^\mu\ln R^2)^2. \tag{70}$$

By substituting (69) and (70) into (66) we get

$$\frac{1}{m}\left(\partial^\mu S'-\frac{e}{c}A^\mu\right)^2-m\left(\frac{\hbar}{2m}\partial^\mu\ln R^2\right)^2-\frac{\hbar^2}{2m}\Box^2\ln R^2=-mc^2. \tag{71}$$

Now, if the identifications, $R^2=|\rho|$ and $S=S'$ are made, (67) and (71) become identical to (44) and (60), respectively. These stochastic results are not totally equivalent to the standard quantum formalism, since the Klein–Gordon equation allows both positive and negative energy eigenvalues. By contrast, in the stochastic model we have

$$E/c=p^0=mv^0\geq mc>0, \tag{72}$$

and, thus, the stochastic derivation has been only for positive energies. However, negative energies can be easily incorporated into the theory if one redefines b_0 (and b_0^*) to be

$$b_0=-c[A(\mathbf{x},t)-B(\mathbf{x},t)], \tag{73}$$

where $A(\mathbf{x},t)$ is the probability that the particle at $(\mathbf{x},t)$ will have positive energy during the time span τ, and $B(\mathbf{x},t)$ is the probability that it will have negative energy. A similar redefinition holds for b_0^* and hence for $v_0=\frac{1}{2}(b_0+b_0^*)$. Alternatively, one may adopt the interpretation that A is the probability the particle will go forward in time and B is the probability that it will go backward in time. Our derivation has thus far assumed $A=1$ and $B=0$. For a negative energy particle, the reverse would be the case.

Aron[3] and de la Pena-Auerbach[4] have also constructed derivations of the Klein–Gordon equation from stochastic concepts. Aron's approach yields two real equations which are not equivalent to the complex Klein–Gordon equation except in special cases.

De la Pena-Auerbach, by studying the properties of stochastic derivatives under time reversal, has arrived by a quite different theoretical route at equations essentially equivalent to our two basic relativistic stochastic equations. Surprisingly, he did not require the particle to have a fixed free speed c but did assume that

$$\frac{1}{2}\frac{\langle(\Delta x)^2\rangle}{\Delta t}=\frac{3\hbar}{2m}, \tag{74}$$

where $\Delta t\approx\tau$.

However, it is easily demonstrable that (74) must be reduced to the more stringent requirement of our random-walk model in order to avoid inconsistency. Indeed if in a particular excursion, we had $|\Delta\mathbf{x}|>c\tau$, the world line of the particle would be spacelike and hence

inadmissible. If, however, we always require $|\Delta\mathbf{x}| \leq c\tau$, then necessarily

$$\frac{1}{2}\frac{\langle(\Delta x)^2\rangle}{\tau} \leq \frac{1}{2}\frac{c^2\tau^2}{\tau} = \frac{3\hbar}{2m}. \tag{75}$$

The equality sign holds only if $|\Delta\mathbf{x}| = c\tau$, for all $\Delta\mathbf{x}$, which corresponds to our stipulation (20).

[1]E. Nelson, Phys. Rev. **150**, 1079 (1966).

[2]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U. P., Princeton, N.J., 1967), p. 141.

[3]J. Aron, Prog. Theor. Phys. **35**, 147 (1966).

[4]L. de la Pena-Auerbach, Rev. Mex. Fis. **19**, 133 (1970).

[5]D. Bohm, Phys.Rev. **166**, (1952).

[6]L. DeBroglie, Ann. Inst. Henri Poincaré **1**, 1 (1964).

[7]K. Ito, *Lectures on Stochastic Processes* (Tata, Bombay, 1961), p. 187.

[8]A. Kracklauer, Phys. Rev. D **10**, 1358 (1974).

[9]W. Lehr, thesis (Washington State University, 1976).

[10]R. Hakim, J. Math. Phys. **9**, 1805 (1968).

[11]P.A.M. Dirac, *The Principles of Quantum Mechanics* (Oxford U. P., Oxford, 1958), 4th ed.

[12]L. deBroglie, *Non-Linear Wave Mechanics* (Elsevier, Amsterdam, 1960), p. 4.

Variational methods for chemical and nuclear reactions*

Oakley H. Crawford

Chemistry Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830
(Received 23 August 1976)

All the variational functionals are derived which satisfy certain criteria of suitability for molecular and nuclear scattering, below the threshold energy for three-body breakup. The existence and uniqueness of solutions are proven. The most general suitable functional is specialized, by particular values of its parameters, to Kohn's $\tan\eta$, Kato's $\cot(\eta-\theta)$, the inverse Kohn $\cot\eta$, Kohn's S matrix, our S matrix, Lane and Robson's functional, and several new functionals, an infinite number of which are contained in the general expression. Four general ways of deriving algebraic methods from a given functional are discussed, and illustrated with specific algebraic results. These include equations of Lane and Robson and of Kohn, the fundamental R matrix relation, and new equations. The relative configuration space is divided as in the Wigner R matrix theory, and trial wavefunctions are needed for only the region where all the particles are interacting. In addition, a version of the general functional is presented which does not require any division of space.

I. INTRODUCTION

Many variational methods have been proposed for scattering theory, as discussed in current reviews,[1-4] and a number of them have recently been employed in studies of a chemical reaction,[5,6] electron scattering from complex atoms,[7-16] and nuclear scattering.[17-21] Such methods are valuable, as they provide a means of calculating scattering matrices whose errors are of second order with respect to the errors in the approximate wavefunctions which go into the computation. Certain methods (which unfortunately require the exact solution of a Schrödinger equation in all open channels) produce a variational result whose error, besides being of second order, is of a known sign.[22-26] This is extremely useful. For one thing, it allows one immediately to choose the best of several results, and, thus, to systematically improve the calculation. However, none of the other known procedures is believed to produce a variational bound. One may hope that the best method remains to be found.

Much work has gone into finding various limited minimum principles[27] and a convergence theorem[28] for Kohn's principle[29] for the tangent of the phase shift and its multichannel extension, and studying the false resonances which often appear.[4,30-32] One would like to see similar results for all variational methods. Unfortunately, the existing methods have not been put into a unified enough formalism that convergence (or nonconvergence) theorems can be applied to them as a class. Furthermore, the known variational methods by no means exhaust the possibilities. To rectify this situation, by putting a broad class of variational methods into a single formal framework, is the purpose of this work.

The heart of any variational method is a functional whose stationary value is to be approximated. It is assumed in this work that functionals must satisfy certain criteria (such as being bilinear in trial functions), in order to be suitable for reactions and other multichannel processes. Working from these criteria, an explicit general expression for *all* suitable variational functionals is derived. It is seen that the suitable functionals form a continuous set, containing the previously known ones and an infinite number of new ones.

Several ways of using the general suitable functional in an algebraic variational calculation are examined in four general classes.

II. THE GENERAL FUNCTIONAL

Consider a collision of two nuclei, or molecules,[33] at an energy below the threshold for three-body breakup. Assume the applicability of nonrelativistic quantum mechanics and the absence or unimportance of all processes of creation or annihilation. Further assume that, for any pair of nuclei α, there exists a finite distance of separation a_α, beyond which any potential acting between the pair α is a function of only the distance r_α between the pair.

In the configuration space of relative coordinates of the nucleons, there exists a set of surfaces S_α defined by $r_\alpha = a_\alpha$. From these surfaces can be constructed a closed surface S. Inside S is the region in which all the nucleons are close together. Outside of S are certain regions, called arrangement channels, in which the nucleons are grouped in one or another of the possible pairs α. The remaining space corresponds to configurations which occur with negligible probability, by assumption.

The above assumptions and the partitioning of space have been discussed in detail by Lane and Thomas.[34]

The wavefunction on and outside of S may now be written as a sum over (two-body) channels c

$$|\psi\rangle = \sum_c |c) q_c(r_c), \quad r_c \geq a_c, \tag{1}$$

in which q_c is a function only of r_c and $|c)$ is a normalized "channel spin wavefunction," which is independent of r_c. The index c stands for α and all the quantum numbers necessary to describe the nuclear states and the relative angular motion in the channel.

Following Bloch,[35] the notation $|\ \rangle$ will be reserved for functions of all the coordinates, while $|\ )$ will denote functions of all the coordinates except r_α. The functions $|c)$ are orthogonal and will be defined to be normalized. Thus,

$$(c|c') = \delta_{cc'}. \tag{2}$$

The above notation implies a scalar product in which the product of the complex conjugate of the first function and the second function is integrated with respect to solid angle ω. The differential of solid angle, $d\omega$, is defined such that the volume element in the α arrangement channel is $d\omega r_\alpha^2 dr_\alpha$. Note that $d\omega$ does not include the r_α^2 factor.

We seek the most general suitable variational functional. So that the functional will be amenable to treatment by matrix methods, it will be bilinear[36] in two trial functions,[37] $\psi_t^{(1)}$ and $\psi_t^{(2)}$, which are defined on and inside the surface S. The functional will contain the Hamiltonian H and the energy E in

$$\langle \psi_t^{(1)} | (H - E)\psi_t^{(2)} \rangle, \tag{3a}$$

and the Hermitian adjoint of the Hamiltonian, in

$$\langle \psi_t^{(1)} | (H^\dagger - E)\psi_t^{(2)} \rangle = \langle (H - E)\psi_t^{(1)} | \psi_t^{(2)} \rangle. \tag{3b}$$

The integration indicated by the angular brackets extends only over the volume inside S (including all singularities on S). Other operators, such as the square of the Hamiltonian, are unnecessary, and we assume that including them in the functional would complicate the expression without conferring a compensating advantage.[38] The surface integrals[39]

$$u_{ct}^{(\mu)} = (c | r_c \psi_t^{(\mu)})_{a_c} \tag{3c}$$

and

$$u_{ct}^{(\mu)\prime} = (c | \mathrm{grad}_n r_c \psi_t^{(\mu)})_{a_c}, \tag{3d}$$

where grad_n stands for the component normal to S (outward) of the gradient, will be needed to make the functional stationary and to relate it to collision matrices. The suffix t stands for "trial."

To be suitable for a variational calculation, the functional must be stationary with respect to all continuous variations of the $\psi_t^{(\mu)}$ if and only if the $\psi_t^{(\mu)}$ satisfy the Schrödinger equation inside S and, perhaps, inhomogenous boundary conditions[40] on S. Finally, we require that the scattering matrix be calculable from the stationary value of the functional (rather than $\psi_t^{(\mu)}$), so that the error in the scattering matrix is of second order in the error in the trial functions.

The most general functional satisfying the criteria of the preceding two paragraphs has the form

$$\begin{aligned} G = &\sum_{rs} [u_{rt}^{(1)*\prime}(P_{1rs} u_{st}^{(2)\prime} + P_{2rs} u_{st}^{(2)} + P_{3r}\delta_{rs}) \\ &+ u_{rt}^{(1)*}(P_{4rs} u_{st}^{(2)\prime} + P_{5rs} u_{st}^{(2)} + P_{6r}\delta_{rs})] \\ &+ \sum_s [P_{7s} u_{st}^{(2)\prime} + P_{8s} u_{st}^{(2)}] + P_9 \\ &- P_{10}\langle \psi_t^{(1)} | (H - E)\psi_t^{(2)} \rangle - P_{11}\langle \psi_t^{(1)} | (H^\dagger - E)\psi_t^{(2)} \rangle, \end{aligned} \tag{4}$$

where the summation extends over all two-body channels, open and closed, the P_i are all constants, and the $u_{ct}^{(\mu)}$ and $u_{ct}^{(\mu)\prime}$ quantities are to be calculated from the trial functions, using Eqs. (3c) and (3d).

To discover what values should be assigned to the P_i, we examine the first-order variation of G, which, in matrix notation, is

$$\begin{aligned} \delta G = &\delta \mathbf{u}^{(1)\dagger\prime}(\mathbf{P}_1 \mathbf{u}_e^{(2)\prime} + [\mathbf{P}_2 + P_{11}(\hbar^2/2\mathbf{m})]\mathbf{u}_e^{(2)} + \mathbf{P}_3) \\ &+ \delta \mathbf{u}^{(1)\dagger}([\mathbf{P}_4 - P_{11}(\hbar^2/2\mathbf{m})]\mathbf{u}_e^{(2)\prime} + \mathbf{P}_5 \mathbf{u}_e^{(2)} + \mathbf{P}_6) \\ &+ (\mathbf{u}_e^{(1)\dagger\prime}\mathbf{P}_1 + \mathbf{u}_e^{(1)\dagger}[\mathbf{P}_4 + \mathbf{P}_{10}(\hbar^2/2\mathbf{m})] + \mathbf{P}_7)\delta \mathbf{u}^{(2)\prime} \\ &+ (\mathbf{u}_e^{(1)\dagger\prime}[\mathbf{P}_2 - P_{10}(\hbar^2/2\mathbf{m})] + \mathbf{u}_e^{(1)\dagger}\mathbf{P}_5 + \mathbf{P}_8)\delta \mathbf{u}^{(2)} \\ &- (P_{10} + P_{11})(\langle \delta\psi^{(1)} | (H - E)\psi_e^{(2)} \rangle \\ &+ \langle (H - E)\psi_e^{(1)} | \delta\psi^{(2)} \rangle), \end{aligned} \tag{5}$$

in which $\mathbf{u}_e^{(\mu)}$ and $\mathbf{u}_e^{(\mu)\prime}$ are column vectors, the dagger represents the transpose of the complex conjugate, and $\mathbf{m}$ is the diagonal matrix of the channel reduced masses. The subscript e denotes an exact wavefunction. In deriving Eq. (5), Green's theorem was used, which gives

$$\begin{aligned} -\langle \psi_e^{(1)} | (H - E)\delta\psi^{(2)} \rangle = &-\langle (H - E)\psi_e^{(1)} | \delta\psi^{(2)} \rangle \\ &+ \mathbf{u}_e^{(1)\dagger}(\hbar^2/2\mathbf{m})\delta \mathbf{u}^{(2)\prime} \\ &- \mathbf{u}_e^{(1)\dagger\prime}(\hbar^2/2\mathbf{m})\delta \mathbf{u}^{(2)}, \end{aligned} \tag{6}$$

which is valid even if $\delta\psi^{(2)}$ contains many-body channel components on S,[41] since the exact wavefunction $\psi_e^{(1)}$ contains, by assumption, only two-body channel components on S. From Eq. (5) emerge the following six equations as necessary and sufficient conditions for G to be stationary with respect to arbitrary infinitesimal continuous variations of $\psi^{(1)}$ and $\psi^{(2)}$:

$$\mathbf{P}_1 \mathbf{u}_e^{(2)\prime} + [\mathbf{P}_2 + P_{11}(\hbar^2/2\mathbf{m})]\mathbf{u}_e^{(2)} + \mathbf{P}_3 = 0, \tag{7a}$$

$$[\mathbf{P}_4 - P_{11}(\hbar^2/2\mathbf{m})]\mathbf{u}_e^{(2)\prime} + \mathbf{P}_5 \mathbf{u}_e^{(2)} + \mathbf{P}_6 = 0, \tag{7b}$$

$$\tilde{\mathbf{P}}_1 \mathbf{u}_e^{(1)*\prime} + [\tilde{\mathbf{P}}_4 + P_{10}(\hbar^2/2\mathbf{m})]\mathbf{u}_e^{(1)*} + \tilde{\mathbf{P}}_7 = 0, \tag{7c}$$

$$[\tilde{\mathbf{P}}_2 - P_{10}(\hbar^2/2\mathbf{m})]\mathbf{u}_e^{(1)*\prime} + \tilde{\mathbf{P}}_5 \mathbf{u}_e^{(1)*} + \tilde{\mathbf{P}}_8 = 0, \tag{7d}$$

$$(P_{10} + P_{11})(H - E)\psi_e^{(1)} = 0, \tag{8a}$$

$$(P_{10} + P_{11})(H - E)\psi_e^{(2)} = 0. \tag{8b}$$

A set of values must be assigned to the constants $\mathbf{P}_i$ in such a way that Eqs. (7) and (8) are appropriate conditions for the exact wavefunctions $\psi_e^{(1)}$ and $\psi_e^{(2)}$. Since it was stipulated above that the Schrödinger equation be among the conditions, it follows from Eq. (8) that $P_{10} + P_{11}$ must be nonzero, but is otherwise arbitrary. For convenience, let

$$P_{10} + P_{11} = 1. \tag{9}$$

Because of the ways in which P_{10} and P_{11} appear in combination with $\mathbf{P}_2$ and $\mathbf{P}_4$ in Eqs. (7), it is easy to see that the eight arrays of coefficients of the $\mathbf{u}_e^{(\mu)\prime}$ and $\mathbf{u}_e^{(\mu)}$ vectors in those equations cannot all be zero. Therefore, Eqs. (7) are boundary conditions[40] on the exact wavefunction.

This concludes the derivation on which the remainder of this work is based.[42] We have the most general suitable functional, Eq. (4), and we see that, in order for it to have its intended properties, values must be assigned to the $\mathbf{P}_i$ parameters in such a way that Eqs. (7) are consistent with the Schrödinger equation. The most straightforward way of doing that is, evidently, to select an appropriate form for the boundary conditions, and determine the $\mathbf{P}_i$ accordingly.

There is some freedom in the form of the boundary conditions. We begin with a fairly general form

$$\mathbf{u}_e^{(1)*\prime} - \gamma^{(1)*}\mathbf{u}_e^{(1)*} = \zeta^{(1)*}, \tag{10a}$$

$$\mathbf{u}_e^{(2)\prime} - \gamma^{(2)}\mathbf{u}_e^{(2)} = \zeta^{(2)}, \tag{10b}$$

where $\gamma^{(\mu)}$ and $\zeta^{(\mu)}$ are arbitrary constant square matrices ($n\times n$) and column vectors ($n\times 1$), respectively (when there are n channels). Other boundary conditions are studied in Appendices A and B.

Equations (10) and the Schrödinger equation together uniquely determine the wavefunctions $\psi_e^{(\mu)}$. (The $\zeta^{(\mu)}$ cannot be zero vectors, or else the solutions exist only at discrete energies, as in R-matrix theory. See Appendix A.) Therefore, Eq. (7a) must be identical to the product of $\mathbf{P}_1$ and Eq. (10b), in order that Eq. (7a) contain no content not found in Eq. (10b). It follows that

$$\mathbf{P}_2 + P_{11}(\hbar^2/2\mathrm{m}) = -\mathbf{P}_1\gamma^{(2)} \tag{11a}$$

and

$$\mathbf{P}_3 = -\mathbf{P}_1\zeta^{(2)}. \tag{11b}$$

Similar reasoning, applied to Eqs. (7b)—(7d), produces

$$\mathbf{P}_5 = -[\mathbf{P}_4 - P_{11}(\hbar^2/2\mathrm{m})]\gamma^{(2)}, \tag{11c}$$

$$\mathbf{P}_6 = -[\mathbf{P}_4 - P_{11}(\hbar^2/2\mathrm{m})]\zeta^{(2)}, \tag{11d}$$

$$\tilde{\mathbf{P}}_4 + P_{10}(\hbar^2/2\mathrm{m}) = -\tilde{\mathbf{P}}_1\gamma^{(1)*}, \tag{11e}$$

$$\tilde{\mathbf{P}}_7 = -\tilde{\mathbf{P}}_1\zeta^{(1)*}, \tag{11f}$$

$$\tilde{\mathbf{P}}_5 = -[\tilde{\mathbf{P}}_2 - P_{10}(\hbar^2/2\mathrm{m})]\gamma^{(1)*}, \tag{11g}$$

$$\tilde{\mathbf{P}}_8 = -[\tilde{\mathbf{P}}_2 - P_{10}(\hbar^2/2\mathrm{m})]\zeta^{(1)*}. \tag{11h}$$

When the two expressions, above, for $\mathbf{P}_5$ are equated, after eliminating $\mathbf{P}_2$ and $\mathbf{P}_4$ with Eqs. (11a) and (11e), one finds that

$$\gamma^{(1)\dagger}(\hbar^2/2\mathrm{m}) = (\hbar^2/2\mathrm{m})\gamma^{(2)}. \tag{12}$$

Therefore, $\gamma^{(1)\dagger}$ and $\gamma^{(2)}$ are both related to the same arbitrary square matrix ϵ,

$$\gamma^{(2)} = (2\mathrm{m}/\hbar^2)\epsilon \tag{13a}$$

and

$$\gamma^{(1)*} = (2\mathrm{m}/\hbar^2)\tilde{\epsilon}. \tag{13b}$$

Letting

$$\mathbf{P}_1 = \xi, \tag{14a}$$

for the other P_i, we find

$$\mathbf{P}_2 = -\xi(2\mathrm{m}/\hbar^2)\epsilon - P_{11}(\hbar^2/2\mathrm{m}), \tag{14b}$$

$$\mathbf{P}_3 = -\xi\zeta^{(2)}, \tag{14c}$$

$$\mathbf{P}_4 = -\epsilon(2\mathrm{m}/\hbar^2)\xi - (\hbar^2/2\mathrm{m})(1 - P_{11}), \tag{14d}$$

$$\mathbf{P}_5 = \epsilon(2\mathrm{m}/\hbar^2)\xi(2\mathrm{m}/\hbar^2)\epsilon + \epsilon, \tag{14e}$$

$$\mathbf{P}_6 = [\epsilon(2\mathrm{m}/\hbar^2)\xi + (\hbar^2/2\mathrm{m})]\xi^{(2)}, \tag{14f}$$

$$\mathbf{P}_7 = -\zeta^{(1)\dagger}\xi, \tag{14g}$$

$$\mathbf{P}_8 = \zeta^{(1)\dagger}[\xi(2\mathrm{m}/\hbar^2)\epsilon + (\hbar^2/2\mathrm{m})]. \tag{14h}$$

For convenience, let P_9, which is merely an additive constant, be

$$P_9 = \zeta^{(1)\dagger}\xi\zeta^{(2)}. \tag{14i}$$

Then, substituting the above P's into Eq. (4) produces the desired functional,

$$\begin{aligned} G = P_{11}\Big[&\mathbf{u}_t^{(1)\dagger}\frac{\hbar^2}{2\mathrm{m}}\zeta^{(2)} \\ &- \Big(\mathbf{u}_t^{(1)\dagger\prime} - \mathbf{u}_t^{(1)\dagger}\epsilon\frac{2\mathrm{m}}{\hbar^2} - \zeta^{(1)\dagger}\Big)\frac{\hbar^2}{2\mathrm{m}}\mathbf{u}_t^{(2)}\Big] \\ &+ (1 - P_{11})\Big[\zeta^{(1)\dagger}\frac{\hbar^2}{2\mathrm{m}}\mathbf{u}_t^{(2)} - \mathbf{u}_t^{(1)\dagger}\frac{\hbar^2}{2\mathrm{m}}\Big(\mathbf{u}_t^{(2)\prime} - \frac{2\mathrm{m}}{\hbar^2}\epsilon\mathbf{u}_t^{(2)} - \zeta^{(2)}\Big)\Big] \\ &+ \Big(\mathbf{u}_t^{(1)\dagger\prime} - \mathbf{u}_t^{(1)\dagger}\epsilon\frac{2\mathrm{m}}{\hbar^2} - \zeta^{(1)\dagger}\Big)\xi\Big(\mathbf{u}_t^{(2)\prime} - \frac{2\mathrm{m}}{\hbar^2}\epsilon\mathbf{u}_t^{(2)} - \zeta^{(2)}\Big) \\ &- (1 - P_{11})\langle\psi_t^{(1)}|(H - E)\psi_t^{(2)}\rangle - P_{11}\langle\psi_t^{(1)}|(H^\dagger - E)\psi_t^{(2)}\rangle. \end{aligned} \tag{15}$$

As implied by its derivation, and proved in Appendix C, Eq. (15) is stationary under all continuous infinitesimal variations of $\psi_t^{(1)}$ and $\psi_t^{(2)}$ if and only if $\psi_t^{(1)}$ and $\psi_t^{(2)}$ satisfy Eqs. (8) and (10), the latter of which are now

$$\mathbf{u}_e^{(2)\prime} - (2\mathrm{m}/\hbar^2)\epsilon\mathbf{u}_e^{(2)} = \zeta^{(2)} \tag{16a}$$

and

$$\mathbf{u}_e^{(1)*\prime} - (2\mathrm{m}/\hbar^2)\tilde{\epsilon}\mathbf{u}_e^{(1)*} = \zeta^{(1)*}. \tag{16b}$$

When the exact wavefunctions which satisfy Eqs. (8) and (16) are used as arguments, G reduces to

$$G[\psi_e^{(1)}, \psi_e^{(2)}] = P_{11}\mathbf{u}_e^{(1)\dagger}\frac{\hbar^2}{2\mathrm{m}}\zeta^{(2)} + (1 - P_{11})\zeta^{(1)\dagger}\frac{\hbar^2}{2\mathrm{m}}\mathbf{u}_e^{(2)}. \tag{17a}$$

As shown in Appendix C, the terms multiplying P_{11} and $(1 - P_{11})$ are equal, so that the above can be simplified to

$$G[\psi_e^{(1)}, \psi_e^{(2)}] = \zeta^{(1)\dagger}\frac{\hbar^2}{2\mathrm{m}}\mathbf{u}_e^{(2)}. \tag{17b}$$

Since the above is the stationary value of G, we may also write, using approximate trial functions,

$$\zeta^{(1)\dagger}\frac{\hbar^2}{2\mathrm{m}}\mathbf{u}_e^{(2)} = G[\psi_t^{(1)}, \psi_t^{(2)}] + O(\Delta\psi^{(1)}\times\Delta\psi^{(2)}). \tag{18}$$

Equations (4), (7), and (15)—(18) constitute the main results of this work.

When the unknown (but second order) error term in Eq. (18) is dropped, the right-hand side can be evaluated for any pairs of trial wavefunctions, to yield approximate values of the $u_{ce}^{(2)}$. These, together with the boundary conditions, Eqs. (16), are precisely the kind of intermediate results one needs in a scattering calculation.

It follows from the derivation that Eq. (15) for G includes all the suitable stationary functionals which have the general form of Eq. (4), assuming Eq. (10) for the boundary conditions. We note that G contains several arrays of constants: $\mathbf{a}$, ϵ, ζ, and the scalar P_{11}. These are parameters of the *method*, not necessarily of the trial functions themselves. Since these parameters may be chosen almost at will, we have here a multidimensional continuum of suitable variational functionals.

A simpler functional results when ϵ is a diagonal matrix. In this case, let us define b_c for all c by

$$b_c = a_c(2m_c/\hbar^2)\epsilon_{cc}, \tag{19}$$

and write Bloch's $\mathcal{L}$ operator,[35]

$$\mathcal{L}(\mathbf{b}) = \sum_c |c\rangle \frac{\hbar^2}{2m_c a_c} \delta(r_c - a_c)\left(\frac{\partial}{\partial r_c} - \frac{b_c}{a_c}\right) r_c \langle c|, \tag{20a}$$

in which the sum extends over all open and closed two-body channels. Also define operators $\mathcal{A}(\mathbf{b})$ and $\bar{\mathcal{A}}$

$$\mathcal{A}(\mathbf{b}) = H + \mathcal{L}(\mathbf{b}) - E, \tag{20b}$$

$$\bar{\mathcal{A}} = (1 - P_{11})\mathcal{A}(\mathbf{b}) + P_{11}\mathcal{A}(\mathbf{b}^*)^\dagger. \tag{20c}$$

Equation (15) is now

$$G[\psi_t^{(1)}, \psi_t^{(2)}] = \sum_c [(\hbar^2/2m_c)(\xi_c^{(1)*} u_{ct}^{(2)} + u_{ct}^{(1)*} \xi_c^{(2)}] + z - \langle \psi_t^{(1)} | \bar{\mathcal{A}} \psi_t^{(2)} \rangle, \tag{21}$$

where z is the term involving ξ. The boundary conditions (16) for the exact wavefunctions become

$$u_{ce}^{(1)*\prime} - (b_c/a_c) u_{ce}^{(1)*} = \xi_c^{(1)*}, \tag{22a}$$

$$u_{ce}^{(2)\prime} - (b_c/a_c) u_{ce}^{(2)} = \xi_c^{(2)}, \tag{22b}$$

or, equivalently,

$$\mathcal{L}(\mathbf{b}^*)\psi_e^{(1)} = \sum_c |c\rangle \frac{\hbar^2}{2m_c a_c} \delta(r_c - a_c)\xi_c^{(1)}, \tag{22c}$$

$$\mathcal{L}(\mathbf{b})\psi_e^{(2)} = \sum_c |c\rangle \frac{\hbar^2}{2m_c a_c} \delta(r_c - a_c)\xi_c^{(2)}. \tag{22d}$$

No such conditions have been assumed for the *trial* wavefunctions.

The boundary conditions, (16), and the simpler ones, (22), are related, in the sense that a wavefunction satisfying the former conditions is a linear combination of up to n wavefunctions satisfying the latter ones, if there are n channels. It is not clear whether any useful flexibility is lost by using the simpler conditions, i.e., by passing from Eqs. (15) and (16) to Eqs. (21) and (22). At any rate, Eq. (21) is the form of the general functional which will be used in the remainder of this work, except in Appendix B, where a different form of the boundary conditions is employed.[43]

A. Wavefunctions outside S

Outside of S, the wavefunction can be expanded, according to the original assumptions, as a sum of two-body terms,

$$\psi = \sum_c |c\rangle r_c^{-1} u_c(r_c), \quad r_c \geq a_c. \tag{23}$$

The radial wavefunctions $u_c(r_c)$ satisfy the following equations when $r_c \geq a_c$:

$$\left[\frac{\hbar^2}{2m_c}\left(-\frac{d^2}{dr_c^2} + \frac{l_c(l_c+1)}{r_c^2} - k_c^2\right) + V_c\right] u_c(r_c) = 0, \tag{24}$$

in which k_c^2 is the square of the channel wavenumber and V_c is the channel two-body interaction potential.

Since the exact wavefunction and its gradient are both continuous across S, it follows that

$$u_c(a_c) = u_c, \tag{25}$$

$$\left[\frac{d}{dr_c} u_c(r_c)\right]_{a_c} = u_c', \tag{26}$$

where u_c and u_c' (when they appear with no argument) are surface quantities defined as in Eq. (3), calculated from the *inside* wavefunction.

B. The R matrix

The relations between the surface quantities $u_c^{(2)}$ and $u_c^{(2)\prime}$ may in general be expressed by the Wigner R matrix,[34,44] which is defined by

$$\left(\frac{\hbar^2}{2m_r a_r}\right)^{1/2} u_r^{(2)} = \sum_s R_{rs}\left(\frac{\hbar^2 a_s}{2m_s}\right)^{1/2}\left(u_s^{(2)\prime} - \frac{b_s}{a_s} u_s^{(2)}\right). \tag{27}$$

Now, consider wavefunctions satisfying the boundary conditions of Eq. (22) with arbitrary $\mathbf{b}$ and the following ξ:

$$\xi_c^{(1)} = \left(\frac{2m_r}{\hbar^2 a_r}\right)^{1/2} \delta_{cr}, \quad \text{all } c, \tag{28a}$$

$$\xi_c^{(2)} = \left(\frac{2m_s}{\hbar^2 a_s}\right)^{1/2} \delta_{cs}, \quad \text{all } c. \tag{28b}$$

According to Eq. (28a), the left-hand side of Eq. (27) is identical to the stationary value of G [see Eq. (17)]. Furthermore, because of Eq. (28b), the right-hand side of Eq. (27) reduces to just R_{rs}. Therefore, R_{rs} may be approximately calculated from the general functional, Eq. (21), which becomes

$$G = R_{rs} = (\hbar^2/2m_s a_s)^{1/2} u_{st}^{(1)*} + (\hbar^2/2m_r a_r)^{1/2} u_{rt}^{(2)} + z - \langle \psi_t^{(1)} | \bar{\mathcal{A}} \psi_t^{(2)} \rangle. \tag{29}$$

The right-hand side of this equation, including the surface quantities [see Eq. (3)], is evaluated from trial wavefunctions. The error in the expression is of the order of the product of the errors of the two trial functions [Eq. (18) and Appendix C].

Note that the values of the b_c were not specified in deriving the above equation. In fact, a different R matrix is defined [Eq. (27)] for each different set of these constants.[34] In a particular case, the choice of the b_c might be dictated either by the ease of application of the associated R matrix or by considerations of accuracy and convenience in the variational calculation. It will ordinarily be useful to choose, for the *closed* channels, the values of the b_c suggested in the next section. When this is done, all the information concerning physical scattering amplitudes is contained in the submatrix (with R) which connects only open channels.

The scattering matrix may be readily obtained[34] from the R matrix, if solutions of the radial wave equations (outside of S) are in hand.

Algebraic methods using Eq. (29) will be discussed in Sec. III. For the time being, we note that, although the exact wavefunctions [which make Eq. (29) stationary with respect to all well-behaved variations] satisfy boundary conditions expressed by Eq. (22), the trial wavefunctions need not be restricted by boundary conditions. This is in marked contrast to the standard R-matrix method, in which the inside wavefunction is expanded in terms of functions which all satisfy the homogenous boundary conditions $\mathcal{L}f = 0$.

C. The generalized collision matrix

Let $I_c(r_c)$ and $O_c(r_c)$ be two linearly independent solutions of the radial wave equation (24) in the c channel (*not* necessarily incoming and outgoing waves). For

closed channels, $O_c^-(r_c)$ will often be chosen as the solution which approaches zero at large r_c. Plus and minus signs will be used as superscripts to distinguish between open and closed channels, respectively, when necessary to do so. Define the generalized collision matrix[34] **X** as that matrix which relates the amplitudes of the I and O terms in the wavefunction. Thus, in an n-channel problem, a particular set of wavefunctions has the following radial wavefunctions:

$$u_{re}^{s}(r_r) = I_s(r_s)\delta_{rs} - O_r(r_r)X_{rs}, \quad s = 1,2,\ldots,n. \tag{30}$$

To find elements of **X**, consider that particular *inside* wavefunction whose associated surface quantities have the same form as in Eq. (30); i.e.,

$$\begin{aligned} u_{ce}^{(2)} &= I_s\delta_{cs} - O_cX_{cs}, \quad \text{all } c, \\ u_{ce}^{(2)\prime} &= I_s'\delta_{cs} - O_c'X_{cs}, \quad \text{all } c. \end{aligned} \tag{31}$$

Since **X** is not known in advance, the above result is to be achieved by selecting values of the constants in the inhomogenous boundary conditions, Eq. (22b), such that the coefficients of the I and I' terms in the surface quantities are the Kronecker deltas given above, regardless of **X**. This requires the following:

$$b_c = a_cO_c'/O_c, \quad \text{all } c, \tag{32a}$$

$$\begin{aligned} \xi_c^{(2)} &= O_s^{-1}(I_s'O_s - O_s'I_s)\delta_{cs}, \quad \text{all } c, \\ &= O_s^{-1}W_s\delta_{cs}, \quad \text{all } c. \end{aligned} \tag{32b}$$

The prime denotes r differentiation, and W_s is a Wronskian. The argument of O_c, O_c', I_c, or I_c' should be understood to be a_c when not specified. Now, let

$$\xi_c^{(1)*} = W_cO_c^{-1}\delta_{cr}, \quad \text{all } c. \tag{33}$$

Then, after using Eq. (17b) to eliminate $u_{ce}^{(2)}$, Eq. (31) may be rearranged to yield the following:

$$X_{rs} = O_r^{-1}I_s\delta_{rs} - (2m_r/\hbar^2)W_r^{-1}G[\psi_e^{(1)}, \psi_e^{(2)}]. \tag{34}$$

Finally, when the exact G is approximated by the corresponding functional, Eq. (21), the following stationary expression for X_{rs} results:

$$\begin{aligned} X_{rs} = O_r^{-1}[I_s\delta_{rs} - u_{rt}^{(2)}] - (2m_r/\hbar^2)W_r^{-1} \\ \times\{(\hbar^2/2m_s)(W_s/O_s)u_{st}^{(1)*} + z - \langle\psi_t^{(1)}|\bar{A}\psi_t^{(2)}\rangle\}. \end{aligned} \tag{35}$$

D. The a-independent functional

Many variational scattering calculations employ infinite channel radii. Although Eq. (35) is valid for arbitrarily large a_c, its terms oscillate, and do not individually have limits as the a_c go to infinity. Therefore, it is of interest to write the variational functional for **X** in a different form, which is directly suited to the large-a limit. It is convenient for this purpose to express the $u_{ct}^{(\mu)}$ and $u_{ct}^{(\mu)\prime}$ values in terms of α and β coefficients defined by the relations

$$\begin{aligned} u_{ct}^{(2)} &= I_c\alpha_c^{(2)} - O_c\beta_c^{(2)}, \\ u_{ct}^{(2)\prime} &= I_c'\alpha_c^{(2)} - O_c'\beta_c^{(2)}, \\ u_{ct}^{(1)*} &= I_c\alpha_c^{(1)*} - O_c\beta_c^{(1)*}, \\ u_{ct}^{(1)*\prime} &= I_c'\alpha_c^{(1)*} - O_c'\beta_c^{(1)*}. \end{aligned} \tag{36}$$

[That complex conjugation should be used, for simplicity, in the latter two equations, is suggested by Eqs. (22a) and (22b), and is a manifestation of an asymmetry between $\psi_e^{(1)}$ and $\psi_e^{(2)}$ which is discussed in Sec. III. E. 3.] Then, Eq. (35) becomes, if (32) and (33) are still in force,

$$\begin{aligned} X_{rs} = \frac{2m_r}{\hbar^2W_r}\Bigg\{ & P_{11}\left[\beta_s^{(1)*}\frac{\hbar^2}{2m_s}W_s - \sum_i\left(\frac{\hbar^2}{2m_i}W_i[\alpha_i^{(1)*} - \delta_{ir}]\beta_i^{(2)}\right)\right] \\ & + (1-P_{11})\left[W_r\frac{\hbar^2}{2m_r}\beta_r^{(2)} - \sum_i\left(\frac{\hbar^2}{2m_i}W_i\beta_i^{(1)*}[\alpha_i^{(2)} - \delta_{is}]\right)\right] \\ & + \sum_{ij}\left[(\alpha_i^{(1)*} - \delta_{ir})\left(\frac{\hbar^2W_iI_i}{2m_iO_i}\delta_{ij} - \frac{W_i}{O_i}\xi_{ij}\frac{W_j}{O_j}\right)(\alpha_j^{(2)} - \delta_{js})\right] \\ & + \langle\psi_t^{(1)}|[(1-P_{11})H + P_{11}H^\dagger - E]\psi_t^{(2)}\rangle\Bigg\}. \end{aligned} \tag{37}$$

The same wavefunctions make both Eqs. (35) and (37) stationary. The boundary conditions, Eqs. (32) and (33), for the exact wavefunctions, have the following simple form:

$$\alpha_{ce}^{(1)} = \delta_{cr}, \quad \alpha_{ce}^{(2)} = \delta_{cs}, \tag{38}$$

for all channels c, if $\alpha_{ce}^{(1)}$ and $\alpha_{ce}^{(2)}$ are defined as in Eq. (36), but for exact wavefunctions. Note that these conditions need not necessarily be enforced on the *trial* wavefunctions of an approximate calculation.

Let us assume that $I_c(r_c)$ and $O_c(r_c)$ are defined independently of a_c. Then, the generalized collision matrix **X** is also independent of the a_c. However, the double sum in Eq. (37) does not in general have a large-a limit. Since the ξ_{ij} have so far remained arbitrary, we may choose them so as to cancel the offending terms. Thus, take

$$\xi_{ij} = \frac{\hbar^2I_iO_i}{2m_iW_i}\delta_{ij} - \frac{O_iO_j}{W_iW_j}\tau_{ij}, \tag{39}$$

in which the τ_{ij} are arbitrary parameters with no a dependence. Then Eq. (37) becomes

$$\begin{aligned} X_{rs} = \frac{2m_r}{\hbar^2W_r}\Bigg\{ & P_{11}\left[\beta_s^{(1)*}\frac{\hbar^2}{2m_s}W_s - \sum_i\left(\frac{\hbar^2}{2m_i}W_i[\alpha_i^{(1)*} - \delta_{ir}]\beta_i^{(2)}\right)\right] \\ & + (1-P_{11})\left[W_r\frac{\hbar^2}{2m_r}\beta_r^{(2)} - \sum_i\left(\frac{\hbar^2}{2m_i}W_i\beta_i^{(1)*}[\alpha_i^{(2)} - \delta_{is}]\right)\right] \\ & + \sum_{ij}[(\alpha_i^{(1)*} - \delta_{ir})\tau_{ij}(\alpha_j^{(2)} - \delta_{js})] + \langle\psi_t^{(1)}|[(1-P_{11})H \\ & + P_{11}H^\dagger - E]\psi_t^{(2)}\rangle\Bigg\}. \end{aligned} \tag{40}$$

The above functional has the desired properties. When all the channel radii are infinite, the trial functions must be chosen such that the surface quantities $\alpha_c^{(\mu)}$ and $\beta_c^{(\mu)}$ approach constant values at large a_c. (The $\alpha_c^{(\mu)}$ and $\beta_c^{(\mu)}$ will be identifiable as constant coefficients in the asymptotic part of the trial wavefunction.) Then, all the terms in Eq. (40) approach constant values at large a_c, so the surface S may be moved out to infinity in any or all radial directions.

We note in passing that if the trial functions themselves satisfy Eq. (38), the double sum in Eq. (37) vanishes. Then, Eqs. (37) and (40) become identical.

E. Comparison with some previously derived functionals

1. tanη

It is of interest to see how some variational methods in current use for nuclear and atomic scattering are

related to the general formulas of the present work. First, consider Kohn's principle[29] for $\tan\eta$, the tangent of the s wave phase shift, which applies to the scattering of two structureless particles, if at least one of them is not charged. Kohn proved that the following is a stationary functional for $\tan\eta$,

$$\tan\eta = \beta - \frac{2m}{\hbar^2 k}\int_0^\infty \psi_t(\mathcal{H}-E)\psi_t(4\pi r^2)dr \tag{41}$$

provided that the trial wavefunction is real, and asymptotically has the form

$$\psi_t \sim (4\pi)^{-1/2} r^{-1}(\alpha \sin kr + \beta \cos kr) \tag{42}$$

with $\alpha = 1$ and β free to vary.

Let us apply the general a-independent stationary functional to the same problem. Let a be infinite, and take I and O to be $\sin kr$ and $-\cos kr$, respectively. For simplicity, use a single real trial function, the same one as above, but leave both constants α and β undetermined. Then, X is $\tan\eta$, and Eq. (40) becomes

$$\tan\eta = \beta(2-\alpha) - (2m/\hbar^2 k)[(\alpha-1)^2\tau + \textstyle\int_0^\infty \psi_t(\mathcal{H}-E)\psi_t(4\pi r^2)dr]. \tag{43}$$

This stationary expression differs from Kohn's, but it reduces to Kohn's equation[29] if α is given the value unity. [Note from Eq. (38) that $\alpha = 1$ is the inhomogenous boundary condition[40] satisfied by the *exact* wavefunction which makes the general functional stationary.]

2. cot (η - θ)

A more general variational principle than Kohn's, for scattering by a center of force, is that of Kato,[45] who proved that the following expression is correct to first order:

$$\cot(\eta-\theta) = \beta + \frac{2m}{\hbar^2 k}\int_0^\infty \psi_t(\mathcal{H}-E)\psi_t(4\pi r^2)dr \tag{44}$$

for real trial functions which asymptotically have the form

$$\psi_t \sim (4\pi)^{-1/2} r^{-1}[\alpha \cos(kr+\theta) + \beta \sin(kr+\theta)] \tag{45}$$

with $\alpha = 1$, where θ is a fixed real constant. If θ is $\pi/2$, or zero, Eq. (44) is respectively Kohn's principle,[29] or the inverse Kohn (or Rubinow) principle,[4] for $\cot\eta$.

The following similar stationary expression results from the general a-independent functional, Eq. (40), if $I(r)$ and $O(r)$ are $\cos(kr+\theta)$ and $-\sin(kr+\theta)$, respectively [so that X is $\cot(\eta-\theta)$] and both α and β are among the variational parameters in the single real trial function ψ_t:

$$\cot(\eta-\theta) = \beta(2-\alpha) + (2m/\hbar^2 k)[(\alpha-1)^2\tau + \textstyle\int_0^\infty \psi_t(\mathcal{H}-E)\psi_t(4\pi r^2)dr]. \tag{46}$$

This is more general than Kato's principle, as α is here permitted to be one of the variational parameters in the trial function. If, however, ψ_t is required to satisfy the same inhomogenous boundary condition that the exact wavefunction does, then $\alpha = 1$, and Eq. (46) reduces to Kato's equation.

Kato's principle may also be derived directly from the general functional, Eq. (15) or (21), by letting $b = ka\cot(ka+\theta)$, and enforcing the boundary condition (22) on the trial wavefunction. When the latter is done, the z term drops out of the general and a-independent expressions, (35) and (40), respectively, which then become equivalent.

3. S matrix

Next, we consider some variational expressions for the S matrix. Let $I_c(r_c)$ and $O_c(r_c)$, in terms of which the generalized collision matrix is defined [Eq. (30)], have the following asymptotic forms in the open channels (assuming there is no r_c^{-1} term in the potential):

$$\begin{aligned} I_c^*(r_c) &\sim (m_c/\hbar k_c)^{1/2}\exp(-ik_c r_c + l_c\pi/2), \\ O_c^*(r_c) &\sim (m_c/\hbar k_c)^{1/2}\exp(+ik_c r_c - l_c\pi/2), \end{aligned} \tag{47}$$

so that $\mathbf{X}$ is the scattering matrix $\mathbf{S}$. Then the general a-independent stationary expression (40) becomes

$$S_{rs} = \beta_r^{(2)} - \sum_i \beta_i^{(1)*}(\alpha_i^{(2)} - \delta_{is}) + (i/\hbar)\sum_{ij}(\alpha_i^{(1)*} - \delta_{ir}) \times \tau_{ij}(\alpha_j^{(2)} - \delta_{js}) + (i/\hbar)\langle\psi_t^{(1)}|(\mathcal{H}-E)\psi_t^{(2)}\rangle, \tag{48}$$

where the α and β constants are the surface quantities defined by Eq. (36) and, for simplicity, we have put $P_{11} = 0$. The wavefunctions which make Eq. (48) exactly stationary satisfy the Schrödinger equation and the boundary conditions, for all c,

$$\alpha_{ce}^{(1)} = \delta_{cr}, \quad \alpha_{ce}^{(2)} = \delta_{cs}. \tag{49}$$

As discussed in Sec. IID, it is not necessary to apply these conditions to the trial functions themselves, and it may not even by desirable to do so. But, if it is done, Eq. (48) reduces to

$$S_{rs} = \beta_r^{(2)} + (i/\hbar)\langle\psi_t^{(1)}|(\mathcal{H}-E)\psi_t^{(2)}\rangle. \tag{50}$$

This equation has been studied by us, using finite radii a_i, for its applicability to chemical reactions.[6] If the radii are infinite, the equation is identical to Kohn's S-matrix method.[29]

4. Lane and Robson's functional

Lane and Robson[46] have obtained interesting and useful results from the following stationary functional M,

$$-M[\psi_t] = \langle\psi_e^{(1)}|\mathcal{A}\psi_t^{(2)}\rangle + \langle\psi_t^{(1)}|\mathcal{A}\psi_e^{(2)}\rangle - \langle\psi_t^{(1)}|\mathcal{A}\psi_t^{(2)}\rangle, \tag{51}$$

in which $\psi_e^{(1)}$ is required to be related to $\psi_e^{(2)}$ (and likewise $\psi_t^{(1)}$ to $\psi_t^{(2)}$) in such a way that the surface quantities associated with them are the complex conjugates of each other, i.e.,

$$u_{ce}^{(1)} = u_{ce}^{(2)*}, \quad u_{ce}^{(1)\prime} = u_{ce}^{(2)\prime *}, \tag{52}$$

for all channels (but see the end of this section).

It is instructive to bring the general functional, Eq. (21), to a form similar to (51). As it stands, Eq. (21) is

$$G = \langle\psi_e^{(1)}|\mathcal{L}(\mathbf{b}^*)^\dagger\psi_t^{(2)}\rangle + \langle\psi_t^{(1)}|\mathcal{L}(\mathbf{b})\psi_e^{(2)}\rangle - \langle\psi_t^{(1)}|\bar{\mathcal{A}}\psi_t^{(2)}\rangle + z. \tag{53}$$

Now, $(\mathcal{H}-E)^\dagger$ and $(\mathcal{H}-E)$, respectively, may be added to the operators in the first two integrals above without changing their values. Also, we note that, as a consequence of Green's theorem [see Eq. (6)],

$$\langle f|\mathcal{A}(\mathbf{b}^*)^\dagger g\rangle = \langle f|\mathcal{A}(\mathbf{b})g\rangle = \langle f|\bar{\mathcal{A}}g\rangle \tag{54}$$

for any pair of functions f and g, as long as at least one of them has, on S, two-body channel components only. Therefore, Eq. (53) is

$$G = \langle \psi_e^{(1)}|\bar{\mathcal{A}}\psi_t^{(2)}\rangle + \langle \psi_t^{(1)}|\bar{\mathcal{A}}\psi_e^{(2)}\rangle - \langle \psi_t^{(1)}|\bar{\mathcal{A}}\psi_t^{(2)}\rangle + z. \tag{55}$$

Finally, if we set all the ξ_{rs} to zero, so z is zero, and let P_{11} be zero also, the above reduces to Lane and Robson's $-M$.

Notice that the relation (52) between $\psi_e^{(1)}$ and $\psi_e^{(2)}$ (and between $\psi_t^{(1)}$ and $\psi_t^{(2)}$) was not assumed in the above derivation, so that such an equation is neither a condition for, nor a consequence of, G being a stationary functional. Indeed, Eq. (52) is true only when $\zeta^{(1)} = \zeta^{(2)*}$ [see Eq. (22)], which is not compatible with our preceding stationary expressions for off-diagonal elements of the matrices **R**, **X**, and **S**.

III. ALGEBRAIC METHODS

This section concerns variational calculations in which the trial wavefunctions are linear combinations of two sets of basis functions $\chi_i^{(1)}$ and $\chi_i^{(2)}$, whose values and gradients are defined everywhere inside and on the surface S:

$$\psi_t^{(\mu)} = \sum_i C_i^{(\mu)}\chi_i^{(\mu)}, \tag{56}$$

where the $C_i^{(\mu)}$ are coefficients, initially undetermined.

Let $u_{ci}^{(\mu)}$ and $u_{ci}^{(\mu)\prime}$ be surface integrals defined as in Eq. (3),[39] but with $\chi_i^{(\mu)}$ in place of $\psi_t^{(\mu)}$, and define

$$\begin{aligned}\bar{\mathcal{A}}_{ij} &= \langle \chi_i^{(1)}|\bar{\mathcal{A}}\chi_j^{(2)}\rangle,\\ f_{ci}^{(1)} &= u_{ci}^{(1)\prime} - (b_c^*/a_c)u_{ci}^{(1)},\\ f_{ci}^{(2)} &= u_{ci}^{(2)\prime} - (b_c/a_c)u_{ci}^{(2)}.\end{aligned} \tag{57}$$

Then G, Eq. (21), is given by

$$\begin{aligned}G = \sum_{rsi}\{&\xi_r^{(1)*}[\delta_{rs}(\hbar^2/2m_s)u_{si}^{(2)} - \xi_{rs}f_{si}]C_i^{(2)}\\ &+ C_i^{(1)*}[u_{ri}^{(1)*}(\hbar^2/2m_r)\delta_{rs} - f_{ri}^{(1)*}\xi_{rs}]\zeta_s^{(2)}\}\\ &+ \sum_{ij}C_i^{(1)*}[\sum_{rs}f_{ri}^{(1)*}\xi_{rs}f_{sj}^{(2)} - \bar{\mathcal{A}}_{ij}]C_j^{(2)}\\ &+ \sum_{rs}\zeta_r^{(1)*}\xi_{rs}\zeta_s^{(2)}.\end{aligned} \tag{58}$$

In matrix notation, Eq. (58) is

$$\begin{aligned}G = \zeta^{(1)\dagger}&[(\hbar^2/2m)\mathbf{u}^{(2)} - \xi\mathbf{f}^{(2)}]\mathbf{C}^{(2)} + \mathbf{C}^{(1)\dagger}[(\mathbf{u}^{(1)\dagger}\hbar^2/2m)\\ &- \mathbf{f}^{(1)\dagger}\xi]\zeta^{(2)} + \mathbf{C}^{(1)\dagger}[\mathbf{f}^{(1)\dagger}\xi\mathbf{f}^{(2)} - \bar{\mathbf{A}}]\mathbf{C}^{(2)} + \zeta^{(1)\dagger}\xi\zeta^{(2)},\end{aligned} \tag{59}$$

where the dagger represents the complex conjugate of the transpose, the $\mathbf{C}^{(\mu)}$ and $\zeta^{(\mu)}$ are column vectors, and the remaining quantities on the right-hand side are rectangular or square matrices.

In the absence of a general extremum principle for G, it is not obvious what is the best procedure for finding the coefficients $C_i^{(\mu)}$ with which to evaluate the above expression. Several ways of choosing them are discussed and illustrated in subsections A–D, where four general classes of methods are distinguished. The first three classes are all based on making G stationary (with respect to allowed variations of the $C_i^{(\mu)}$), but differ as to how the boundary conditions are handled. In the fourth class, other criteria are employed for selecting the $C_i^{(\mu)}$. Section E concerns the meaning of the parameters of the functional, the $\zeta_c^{(\mu)}$, a_c, b_c, $\xi_{cc'}$, and P_{11}, and how their values affect the outcome of a calculation. The relative merits of the first three classes of methods are considered in Sec. IV.

A. The first class

Since the general functional G was deliberately derived so as to be stationary about its exact value with respect to all continuous variations of the trial wavefunctions, the most pleasing way of treating the algebraic expression (59) for G is to find the value of the right-hand side which is stationary with respect to all variations of the coefficients. Equating to zero the derivative of the right-hand side with respect to each $C_i^{(1)*}$ and solving gives

$$\mathbf{C}^{(2)} = [\bar{\mathbf{A}} - \mathbf{f}^{(1)\dagger}\xi\mathbf{f}^{(2)}]^{-1}[\mathbf{u}^{(1)\dagger}\hbar^2/2m - \mathbf{f}^{(1)\dagger}\xi]\zeta^{(2)}, \tag{60}$$

and, with this $\mathbf{C}^{(2)}$, Eq. (59) becomes

$$\begin{aligned}G = \zeta^{(1)\dagger}\{&[(\hbar^2/2m)\mathbf{u}^{(2)} - \xi\mathbf{f}^{(2)}][\bar{\mathbf{A}} - \mathbf{f}^{(1)\dagger}\xi\mathbf{f}^{(2)}]^{-1}\\ &\times[(\mathbf{u}^{(1)\dagger}\hbar^2/2m) - \mathbf{f}^{(1)\dagger}\xi] + \xi\}\zeta^{(2)}.\end{aligned} \tag{61}$$

(Precisely the same G results from making the derivatives with respect to each $C_i^{(2)}$ be zero.)

According to the developments of Secs. IIB and IIC, the R matrix and the generalized scattering matrix **X** may be obtained directly from G when $\zeta^{(1)}$ and $\zeta^{(2)}$ are given by Eqs. (28). When this is done, using the above value of G, the results are as follows.

$$\begin{aligned}\mathbf{R} = (2m/\hbar^2\mathbf{a})^{1/2}\{&[(\hbar^2/2m)\mathbf{u}^{(2)} - \xi\mathbf{f}^{(2)}][\bar{\mathbf{A}} - \mathbf{f}^{(1)\dagger}\xi\mathbf{f}^{(2)}]^{-1}\\ &\times[(\mathbf{u}^{(1)\dagger}\hbar^2/2m) - \mathbf{f}^{(1)\dagger}\xi] + \xi\}(2m/\hbar^2\mathbf{a})^{1/2},\end{aligned} \tag{62a}$$

for arbitrary **b** (in the definition of **R** and in $\bar{\mathbf{A}}$, $\mathbf{f}^{(1)}$, and $\mathbf{f}^{(2)}$). The above has the following simple form when one chooses $\xi = 0$:

$$\mathbf{R} = (\hbar^2/2m\mathbf{a})^{1/2}\mathbf{u}^{(2)}\bar{\mathbf{A}}^{-1}\mathbf{u}^{(1)\dagger}(\hbar^2/2m\mathbf{a})^{1/2}. \tag{62b}$$

When a single basis set is used, the last equation becomes identical to one derived by Lane and Robson,[46] who argued that it is the best expression for calculating the scattering matrix from a given basis set.[47]

Similarly, the generalized scattering matrix is approximated by

$$\begin{aligned}\mathbf{X} = \mathbf{O}^{-1}\mathbf{I} - \mathbf{O}^{-1}(2m/\hbar^2)\{&[(\hbar^2/2m)\mathbf{u}^{(2)} - \xi\mathbf{f}^{(2)}]\\ &\times[\bar{\mathbf{A}} - \mathbf{f}^{(1)\dagger}\xi\mathbf{f}^{(2)}]^{-1}[(\mathbf{u}^{(1)\dagger}\hbar^2/2m) - \mathbf{f}^{(1)\dagger}\xi] + \xi\}\mathbf{W}\mathbf{O}^{-1},\end{aligned} \tag{63}$$

where $\mathbf{b} = a\mathbf{O}'\mathbf{O}^{-1}$ (in $\mathbf{f}^{(1)}$, $\mathbf{f}^{(2)}$, and $\bar{\mathbf{A}}$) and **I**, **I′**, **O**, **O′**, and **W** are diagonal matrices of quantities defined in Sec. IIC.

The corresponding matrix equation for the a-independent calculation of **X** can be derived from either Eq. (40) or Eq. (63), with the result

$$\begin{aligned}\mathbf{X} = -(2m/\hbar^2\mathbf{W})&[(\hbar^2\mathbf{W}/2m)\beta^{(2)} - \tau\alpha^{(2)}][\mathbf{M} - P_{11}\boldsymbol{\alpha}^{(1)\dagger}\\ &\times(\hbar^2\mathbf{W}/2m)\beta^{(2)} - (1 - P_{11})\boldsymbol{\beta}^{(1)\dagger}(\hbar^2\mathbf{W}/2m)\boldsymbol{\alpha}^{(2)}\\ &+ \boldsymbol{\alpha}^{(1)\dagger}\tau\boldsymbol{\alpha}^{(2)}]^{-1}[\boldsymbol{\beta}^{(1)\dagger}(\hbar^2/2m)\mathbf{W} - \boldsymbol{\alpha}^{(1)\dagger}\tau]\\ &+ (2m/\hbar^2\mathbf{W})\tau,\end{aligned} \tag{64}$$

where τ is the square matrix of the constants τ_{ij}, and $\alpha^{(\mu)}$ and $\beta^{(\mu)}$ are rectangular matrices of constants $\alpha_{ci}^{(\mu)}$ and $\beta_{ci}^{(\mu)}$ defined as in Eq. (36), but with $u_{ci}^{(\mu)}$ (and $u_{ci}^{(\mu)\prime}$) in place of $u_{ct}^{(\mu)}$ (and $u_{ct}^{(\mu)\prime}$), and where M is defined by

$$M_{ij}=\langle\chi_i^{(1)}|(1-P_{11})H+P_{11}H^{\dagger}-E|\chi_j^{(2)}\rangle. \tag{65}$$

Equations (61)—(64) are each stationary with respect to variations in the space of the basis functions. Furthermore, their error is second order with respect to functions omitted from the bases. Therefore, approximate collision matrices and cross sections calculated from any one of these equations should be of better quality than the basis sets themselves.

Equations (61)—(64) are equivalent. That is, they yield the same approximate scattering matrix, if the same bases and the same values of a, b, ξ, and P_{11} are used in each [assuming that $O_c(r_c)$ is determined by b_c, via Eq. (32a), rather than the other way around]. This is because Eqs. (62)—(64) are derivable from (61) using only (i) the boundary conditions, Eq. (22), which are intended to be used along with G, (ii) the relations (27) and (30) which define R and **X**, and (iii) the linearity of the Hamiltonian operator. On the other hand, different choices of the parameters of the method, **a**, **b**, ξ, and P_{11}, produce different approximations to the scattering matrix. These parameters are examined in Sec. IIE.

The calculable theories of Lane and Robson

An interesting expression arises when the exact stationary value of G, Eq. (17b), is equated to its variational approximation, Eq. (61), with the $\xi_{rs}=0$, using forms which fully utilize L:

$$\langle L(\mathbf{b}^*)\psi_e^{(1)}|\psi_e^{(2)}\rangle=\sum_{ij}\langle L(\mathbf{b}^*)\psi_e^{(1)}|\chi_i^{(2)}\rangle(\bar{\mathbf{A}}^{-1})_{ij}\times\langle\chi_j^{(1)}|L(\mathbf{b})\psi_e^{(2)}\rangle, \tag{66}$$

which, since it applies for all $\psi_e^{(1)}$, implies the following approximate relation for the wavefunction on S:

$$|\psi_e\rangle_a=\sum_{ij}|\chi_i^{(2)}\rangle_a(\bar{\mathbf{A}}^{-1})_{ij}\langle\chi_j^{(1)}|L\psi_e\rangle. \tag{67}$$

All terms on the right-hand side are known quantities. This approximation (with a single basis set) was first derived and studied by Lane and Robson,[41] and several nuclear scattering calculations have been performed with it.[19–21] We emphasize that the error in the approximation is only of second order with respect to the truncation error when a finite number of basis functions are used. An alternate derivation of the last equation is given in Appendix D.

B. The second class

A second general way of using the algebraic expression for G, Eq. (59), is to impose, on the trial wavefunctions, the boundary conditions satisfied by the exact wavefunctions [i.e., (22)], and make G stationary with respect to all variations (of the $C_i^{(\mu)}$) which preserve the boundary conditions.

Assume that the two basis sets each contain at least as many members as there are channels, and that the bases have been transformed such that

$$f_{ci}^{(1)}=f_{ci}^{(2)}=\delta_{ci},\quad \text{all } c,i. \tag{68}$$

The bases now divide into two classes: the first n functions (for n channels), and the remainder. It is convenient to label these groups μ and ν, respectively, and to divide the matrices appearing in G, Eq. (59), into blocks, accordingly.

The boundary conditions (22) now become equivalent to

$$\mathbf{C}_{\mu}^{(1)}=\boldsymbol{\zeta}^{(1)},\quad \mathbf{C}_{\mu}^{(2)}=\boldsymbol{\zeta}^{(2)}, \tag{69}$$

and Eq. (59) becomes

$$\begin{aligned}G=&\boldsymbol{\zeta}^{(1)\dagger}[-\bar{\mathbf{A}}_{\mu\mu}+(\hbar^2/2\mathbf{m})\mathbf{u}_{\mu}^{(2)}+\mathbf{u}_{\mu}^{(1)\dagger}\hbar^2/2\mathbf{m}]\boldsymbol{\zeta}^{(2)}\\&+\boldsymbol{\zeta}^{(1)\dagger}[-\bar{\mathbf{A}}_{\mu\nu}+(\hbar^2/2\mathbf{m})\mathbf{u}_{\nu}^{(2)}]\mathbf{C}_{\nu}^{(2)}\\&+\mathbf{C}_{\nu}^{(1)\dagger}[-\bar{\mathbf{A}}_{\nu\mu}+\mathbf{u}_{\nu}^{(1)\dagger}\hbar^2/2\mathbf{m}]\boldsymbol{\zeta}^{(2)}-\mathbf{C}_{\nu}^{(1)\dagger}\bar{\mathbf{A}}_{\nu\nu}\mathbf{C}_{\nu}^{(2)}.\end{aligned} \tag{70}$$

Now, the right-hand side of the above is stationary with respect to arbitrary variations of $\mathbf{C}_{\nu}^{(1)}$ if and only if

$$(\bar{\mathbf{A}}_{\nu\mu}-\mathbf{u}_{\nu}^{(1)\dagger}\hbar^2/2\mathbf{m})\boldsymbol{\zeta}^{(2)}+\bar{\mathbf{A}}_{\nu\nu}\mathbf{C}_{\nu}^{(2)}=0. \tag{71}$$

Solving for $\mathbf{C}_{\nu}^{(2)}$ and substituting into Eq. (70) gives the stationary value of G:

$$\begin{aligned}G=&\boldsymbol{\zeta}^{(1)\dagger}\{-\bar{\mathbf{A}}_{\mu\mu}+(\hbar^2/2\mathbf{m})\mathbf{u}_{\mu}^{(2)}+\mathbf{u}_{\mu}^{(1)\dagger}\hbar^2/2\mathbf{m}\\&+[\mathbf{A}_{\mu\nu}-(\hbar^2/2\mathbf{m})\mathbf{u}_{\nu}^{(2)}](\bar{\mathbf{A}}_{\nu\nu})^{-1}(\bar{\mathbf{A}}_{\nu\mu}-\mathbf{u}_{\nu}^{(1)\dagger}\hbar^2/2\mathbf{m})\}\boldsymbol{\zeta}^{(2)}\end{aligned} \tag{72a}$$

$$=-\boldsymbol{\zeta}^{(1)\dagger}[(\mathbf{N}^{-1})_{\mu\mu}]^{-1}\boldsymbol{\zeta}^{(2)}, \tag{72b}$$

where

$$\mathbf{N}=\bar{\mathbf{A}}-\mathbf{u}^{(1)\dagger}(\hbar^2/2\mathbf{m})\mathbf{f}^{(2)}-\mathbf{f}^{(1)\dagger}(\hbar^2/2\mathbf{m})\mathbf{u}^{(2)}. \tag{73}$$

We note in passing that **N** is the matrix of the operator,

$$\mathcal{N}(\mathbf{b})=\bar{\mathcal{A}}(\mathbf{b})-L(\mathbf{b})-L(\mathbf{b}^*)^{\dagger}.$$

Equation (72) yields variational methods which are alternatives to those discussed in Sec. IIIA. For example, by using the $\xi_c^{(\mu)}$ values of Eq. (28), one obtains a stationary expression for R:

$$\mathbf{R}=-(2\mathbf{m}/\hbar^2\mathbf{a})^{1/2}[(\mathbf{N}^{-1})_{\mu\mu}]^{-1}(2\mathbf{m}/\hbar^2\mathbf{a})^{1/2}, \tag{74}$$

for arbitrary b_c. [The b_c dependence is subsumed in the initial transformation of the basis sets, to make them satisfy Eq. (68).]

The second type of method may be applied to the a-independent functional, Eq. (40), also, in a similar way. The boundary conditions, Eq. (38), are to be imposed on the trial wavefunctions. To simplify the result, transform the bases somewhat differently from above, such that

$$\alpha_{ci}^{(1)}=\alpha_{ci}^{(2)}=\delta_{ci},\quad \text{all } c \text{ and } i,$$

and divide the matrices into blocks, just as before. We find the following:

$$[\mathbf{M}_{\nu\mu}+P_{11}\beta_{\nu}^{(1)\dagger}(\hbar^2/2\mathbf{m})\mathbf{W}]\mathbf{O}\mathbf{W}^{-1}\boldsymbol{\zeta}^{(2)}+\mathbf{M}_{\nu\nu}\mathbf{C}^{(2)}=0, \tag{75a}$$

$$\begin{aligned}\mathbf{X}=&\mathbf{M}_{\mu\mu}+(2\mathbf{m}/\hbar^2)\mathbf{W}^{-1}\{(1-P_{11})\mathbf{W}(\hbar^2/2\mathbf{m})\beta_{\mu}^{(2)}+P_{11}\beta_{\mu}^{(1)\dagger}\\&\times(\hbar^2/2\mathbf{m})\mathbf{W}-[\mathbf{M}_{\mu\nu}+(1-P_{11})\mathbf{W}(\hbar^2/2\mathbf{m})\beta_{\nu}^{(2)}][\mathbf{M}_{\nu\nu}]^{-1}\\&\times[\mathbf{M}_{\nu\mu}+P_{11}\beta_{\nu}^{(1)\dagger}(\hbar^2/2\mathbf{m})\mathbf{W}]\},\end{aligned} \tag{75b}$$

where **M** is the matrix defined by Eq. (65). Several special cases of the above formula have previously been derived.[1–4]

Note from Eqs. (71), (72a), (34), and (75) that **C**, G, and **X** each has a pole when $\bar{\mathbf{A}}_{\nu\nu}(=\mathbf{M}_{\nu\nu})$ is singular [unless

the pole is cancelled by one of the matrices which multiplies $(\bar{\mathbf{A}}_{\nu\nu})^{-1}$ in Eq. (72a)]. Some of these poles are legitimate, occurring also in the exact result, but the others, first studied by Schwartz, are spurious.[20–32] Since the latter must be reckoned with in any variational method, we will briefly summarize some previous work on the problem, and relate that work to the present formalism.

Brownstein and McKinley,[31] analyzing the finite radius case, considered Kohn's functional for the tangent of the s-wave phase shift, Eq. (41), which is equivalent to Eq. (72) with $b = -ka\tan ka$. They found that if the basis set is good enough, for each pole in b there is, at a slightly higher energy, a spurious pole in $\tan\eta$. However, these poles may be avoided by using a constant b.[22]

With infinite radii, the problem is more severe, as constant b_c cannot be employed (cf. Sec. IID). Two basically different approaches have been used. In the first, one avoids the poles, by choosing the boundary conditions [i.e., by adjusting the definitions of $I_c^*(r_c)$ and $O_c^*(r_c)$] at each energy such that $\mathbf{M}_{\nu\nu}$ is nonsingular. Examples are Nesbet's anomaly-free method,[32] and a method of Seraph, Seaton, and Shemming.[49] The second approach is to make the pole strength vanish, by choosing the boundary conditions such that the factor enclosed in parentheses in Eq. (75a) is zero. This is done in Nesbet and Oberoi's optimized anomaly-free method.[50–51]

The methods of the second type are shown in Sec. IIIE4 to be limiting cases (the large ξ limit) of methods of the first type.

C. The third class

A third general way of using the algebraic expression (59) for G is to make it stationary within a finite space of basis functions $\chi_i^{(\mu)}$, all of which satisfy *homogenous* boundary conditions on S. That is,

$$L\chi_i^{(1)*} = L\chi_i^{(2)} = 0, \quad \text{all } i. \tag{76}$$

According to Eq. (62a), when the general functional is made stationary with respect to variations in this basis, the result is

$$\mathbf{R} = (\hbar^2/2\mathbf{ma})^{1/2}\mathbf{u}^{(2)}\mathbf{M}^{-1}\mathbf{u}^{(1)\dagger}(\hbar^2/2\mathbf{ma})^{1/2}, \tag{77}$$

where $\mathbf{M}$ is the matrix defined in Eq. (65). $\mathbf{R}$ depends on the b_c, which appear in L in Eq. (76). In addition, suppose the basis functions are eigenfunctions of the Hamiltonian, normalized inside S, and with real $u_{ci}^{(1)} = u_{ci}^{(2)}$. This is the standard basis set of the R-matrix method, and, indeed, the stationary expression above is, in this case, equivalent to the standard R-matrix expression[34,44]

$$R_{cc'} = \sum_\lambda \gamma_{c\lambda}(E_\lambda - E)^{-1}\gamma_{c'\lambda}, \tag{78}$$

where the E_λ are the eigenvalues of H (and $H + L$) in the above space, and

$$\gamma_{c\lambda} = (\hbar^2/2m_c a_c)^{1/2} u_{c\lambda}.$$

Another variational derivation of Eq. (78) has been given by Jackson.[52]

It must be emphasized that the above derivation shows Eqs. (77) and (78) to be stationary in a space of only functions satisfying the *homogenous* boundary conditions, Eq. (76). However, the exact wavefunctions satsify *inhomogenous* boundary conditions only, except at discrete level energies, E_λ (for fixed b_c), or for "natural" values[53] of the b_c (for fixed E).

D. The fourth class

A convenient measure of the error of any particular trial wavefunction is the following sum:

$$\sum_i |\langle \Phi_i | (H - E)\psi_t \rangle|^2. \tag{79}$$

In what is called a least squares method, one chooses some of the linear coefficients of ψ_t so as to minimize the above sum, subject to linear relations which the remaining coefficients are forced to satisfy. (Note that the other relations must include inhomogenous boundary conditions, or other normalization, so that other than trivial solutions are obtained.)

The first to suggest the above procedure were Harris and Michels.[1,54] In their minimum norm method, the Φ_i are a certain subset of the basis set, and (79) is used to determine the coefficients of only the asymptotic part part of the wavefunction. The sum (79) contains implicitly an arbitrary set of weight parameters related to the normalization of the Φ_i functions. Another version of this method was suggested by Nesbet and Oberoi.[50]

In Wladawsky's variational least squares method,[55] n coefficients (for an n-channel problem) are taken from the boundary conditions, and all the rest are determined by minimizing expression (79), letting the Φ_i consist of the entire basis set.

Callaway and Wooten,[56] comparing a least squares method with some other methods, do not come to a marked preference. However, Abdallah and Truhlar,[57] in sample calculations, have found Wladawsky's method to work better than several others tried.

Schmid and co-workers[58–61] have shown that results improve when the set Φ_i is augmented with functions not in the basis set of ψ_t.

In the methods described above, the trial functions, once obtained, are inserted into a multichannel version of Kato's relation (30), or some other class two functional, with infinite radii, a_c. One could use a least squares procedure in conjunction with any of the other general methods discussed in the preceding three sections, as well, with either finite or infinite a_c. Furthermore, the least squares process should be especially useful in optimizing nonlinear parameters.

The method of Zvijac, Heller, and Light[43] consists of an R-matrix calculation, a Buttle correction, and a variational correction. The latter is accomplished essentially by modifying the energies of certain R-matrix levels.

E. The influence of the parameters

Within the framework of the general variational functional, as written in Eq. (21), any calculational method can be described by the values chosen for the param-

eters $\xi_c^{(\mu)}$, a_c, b_c, $\xi_{cc'}$, and P_{11}, and the choice of trial functions. Let us consider the selection and influence of these parameters in algebraic variational calculations.

1. The $\xi_c^{(1)}$ and $\xi_c^{(2)}$

Since the $\xi_c^{(\mu)}$ serve as the inhomogenous parts of boundary conditions,[40] they establish (for given b_c and a_c) the normalization of the exact wavefunctions. As a consequence, the values of the $\xi_c^{(\mu)}$ determine what particular element (or linear combination of elements) of a generalized scattering matrix is being approximated [cf. Eqs. (28), (29), (32)–(34), and Sec. 3 below].

The $\xi_c^{(\mu)}$ were put proportional to Kronecker deltas in the preceding derivations when formulas were derived for individual elements of collision matrices. However, since both the exact stationary value of G and the values of G determined by algebraic variational calculations transform in the same way with the $\xi_c^{(1)}$ and $\xi_c^{(2)}$, the approximate collision matrices are independent of the choice of the $\xi_c^{(\mu)}$.

2. The a_c

The significance of the radii a_c is well known from their use in R matrix theory. The two extremes of the a_c are the most used—as small as possible (consistent with the assumptions of Sec. II), or infinite. In the latter case, one must use the a-independent functional, Eq. (64). Philpott and George[62] have studied the dependence of calculated results on the a_c, and have proposed a criterion for the optimal choice of these parameters.

3. The b_c

The parameters b_c determine (for given a_c) what part of each radial wavefunction is directly affected by the Bloch operator and the boundary condition. For example suppose $b_c = a_c O_c'/O_c$ for all c, as in Eq. (32). Then, $L\psi_e^{(2)}$ contians no $O_c(r_c)$ waves on S, and the boundary condition (22b), $u_{ce}^{(2)\prime} - (b_c/a_c)u_{ce}^{(2)} = \xi_c^{(2)}$, implies that the radial wavefunctions $u_{ce}^{(2)}(r_c)$ (defined in $r_c \geq a_c$) have the form $(O_c/W_c)\xi_c^{(2)} I_c(r_c) - \beta_c^{(2)} O_c(r_c)$, where $\beta_c^{(2)}$ is undetermined. That is, the above boundary condition specifies the coefficients of the $I_c(r_c)$ components of the wavefunction, and only those coefficients.

There is an unavoidable asymmetry in the boundary conditions, Eqs. (22a) and (22b), in that, whereas one equation contains $u_{ce}^{(2)\prime}$ and $u_{ce}^{(2)}$, and the other has the *complex conjugates*, $u_{ce}^{(1)*\prime}$ and $u_{ce}^{(1)*}$, both equations use the same b_c (not complex conjugated). Referring to the example of the previous paragraph, the values of b_c stipulated there imply that the functions $u_{ce}^{(1)}(r_c)$ (in $r_c \geq a_c$) have the form $(O_c^*/W_c^*)\xi_c^{(1)} I_c^*(r_c) - \beta_c^{(1)} O_c^*(r_c)$. Now, the functions $I_c(r_c)$ have purposely not been completely specified. However, it can be seen, by comparing the results of this and the preceding paragraph, that if the boundary conditions are chosen so as to control the *incoming* portions of $\psi_e^{(2)}$, then it is the *outgoing* portions of $\psi_e^{(1)}$ which are controlled, and *vice versa*.

The exact G, $\mathbf{R}$, and $\mathbf{X}$ each has a b_c dependence in its definition, but the scattering matrix $\mathbf{S}$, or course, does not. [In the case of $\mathbf{X}$, the b_c dependence arises if one uses the relation $O_c'/O_c = b_c/a_c$ in defining $O_c(r_c)$, for arbitrary complex b_c.] One can easily show (as did Mori[63] for Lane and Robson's theory) that the S calculated from approximate $G(\mathbf{b})$, $R(\mathbf{b})$, or $X(\mathbf{b})$ is independent of the b_c if the $G(\mathbf{b})$, $\mathbf{R}(\mathbf{b})$, or $\mathbf{X}(\mathbf{b})$ is the result of a variational calculation of the first class (Sec. IIIA) performed with a b_c-independent basis and with $\xi = 0$. However, the class two and three variational methods (and class one if $\xi \neq 0$) do not have this type of invariance with the b_c.

The avoidance of spurious singularities in class two calculations by the judicious choice of the b_c was discussed in Sec. IIIB.

4. The ξ_{rs}

The effect of the ξ_{rs} is to control the relative importance of the boundary conditions vs. that of the Schrödinger equation in a class-one variational calculation of wavefunctions and G. Thus, suppose ξ is nonsingular, and replace it by $\lambda\xi$, where λ is a scalar. Then, $\delta G = 0$ becomes

$$\delta G^0 + \lambda \delta z = 0, \tag{80}$$

where G^0 is that part of G, Eq. (15) or (59), which is independent of the ξ_{rs}. Now, $\delta G^0 = 0$ if and only if both (i) the Schrödinger equation and (ii) the boundary conditions, Eq. (22), are satisfied (Appendix B), whereas $\delta z = 0$ has only the boundary conditions as necessary and sufficient conditions. Therefore, according to Eq. (80), the larger (in absolute value) λ is, the more weight is attached to satisfying the boundary conditions, independently of the Schrödinger equation. We have observed this effect in calculations.[64]

The above suggests the hypothesis that algebraic methods of the first class (Sec. IIIA) become equivalent, in the limit of infinite λ, to methods of the second class (Sec. IIIB), since the boundary conditions are satisfied exactly in the latter methods. This will now be proven.

Assume bases which are flexible enough to be able to satisfy the appropriate boundary conditions. Then, without loss of generality, assume that Eq. (68) holds, and divide the bases into two classes, μ and ν, as done in Sec. IIIB. Now consider Eq. (61), the stationary G of the first algebraic method. Perform the matrix inversion indicated in Eq. (61), by blocks, using the four identities of the forms

$$\begin{aligned} (\mathbf{B}^{-1})_{\mu\mu} &= (\mathbf{B}_{\mu\mu} - \mathbf{B}_{\mu\nu}[\mathbf{B}_{\nu\nu}]^{-1}\mathbf{B}_{\nu\mu})^{-1}, \\ (\mathbf{B}^{-1})_{\mu\nu} &= -(\mathbf{B}^{-1})_{\nu\nu}\mathbf{B}_{\nu\mu}(\mathbf{B}_{\mu\mu})^{-1}, \end{aligned} \tag{81}$$

which are valid for nonsingular block-divided $\mathbf{B}$, and using the general relation

$$(\lambda\mathbf{M} + \mathbf{N})^{-1} = \mathbf{M}^{-1}\lambda^{-1} - \mathbf{M}^{-1}\mathbf{N}\mathbf{M}^{-1}\lambda^{-2} + O(\lambda^{-3}), \tag{82}$$

which is valid for any nonsingular $\mathbf{M}$ if λ is large enough in absolute value. Then, expand the right-hand side of Eq. (61) in decreasing powers of λ. The result is

$$G = -\xi^{(1)\dagger}\{[(\mathbf{N}^{-1})_{\mu\mu}]^{-1} + O(\lambda^{-1})\}\xi^{(2)}. \tag{83}$$

In the limit of infinite λ, the above becomes identical to the result, Eq. (72), of the class two treatment. The collision matrices will therefore become identical, also, which is what we intended to prove.

Three interesting special cases examined in this paper can be distinguished according to the values adopted for the ξ_{rs}. Thus, the general expressions for methods of the first kind, (61), (62), or (63), include the following: (i) the basic equation of the Lane and Robson calculable theory,[41] if $\xi = 0$, (ii) the class two methods, if ξ is nonsingular and goes uniformly to infinity, and (iii) the a-independent methods, if ξ satisfies Eq. (39).

5. The P_{11}

The scalar constant P_{11} was introduced as an arbitrary parameter in Eq. (4), and appears in the general functional, Eq. (21), in $\bar{A} = (1 - P_{11})A(\mathbf{b}^*) + P_{11}A(\mathbf{b})$. The best value for P_{11} is undoubtedly $\frac{1}{2}$, as this gives the most symmetric theory. In this case, if the b_c are real, the operator $\bar{A}$ is Hermitian.

If one or both trial wavefunctions have only two-body components on the boundary, it follows from Eq. (54b) that the matrix elements of $\bar{A}$ are independent of P_{11}, and $\bar{A}$ may be replaced by A.

IV. SUMMARY AND CONCLUSIONS

A new functional for the variational calculation of chemical and nuclear reaction cross sections has been derived. It is the most general stationary functional of the bilinear form, Eq. (4), which form was chosen to permit matrix solutions, and, thus, it may well be the most general suitable variational functional for scattering below the three-body threshold. The existence and uniqueness of the solutions was proved.

The derivation assumes the same division of configuration space as made in the R-matrix theory.[34,44] The interaction region, in which all the particles are near each other, is enclosed by a surface S. The dimensions of S are large enough that, outside of it, only two-body channels are significant. The emphasis is placed on approximating wavefunctions inside S, and, especially, on relating the value of the exact physical wavefunction on S to its gradient on S, by a stationary formula. Once relations between those quantities are calculated, it is a relatively simple matter to obtain asymptotic amplitudes of wavefunctions, and, hence, the scattering matrix.

Three points should be made, having to do with the lack of restrictions on trial wavefunctions. First, trial functions are defined inside and on S, only, so their asymptotic behavior is not important. Second, the trial functions do not need to be multichannel expansions, as no such form is ever assumed inside S. Third, although the exact wavefunctions have, by assumption, no three-body channel components on S, the same need not necessarily be true of trial functions. Leaving aside the problem of motion of the center of mass, the latter point means that the theory accomodates trial functions of the shell model type.

From the general functional were derived stationary expressions for the R matrix and the generalized collision matrix $\mathbf{X}$. The latter may be specialized to the reactance matrix or the scattering matrix.

For those situations where it is not desirable to divide the configuration space, the general variational functional for $\mathbf{X}$ was specialized to a form, called the a-independent functional, which has no dependence on the radii of S. In using this functional, the surface S may be moved out to infinity, where it effectively disappears from the theory.

It was shown how the general functional specializes to a number of stationary expressions which have previously been studied: Kohn's $\tan\eta$,[29] the inverse Kohn $\cot\eta$,[4] Kato's $\cot(\eta - \theta)$,[45] our S matrix,[6] Kohn's S matrix,[29] and Lane and Robson's functional.[46]

Several algebraic methods were derived from the general functional, by considering each trial function to be expanded in terms of suitable known functions which are defined inside and on S, and making the functionals stationary with respect to variation of the expansion coefficients. Four classes of methods were considered. In the first, the trial functions are not restricted with respect to boundary conditions on S. The basic equation of the calculable theories of Lane and Robson[41] is included among the methods which were derived of this kind. In the second kind of method, the trial functions are restricted by the same inhomogenous boundary conditions on S that the exact wavefunctions satisfy. Various algebraic results of the Kohn type are derivable from the general functional in this way. In the third kind of method, the basis functions are made to satisfy *homogenous* boundary conditions on S. When, in addition, the basis functions are eigenfunctions of the Hamiltonian, the variational result becomes identical to the fundamental R-matrix relation. In the fourth class are methods which use other criteria for the expansion coefficients.

If they begin with the same primitive basis set, the above methods make the general functional stationary with respect to variations in progressively more restricted function spaces as one passes from class one to two to three. *A priori*, computed cross sections should similarly get worse in the same progression. Indeed, it is fairly well established that Lane and Robson's calculable theory (class one) converges faster than the Wigner–Eisenbud R-matrix method (class three).[53,62] However, the class one and two methods must often give results of similar accuracy. This is because the boundary conditions on the $\psi_t^{(\mu)}$ in class two are the correct ones (in contrast to class three), and imposing them removes only one degree of freedom per channel from the permitted variations, which may or may not be serious.

One situation where a class one procedure would be expected to excel over class two is when the basis functions project into many two-body channels which themselves are of little importance in the scattering problem. In class one, these channels merely contribute terms to the functional, through L, whereas, in class two, the number of basis functions is effectively reduced by one per each of these channels. This suggests that a class two method requires (at least in this case) a significantly larger basis set to give the same accuracy as a class one calculation.

In making the above comparison between class one and class two, we have implicitly assumed that the ξ_{rs}

are zero or small. As shown in Sec. IIIE4, there is no distinction between the two classes when ξ_{rs} are very large. Therefore, the argument implies that zero or small values of the ξ_{rs} are to be preferred. Another advantage of class one methods with the $\xi_{rs}=0$ is that they yield cross section values which are independent of the values assigned to the b_c (Sec. IIIE3).

Choosing a variational method has been shown to reduce to the selection of (i) values for the parameters of the general functional, and (ii) trial wavefunctions. A number of alternate choices have been examined in this work, and the superiority of some has been argued. We hope that these results will facilitate practical calculations, and also lead to further advances in approximate scattering theory.

ACKNOWLEDGMENTS

The author thanks R. L. Becker and D. G Truhlar for helpful comments concerning the manuscript, and L. Spruch and L. Rosenberg for a stimulating discussion.

APPENDIX A

This appendix concerns linear boundary conditions in general, and the existence and uniqueness of solutions of the time-independent Schrödinger equation with boundary conditions. We consider only these solutions having no amplitude in many-body channels on S. The existence of solutions satisfying the latter restriction is a basic assumption of this work.

The Schrödinger equation in an n-channel problem is equivalent to n homogenous independent linear relations, Eq. (27), between the $u'_{c\mathscr{E}}$ and the $u_{c\mathscr{E}}$ defined as in Eqs. (3c) and (3d). Since there are $2n$ of these quantities, the most general set of linear boundary conditions one might consider would contain precisely n $(=2n-n)$ independent linear relations between them. The Schrödinger equation and the boundary conditions together are therefore formally equivalent to the matrix equation

$$\mathbf{B}\mathbf{x}=\mathbf{y}, \tag{A1}$$

in which $\mathbf{B}$ is square, of order $2n$, $\mathbf{x}$ is a column vector of both the $u'_{c\mathscr{E}}$ and the $u_{c\mathscr{E}}$, and $\mathbf{y}$ is a column vector of constants. Let $\mathbf{B}$ be arranged such that the top and bottom halves correspond to the Schrödinger equation and the homogenous part of the boundary conditions, respectively. Then, the top half of $\mathbf{y}$ is zero, while the bottom half of $\mathbf{y}$ consists of the inhomogenous part, if any, of the boundary conditions.

If the boundary conditions are independent of energy, $\mathbf{B}$ is singular at isolated energies E_λ only. (These energies are analogous to the level energies of R-matrix theory.) This is also true, for all practical purposes, when the boundary conditions are energy dependent (assuming, as the usual case, that their values are assigned before any solutions of the Schrödinger equation are in hand), since (i) the n rows of $\mathbf{B}$ within each half of $\mathbf{B}$ are linearly independent, and (ii) the top half of $\mathbf{B}$ is not known until the Schrödinger equation is solved.

It follows that, at all energies except the discrete E_λ, the Schrödinger equation and the boundary conditions together have a unique solution. However, if $\mathbf{y}=0$, the solution is the trivial one $\mathbf{x}=0$. Therefore, for a nontrivial solution to exist, the boundary conditions must be inhomogenous (meaning that an inhomogenous term must appear in at least one channel).

At energies E_λ, the solutions of Eq. (A1) are not unique.

APPENDIX B

The boundary conditions assumed in the main body of this work, Eq. (10), are such that the $u_{c\mathscr{E}}^{(\mu)\prime}$ can be calculated from the $u_{c\mathscr{E}}^{(\mu)}$ and $\zeta_c^{(\mu)}$. In this appendix, the general functional, Eq. (4), will be specialized to a different set of boundary conditions, in which the values of the $u_{c\mathscr{E}}^{(\mu)}$ are specified. In terms of column vectors, these conditions are

$$\mathbf{u}_{\mathscr{E}}^{(\mu)}=\boldsymbol{\zeta}^{(\mu)}. \tag{B1}$$

As discussed in Sec. II, each of Eqs. (7) must be identical to the product of a square matrix and a boundary condition equation, Eq. (85) in this case. Thus, from Eq. (7a) we must put

$$\mathbf{P}_1=0 \tag{B2a}$$

and

$$\mathbf{P}_3=-[\mathbf{P}_2+P_{11}(\hbar^2/2\mathrm{m})]\boldsymbol{\zeta}^{(2)}. \tag{B2b}$$

Similar reasoning applied to the rest of Eq. (7) produces

$$\mathbf{P}_4=P_{11}(\hbar^2/2\mathrm{m}), \tag{B2c}$$

$$\mathbf{P}_6=-\mathbf{P}_5\boldsymbol{\zeta}^{(2)}, \tag{B2d}$$

$$\mathbf{P}_7=-\boldsymbol{\zeta}^{(1)\dagger}(\hbar^2/2\mathrm{m}), \tag{B2e}$$

$$\mathbf{P}_2=P_{10}(\hbar^2/2\mathrm{m}), \tag{B2f}$$

$$\mathbf{P}_8=-\boldsymbol{\zeta}^{(1)\dagger}\mathbf{P}_5. \tag{B2g}$$

In addition, we substitute the symbol ξ for $\mathbf{P}_5$

$$\mathbf{P}_5=\boldsymbol{\xi}, \tag{B2h}$$

and assign the following to $\mathbf{P}_9$:

$$\mathbf{P}_9=\boldsymbol{\zeta}^{(1)\dagger}\boldsymbol{\xi}\boldsymbol{\zeta}^{(2)}. \tag{B2i}$$

As before, we have used

$$P_{10}=1-P_{11}. \tag{B2j}$$

With the above parameters, the general form of G, Eq. (4) becomes

$$\begin{aligned}G'=\mathbf{u}_t^{(1)\dagger\prime}(\hbar^2/2\mathrm{m})([1-P_{11}]\mathbf{u}_t^{(2)}-\boldsymbol{\zeta}^{(2)})+(P_{11}\mathbf{u}_t^{(1)\dagger}-\boldsymbol{\zeta}^{(1)\dagger})\\ \times(\hbar^2/2\mathrm{m})\mathbf{u}_t^{(2)\prime}+(\mathbf{u}_t^{(1)\dagger}-\boldsymbol{\zeta}^{(1)\dagger})\boldsymbol{\xi}(\mathbf{u}_t^{(2)}-\boldsymbol{\zeta}^{(2)})\\ -\langle\psi_t^{(1)}[(1-P_{11})H+P_{11}H^\dagger-E]\psi_t^{(2)}\rangle.\end{aligned} \tag{B3}$$

When the exact wavefunctions are used as arguments, the above reduces to its stationary value, which is

$$\begin{aligned}G'[\psi_{\mathscr{E}}^{(1)},\psi_{\mathscr{E}}^{(2)}]=-P_{11}\mathbf{u}_{\mathscr{E}}^{(1)\dagger\prime}(\hbar^2/2\mathrm{m})\boldsymbol{\zeta}^{(2)}-(1-P_{11})\boldsymbol{\zeta}^{(1)\dagger}\\ \times(\hbar^2/2\mathrm{m})\mathbf{u}_{\mathscr{E}}^{(2)\prime}.\end{aligned} \tag{B4}$$

Since the terms multiplying $(1-P_{11})$ and P_{11} are equal (for exact wavefunctions, as shown in Appendix C), an

equivalent result is

$$G'[\psi_e^{(1)},\psi_e^{(2)}]=-\zeta^{(1)\dagger}(\hbar^2/2m)\mathbf{u}_e^{(2)\prime}. \tag{B5}$$

To use the functional, we write (using the stationary property)

$$\zeta^{(1)\dagger}(\hbar^2/2m)\mathbf{u}_e^{(2)\prime}=-G'[\psi_t^{(1)},\psi_t^{(2)}]+O(\Delta\psi^{(1)}\times\Delta\psi^{(2)}) \tag{B6}$$

and drop the unknown (but second order) error term. Since the parameters $\zeta_c^{(1)}$ may be assigned any values, we have here an approximate method for calculating any one of the $u_{ce}^{(2)\prime}$.

Methods for calculating the various collision matrices may be developed from Eqs. (B1), (B3), and (B6), in a similar manner as in the main body of the paper. However, the present functional, Eq. (B3) and the earlier one, Eq. (15) or (21), are not interchangeable, as one is proportional to the $u_{ce}^{(\mu)\prime}$, and the other, to the $u_{ce}^{(\mu)}$.

It is interesting to note that if boundary conditions of the form of Eq. (B1) are used in G for one of the wavefunctions, the same form automatically applies to the other wavefunction, also. For example suppose Eq. (B1) is applied to Eq. (7) for $\psi_t^{(2)}$. Then $\mathbf{P}_1=0$, and $\mathbf{P}_4=P_{11}\times(\hbar^2/2m)$. Then, Eq. (7c) becomes

$$(P_{10}+P_{11})(\hbar^2/2m)\mathbf{u}_e^{(1)*}+\tilde{\mathbf{P}}_7=0. \tag{B7}$$

Rosenberg and Spruch[22] have proved that a class two variational calculation using the boundary condition (B1) yields a lower bound on the absolute phase shift, for scattering from a structureless center of force.

APPENDIX C

In this appendix are derived the necessary and sufficient conditions for the general functional, as given in Eq. (15), to be stationary. Also derived are the stationary form of G, and the second order error term in G.

The trial wave functions $\psi_t^{(1)}$ and $\psi_t^{(2)}$ may be written in the form

$$\psi_t^{(\mu)}=\Phi^{(\mu)}+\lambda^{(\mu)}\Delta^{(\mu)}, \tag{C1}$$

where the $\lambda^{(\mu)}$ are constants, and $\Phi^{(\mu)}$ and $\Delta^{(\mu)}$ are continuous functions, locally square integrable, and well enough behaved that the integrals appearing below all exist. Since G is bilinear in $\psi_t^{(1)*}$ and $\psi_t^{(2)}$, it follows that its dependence on $\lambda^{(1)}$ and $\lambda^{(2)}$ is

$$G[\psi_t^{(1)},\psi_t^{(2)}]=G_0+G_1\lambda^{(1)*}+G_1'\lambda^{(2)}+G_2\lambda^{(1)*}\lambda^{(2)}. \tag{C2}$$

We seek the conditions on $\Phi^{(1)}$, $\Phi^{(2)}$, $\lambda^{(1)}$, and $\lambda^{(2)}$, under which

$$\left(\frac{\partial G}{\partial\lambda^{(1)*}}\right)_{\lambda^{(2)}}=\left(\frac{\partial G}{\partial\lambda^{(2)}}\right)_{\lambda^{(1)}}=0 \quad \text{for all} \quad \Delta^{(1)},\ \Delta^{(2)}. \tag{C3}$$

These derivatives are given by

$$\left(\frac{\partial G}{\partial\lambda^{(1)*}}\right)_{\lambda^{(2)}}=G_1+G_2\lambda^{(2)} \tag{C4a}$$

and

$$\left(\frac{\partial G}{\partial\lambda^{(2)}}\right)_{\lambda^{(1)}}=G_1'+G_2\lambda^{(1)*}. \tag{C4b}$$

Now, of the different functions in Eq. (C1), G_1 depends only on $\Delta^{(1)}$ and $\Phi^{(2)}$, G_1' depends only on $\Delta^{(2)}$ and $\Phi^{(1)}$, while G_2 depends only on $\Delta^{(1)}$ and $\Delta^{(2)}$. Therefore, Eq. (C3) is satisfied if and only if both

$$\lambda^{(1)}=\lambda^{(2)}=0 \tag{C5a}$$

and

$$G_1=G_1'=0 \quad \text{for all } \Delta^{(1)},\ \Delta^{(2)}. \tag{C5b}$$

We now find the necessary and sufficient conditions on the $\Phi^{(\mu)}$ for Eq. (C5b) to be true.

Expand the functions $\Phi^{(\mu)}$ and $\Delta^{(\mu)}$ in a neighborhood of S as a sum of two-body terms and a remainder,

$$\Phi^{(\mu)}=\sum_c|c)r_c^{-1}u_{c\Phi}(r_c)+\mathcal{Q}\Phi^{(\mu)}, \tag{C6a}$$

$$\Delta^{(\mu)}=\sum_c|c)r_c^{-1}u_{c\Delta}(r_c)+\mathcal{Q}\Delta^{(\mu)}, \tag{C6b}$$

where $\mathcal{Q}$ is a projection operator, defined such that $\mathcal{Q}f$ and $\mathrm{grad}_n(\mathcal{Q}f)$ for any function f are orthogonal (in a surface integration S) to all the two-body channels. Define surface quantities

$$u_{c\Phi}^{(\mu)}=(c|r_c\Phi^{(\mu)})_{a_c} \tag{C7a}$$

and

$$u_{c\Phi}^{(\mu)\prime}=(c|\mathrm{grad}_n r_c\Phi^{(\mu)})_{a_c} \tag{C7b}$$

and similar quantities $u_{c\Delta}^{(\mu)}$ and $u_{c\Delta}^{(\mu)\prime}$, with Δ replacing Φ.

Next, expand the terms in G, Eq. (15), which contain integrals over $H-E$, using Green's theorem twice:

$$\begin{aligned}
&-(1-P_{11})\langle\psi_t^{(1)}|(H-E)\psi_t^{(2)}\rangle-P_{11}\langle(H-E)\psi_t^{(1)}|\psi_t^{(2)}\rangle\\
&=-(1-P_{11})\langle\Phi^{(1)}|(H-E)\Phi^{(2)}\rangle-P_{11}\langle(H-E)\Phi^{(1)}|\Phi^{(2)}\rangle\\
&\quad+\lambda^{(1)*}[-\langle\Delta^{(1)}|(H-E)\Phi^{(2)}\rangle+P_{11}\textstyle\int_S(\hbar^2r_\alpha^2/2m_\alpha)\\
&\quad\times(\Phi^{(2)}\mathrm{grad}_n\Delta^{(1)*}-\Delta^{(1)*}\mathrm{grad}_n\Phi^{(2)})d\omega]\\
&\quad+\lambda^{(2)}[-\langle(H-E)\Phi^{(1)}|\Delta^{(2)}\rangle+(1-P_{11})\textstyle\int_S(\hbar^2r_\alpha^2/2m_\alpha)\\
&\quad\times(\Phi^{(1)*}\mathrm{grad}_n\Delta^{(2)}-\Delta^{(2)}\mathrm{grad}_n\Phi^{(1)*})d\omega]\\
&\quad+\lambda^{(1)*}\lambda^{(2)}[-(1-P_{11})\langle\Delta^{(1)}|(H-E)\Delta^{(2)}\rangle\\
&\quad-P_{11}\langle(H-E)\Delta^{(1)}|\Delta^{(2)}\rangle],
\end{aligned} \tag{C8}$$

where the integral with respect to ω goes over all of S ($d\omega=r_\alpha^{-2}dS$). Using Eqs. (C6) and (C7), the first surface integral which appears in the last equation is

$$\begin{aligned}
&\sum_c(\hbar^2/2m_c)(u_{c\Delta}^{(1)*\prime}u_{c\Phi}^{(2)}-u_{c\Delta}^{(1)*}u_{c\Phi}^{(2)\prime})\\
&\quad+\textstyle\int_S(\hbar^2/2m_\alpha)[(\mathcal{Q}\Phi^{(2)})\mathrm{grad}_n\Delta^{(1)*}-\Delta^{(1)*}\mathrm{grad}_n(\mathcal{Q}\Phi^{(2)})]d\omega.
\end{aligned}$$

It now becomes a simple matter to write out the different terms in G. Thus, grouping the terms which are proportional to $\lambda^{(1)*}$, but not to $\lambda^{(2)}$, we find

$$\begin{aligned}
G_1=&\{\mathbf{u}_\Delta^{(1)\dagger\prime}\xi-\mathbf{u}_\Delta^{(1)\dagger}[\hbar^2/2m+\epsilon(2m/\hbar^2)\xi]\}\\
&\times[\mathbf{u}_\Phi^{(2)\prime}-(2m/\hbar^2)\epsilon\mathbf{u}_\Phi^{(2)}-\zeta^{(2)}]\\
&+P_{11}\textstyle\int_S(\hbar^2/2m_\alpha)[(\mathcal{Q}\Phi^{(2)})\mathrm{grad}_n\Delta^{(1)*}-\Delta^{(1)*}\\
&\times\mathrm{grad}_n(\mathcal{Q}\Phi^{(2)})]d\omega-\langle\Delta^{(1)}|(H-E)\Phi^{(2)}\rangle.
\end{aligned} \tag{C9}$$

The three terms above depend on different aspects of $\Delta^{(1)}$; i.e., the two-body part on S, the many-body part on S, and the function inside S, in that order. Clearly, for G_1 to be zero for all well-behaved $\Delta^{(1)}$, each of the three terms must be identically zero. In the first term, note that there is no combination of ξ and ϵ which makes the first factor identically zero. Therefore, we have

$$\mathbf{u}_\Phi^{(2)\prime}-(2m/\hbar^2)\epsilon\mathbf{u}_\Phi^{(2)}-\zeta^{(2)}=0. \tag{C10}$$

From the second term we have either

$$\mathcal{Q}\Phi^{(2)} = \mathrm{grad}_n(\mathcal{Q}\Phi^{(2)}) = 0 \quad \text{on } S \tag{C11a}$$

or

$$P_{11} = 0, \tag{C11b}$$

while, from the third term,

$$(H - E)\Phi^{(2)} = 0. \tag{C12}$$

Similar work with G_1' produces:

$$\mathbf{u}_\phi^{(1)\dagger\prime} - \mathbf{u}_\phi^{(1)\dagger}\epsilon 2m/\hbar^2 - \boldsymbol{\zeta}^{(1)\dagger} = 0; \tag{C13}$$

either

$$\mathcal{Q}\Phi^{(1)} = \mathrm{grad}_n(\mathcal{Q}\Phi^{(1)}) = 0 \quad \text{on } S \tag{C14a}$$

or

$$P_{11} = 1; \tag{C14b}$$

and

$$(H - E)\Phi^{(1)} = 0. \tag{C15}$$

Thus, it has been proven that Eqs. (C5a), (C10)–(C15) are, together, necessary and sufficient for Eq. (C3) to hold. This means that $G[\psi_t^{(1)}, \psi_t^{(2)}]$, Eq. (15), is stationary with respect to all continuous variations of $\psi_t^{(1)}$ and $\psi_t^{(2)}$ if and only if these functions satisfy the Schrödinger equation and the boundary conditions, Eqs. (C10) and (C13), and have no many-body components on S, except that the latter condition is absent with respect to $\psi_t^{(1)}$ or $\psi_t^{(2)}$ if P_{11} is one or zero, respectively. The existence and uniqueness of such wavefunctions is proven in Appendix A.

The stationary value of G, obtained by using the exact wavefunctions, $\psi_e^{(1)}$ and $\psi_e^{(2)}$, which satisfy the above conditions, in either G or G_0, is

$$G[\psi_e^{(1)}, \psi_e^{(2)}] = P_{11}\mathbf{u}_e^{(1)\dagger}(\hbar^2/2m)\boldsymbol{\zeta}^{(2)} + (1 - P_{11})\boldsymbol{\zeta}^{(1)\dagger} \times (\hbar^2/2m)\mathbf{u}_e^{(2)}, \tag{C16}$$

which can be simplified as follows. From Green's theorem comes the relation

$$\langle (H - E)\psi_e^{(1)} | \psi_e^{(2)} \rangle - \langle \psi_e^{(1)} | (H - E)\psi_e^{(2)} \rangle = -\mathbf{u}_e^{(1)\dagger\prime}(\hbar^2/2m)\mathbf{u}_e^{(2)} + \mathbf{u}_e^{(1)\dagger}(\hbar^2/2m)\mathbf{u}_e^{(2)\prime} \tag{C17}$$

(for which it is only necessary to assume that one of the wavefunctions has no many-body components on S). Using the boundary conditions, Eqs. (C10) and (C13), to eliminate the primed quantities in the right-hand side and seeing that the left-hand side is zero by the Schrödinger equation, we find

$$\mathbf{u}_e^{(1)\dagger}(\hbar^2/2m)\boldsymbol{\zeta}^{(2)} = \boldsymbol{\zeta}^{(1)\dagger}(\hbar^2/2m)\mathbf{u}_e^{(2)}. \tag{C18}$$

[This result may also be derived from Eq. (53).] Therefore,

$$G[\psi_e^{(1)}, \psi_e^{(2)}] = \mathbf{u}_e^{(1)\dagger}(\hbar^2/2m)\boldsymbol{\zeta}^{(2)} = \boldsymbol{\zeta}^{(1)\dagger}(\hbar^2/2m)\mathbf{u}_e^{(2)}. \tag{C19}$$

Since the trial functions may always be written in the form

$$\psi_t^{(\mu)} = \psi_e^{(\mu)} + \lambda^{(\mu)}\Delta^{(\mu)}, \tag{C20}$$

in which $\Delta^{(1)}$ and $\Delta^{(2)}$ are ordinarily unknown, G is, formally,

$$G[\psi_t^{(1)}, \psi_t^{(2)}] = G[\psi_e^{(1)}, \psi_e^{(2)}] + \lambda^{(1)*}\lambda^{(2)}G_2[\Delta^{(1)}, \Delta^{(2)}], \tag{C21}$$

so, the error term in G is second order and proportional to G_2. For possible future reference,

$$G_2 = -P_{11}\{\langle (H - E)\Delta^{(1)} | \Delta^{(2)} \rangle + (\mathbf{u}_\Delta^{(1)\dagger\prime} - \mathbf{u}_\Delta^{(1)\dagger}\epsilon 2m/\hbar^2) \times (\hbar^2/2m)\mathbf{u}_\Delta^{(2)}\} - (1 - P_{11})\{\langle \Delta^{(1)} | (H - E)\Delta^{(2)} \rangle + \mathbf{u}_\Delta^{(1)\dagger\prime}(\hbar^2/2m)[\mathbf{u}_\Delta^{(2)\prime} - (2m/\hbar^2)\epsilon\mathbf{u}_\Delta^{(2)}]\} + (\mathbf{u}_\Delta^{(1)\dagger\prime} - \mathbf{u}_\Delta^{(1)\dagger}\epsilon 2m/\hbar^2)\xi(\mathbf{u}_\Delta^{(2)\prime} - (2m/\hbar^2)\epsilon\mathbf{u}_\Delta^{(2)}). \tag{C22}$$

APPENDIX D

In this appendix, we show that Eq. (67) is derivable from a matrix approximation due to Schwartz.

Any eigenfunction $|\psi\rangle$ of $H - E$ satisfies

$$(H + L - E)|\psi\rangle = L|\psi\rangle \tag{D1}$$

and so

$$|\psi\rangle = (H + L - E)^{-1}|L\psi\rangle \tag{D2}$$

(which, incidentally, demonstrates that $H + L - E$ is the inverse of a Green's function[35,65,66]). When expanded in two finite basis sets[67] which are separately orthonormal over the inside of S, the latter equation becomes[35]

$$|\psi\rangle \simeq \sum_{ij} |\chi_i^{(2)}\rangle\langle\chi_i^{(2)}|\mathcal{A}^{-1}|\chi_j^{(1)}\rangle\langle\chi_j^{(1)}|L\psi\rangle. \tag{D3}$$

To avoid a difficult calculation, replace in the above equation the matrix of the inverse of $\mathcal{A}$ by the inverse of the (finite rank) matrix of $\mathcal{A}$. This is an example of a more general matrix approximation which Schwartz[68] has shown to have errors of second order, only, under certain conditions. The result is

$$|\psi\rangle \simeq \sum_{ij} |\chi_i^{(2)}\rangle (\mathbf{A}^{-1})_{ij}\langle\chi_j^{(1)}|L\psi\rangle. \tag{D4}$$

Since ψ contains, on S, two-body components only, it follows from Eq. (53) that $\mathcal{A}$ may be replaced by $\bar{\mathcal{A}}$ in Eq. (D1), and, therefore, in Eq. (D4), which then becomes Eq. (67), when evaluated on S.

The right-hand side of the last equation is easily seen to be invariant with respect to nonsingular linear transformations of the bases, so the restriction of orthonormality may be removed.

*Research sponsored by the Division of Physical Research of the U.S. Energy Research and Development Administration under contract with Union Carbide Corporation. A large portion of this work was performed at Barnard College, where it was supported by a Cottrell College Science Grant from Research Corporation.

[1]F.E. Harris and H.H. Michels, Meths. Comp. Phys. **10**, 143 (1971).
[2]L.M. Delves, Adv. Nucl. Phys. **5**, 1 (1972); M.R.H. Rudge, J. Phys. B: Atom. Molec. Phys. **6**, 1788 (1973).
[3]B.L. Moiseiwitsch, *Variational Principles* (Interscience, London, 1966); Yu.N. Demkov, *Variational Principles in the Theory of Collisions* (Macmillan, New York, 1963).
[4]D.G. Truhlar, J. Abdallah, Jr., and R.L. Smith, Adv. Chem. Phys. **25**, 211 (1974).
[5]E.M. Mortensen, Adv. Chem. Phys. **21**, 127 (1971).
[6]O.H. Crawford, J. Chem. Phys. **55**, 2563, 2571 (1971).
[7]P. McCavert and M.R.H. Rudge, J. Phys. B: Atom. Molec. Phys. **3**, 1286 (1970).
[8]A.L. Sinfailam and R.K. Nesbet, Phys. Rev. A **6**, 2118 (1972).
[9]J.D. Lyons, R.K. Nesbet, C.C. Rankin, and A.C. Yates, J. Comput. Phys. **13**, 229 (1973).
[10]L.D. Thomas, R.S. Oberoi, and R.K. Nesbet, Phys. Rev. A **10**, 1605 (1974).

[11]R.K. Nesbet, Phys. Rev. A 12, 444 (1975).
[12]L.D. Thomas and R.K. Nesbet, Phys. Rev. A 12, 2369, 2378 (1975).
[13]R.I. Campeanu and J.W. Humberston, J. Phys. B: Atom. Molec. Phys. 8, 1244 (1975).
[14]D. Kracht and J.J. Chang, J. Phys. B: Atom. Molec. Phys. 8, L29 (1976).
[15]D.G. Truhlar, to be published in Modern Theoretical Chemistry, edited by G.A. Segal (Plenum, New York).
[16]P.G. Burke and W.D. Robb, Adv. Atomic Molec. Phys. 11, 143 (1975).
[17]R.A. Chatwin and J.E. Purcell, J. Math. Phys. 12, 2024 (1971).
[18]I.A. Sloan and T.J. Brady, Phys. Rev. C 6, 701 (1972).
[19]A. Mori, Prog. Theoret. Phys. 49, 1960 (1973).
[20]R.J. Philpott and J. George, Nucl. Phys. A 233, 164 (1974). This article includes a review of earlier calculations.
[21]R.J. Philpott, Nucl. Phys. A 243, 260 (1975).
[22]L. Rosenberg and L. Spruch, Phys. Rev. 120, 474 (1960). See also Sec. 4 of the following reference.
[23]L. Rosenberg and L. Spruch, Phys. Rev. 125, 1407 (1962).
[24]Y. Hahn, T.F. O'Malley, and L. Spruch, Phys. Rev. 130, 381 (1963); 134, B911 (1964).
[25]R. Sugar and R. Blankenbecler, Phys. Rev. 136, B472 (1964).
[26]Y. Hahn and L. Spruch, Phys. Rev. 153, 1159 (1967).
[27]See, for example, Y. Hahn, Phys. Rev. A 4, 1881 (1971); H.J. Kolker, J. Chem. Phys. 58, 2288 (1973); D.G. Truhlar and R.L. Smith, Phys. Rev. A 6, 233 (1972), and the references therein.
[28]J. Nuttall, Ann. Phys. (N.Y.) 52, 428 (1969).
[29]W. Kohn, Phys. Rev. 74, 1763 (1948).
[30]C. Schwartz, Ann. Phys. (N.Y.) 16, 36 (1961).
[31]K. R. Brownstein and W.A. McKinley, Phys.Rev. 170, 1255 (1968).
[32]R.K. Nesbet, Phys. Rev. 175, 134 (1968); 179, 60 (1969).
[33]For simplicity, in the remainder of the paper, we will refer only to nuclei, although the theory applies to atomic and molecular scattering as well.
[34]A.M. Lane and R.G. Thomas, Rev. Mod. Phys. 30, 257 (1958).
[35]C. Bloch, Nucl. Phys. 4, 503 (1957).
[36]For examples of stationary functionals which are not bilinear or quadratic, and methods for deriving an infinite number of them, see M. Moe and D.S. Saxon, Phys. Rev. 111, 950 (1958), and M. Moe, thesis (University of California, Los Angeles, 1957).
[37]For a different approach, in which the T matrix is varied, instead of wavefunctions, see H. Rabitz and R. Conn, Phys. Rev. A 7, 577 (1973).
[38]An operator which approximates $|H-E|^2$ is sometimes used in refining trial wavefunctions, but not in the stationary functional for scattering parameters. See Sec. IIID.
[39]The right-hand sides of Eqs. (3c) and (3d) are integrals over solid angle, ω, evaluated at $r_c = a_c$. Strictly speaking, they are surface integrals multiplied by a_c^2, because of the definition of $d\omega$, which follows Eq. (2).
[40]As discussed by Kohn,[29] scattering wavefunctions are highly degenerate, and an inhomogenous boundary condition serves to single out (and establish the normalization of) a single wavefunction, from the degenerate set. Boundary conditions are studied in Appendices A and B.
[41]A.M. Lane and D. Robson, Phys. Rev. 185, 1403 (1969).
[42]Another general method for deriving stationary functionals, suitable for a wide range of variational problems, has been described by E. Gerjuoy, A.R.P. Rau, and L. Spruch, Phys. Rev. A 8, 662 (1973).
[43]Nondiagonal boundary conditions of the form of Eq. (16) are used by D.V. Zvijac, E.J. Heller, and J.C. Light, J. Phys. B: Atom. Molec. Phys. 8, 1016 (1975).
[44]E.P. Wigner and L. Eisenbud, Phys. Rev. 72, 29 (1947).
[45]T. Kato, Phys. Rev. 80, 475 (1950).
[46]A.M. Lane and D. Robson, Phys. Rev. 178, 1715 (1969). For a correction, see R.A. Chatwin, Phys. Rev. C 2, 1167 (1970).
[47]The Wigner-Eisenbud R-matrix method has the virtues that the E_λ and $\gamma_{c\lambda}$ often have physical meaning, and that, once the initial computation is done, the energy variation of cross sections may be found by manipulating matrices whose rank equals the number of channels only (instead of the number of basis functions). The other methods discussed in this work do not have these advantages (at least in the form in which they are presented). However, Lane and Robson's theory [Eq. (62b) with a single basis] may be reduced to standard R-matrix form [Eq. (78)] by a congruent transformation which simultaneously diagonalizes the matrix of $H + L$ and transforms the overlap matrix to the unit matrix.[48]
[48]R.J. Philpott, Nucl. Phys. A 243, 260 (1975).
[49]H.E. Saraph, M.J. Seaton, and J. Shemming, Proc. Phys. Soc. 89, 27 (1966).
[50]R.K. Nesbet and R.S. Oberoi, Phys. Rev. A 6, 1855 (1972).
[51]J.D. Lyons, R.K. Nesbet, C.C. Rankin, and A.C. Yates, J. Comp. Phys. 13, 229 (1973).
[52]J.L. Jackson, Phys.Rev. 83, 301 (1951).
[53]The rate of convergence of the R-matrix method is considerably improved by using the natural boundary conditions. However, their values are energy-dependent, and are not known in advance, except for the closed channels. They must be determined by an iterative (or similar) procedure at each energy of interest. See S.S. Ahmad, R.F. Barrett, and B.A. Robson, Nucl. Phys. A 257, 378 (1976).
[54]F.E. Harris and H.H. Michels, Phys. Rev. Lett. 22, 1036 (1969).
[55]I. Wladawsky, J. Chem. Phys. 58, 1826 (1973).
[56]J. Callaway and J.W. Wooten, Phys. Rev. A 9, 1924 (1974).
[57]J. Abdallah, Jr. and D.G. Truhlar, J. Chem. Phys. 60, 4670 (1974).
[58]E.W. Schmid and K.H. Hoffmann, Nucl. Phys. A 175, 443 (1971).
[59]E.W. Schmid and J. Schwager, Nucl. Phys. A 180, 434 (1972).
[60]J. Schwager and E.W. Schmid, Nucl. Phys. A 205, 168 (1973).
[61]E.W. Schmid, Nuovo Cimento A 18, 771 (1973); J. Schwager, Nuovo Cimento A 18, 787 (1973).
[62]R.J. Philpott and J. George, Nucl. Phys. A 233, 164 (1974).
[63]A. Mori, Phys. Rev. C 5, 1795 (1972).
[64]O.H. Crawford, unpublished results.
[65]A.M. Lane and D. Robson, Phys. Rev. 151, 774 (1966).
[66]B.A. Robson and D. Robson, Phys. Lett. B 25, 504 (1967).
[67]That A may be realized in a basis set in coordinate space was proven by Lane and Robson. It follows from the fact that $H + L(0)$ is Hermitian (in the absence of many-body channels), which is why Bloch was led to consider the L operator in the first place.[35,41,65]
[68]C. Schwartz, J. Comput. Phys. 2, 90 (1967).

Invariants for the time-dependent harmonic oscillator*

Jack W. Macki

Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada
(Received 29 November 1976)

Lewis showed that in the case $p(t)\equiv 1$, $h=1$,

$$I(t)=(1/2)\{p^2(t)[\rho(t)y'(t)-y(t)\rho'(t)]^2+h^2y^2(t)/\rho^2(t)\}$$

is constant in time if $y(t)$ solves $(p(t)y')'+q(t)y=0$, and $\rho(t)$ solves $p(t)\rho^3(t)L[\rho]=h^2$ (h constant). Recently, Eliezer and Gray showed that $I(t)=\text{const}$ is just the conservation of angular momentum in an appropriate physical interpretation. We show, using a change of variable technique, that $I(t)=\text{const}$ reduces to $\sin^2\theta+\cos^2\theta=1$. We discuss uniqueness and extendability of solutions to the above equation in ρ.

In this paper we investigate the relation between the differential equations

$$p(t)\rho^3(t)L[\rho]=h^2, \tag{1}$$

$$L[y]\equiv[p(t)y']'+q(t)y=0, \tag{2}$$

and the invariant

$$I(t)=\tfrac{1}{2}[p^2(t)[\rho(t)y'(t)-y(t)\rho'(t)]^2+h^2y^2(t)/\rho^2(t)\}.$$

Here and throughout, h is constant, and we assume $p(t)>0$ with both p and q continuous on some given t-interval J. For the case $p(t)\equiv 1$, $h=1$, Lewis[1] (cf. also Lewis and Riesenfeld[2]) discovered that $I(t)$ is constant whenever ρ, y solve (1) and (2), respectively, thus each solution ρ of (1) generates an invariant for the motion of the corresponding adiabatic oscillator. For the case $p(t)\equiv 1$, Eliezer and Gray[3] gave a kinematic interpretation of this relationship, and extensions of their work are given in recent papers of Leach[4] and Günther and Leach.[5]

We shall show that the relationship between (1), (2) and $I(t)$ can be interpreted in terms of a change of variables technique described in a paper by the author.[6] This point of view allows us to present another invariant for the adiabatic oscillator (cf. Theorem 4). We also discuss uniqueness and extendability of real-valued solutions of (1) (for a discussion of complex-valued solutions, see Lewis[7]), and relate this to the disconjugacy of the operator L. L is disconjugate on J if and only if each solution of (2) has at most one zero on J.

As is well known,[6] if we define

$$s=\int_a^t\gamma(\xi)\,d\xi,\quad u(s)=y(t)/\beta(t)\quad(a\in J,\ t\in J), \tag{3}$$

with $\gamma(t)>0$, $\beta(t)>0$, $\gamma\in C(J)$, $\beta\in C^2(J)$, then $y(t)$ solves (2) on J if and only if $u(s)$ defined by (3) solves

$$\frac{d}{ds}\left(\gamma(t)\beta^2(t)p(t)\frac{du}{ds}\right)+\{\beta(t)L[\beta]/\gamma(t)\}u(s)=0,$$

on the corresponding s-interval. If for a given β, we *define* $\gamma(t)=1/\beta^2p$, then the equation for $u(s)$ is $(d^2u/ds^2)+p\beta^3L[\beta]u(s)=0$. Note that $p\beta^3L[\beta]$ is based entirely on functions given originally as functions of $t=t(s)$, and that the coefficient of $u(s)$ is just the left side of (1).

Theorem 1 (*Lewis*): If $\rho(t)$ solves (1) and $y(t)$ solves (2), then $I(t)$ is a constant for $t\in J$.

Proof: We use the change of variable (3) with $\beta=\rho(t)$, $\gamma=1/p\rho^2$, which maps Eq. (2) into the equation $d^2u/ds^2+h^2u=0$; the general solution of the latter is $u(s)=A\sin(hs+B)$. Since $p(t)\rho^2(t)\,ds=dt$, it is easy to see that

$$I(t)=\frac{1}{2}\left[p^2\left(\frac{d(\rho u)}{dt}\rho-\rho u\frac{d\rho}{dt}\right)^2+h^2\rho^2u^2/\rho^2\right]$$

$$=\frac{1}{2}\left[p^2\rho^4\left(\frac{du}{dt}\right)^2+h^2u^2\right]=\frac{1}{2}\left[\left(\frac{du}{ds}\right)^2+h^2u^2\right]$$

$$=\tfrac{1}{2}h^2.$$ QED

Lewis also pointed out that if y_1, y_2 solve (2) with Wronskian $W(y_1,y_2)=c/p(t)$, c constant, then the general solution of

$$p(t)\rho^3L[\rho]=\lambda \tag{4}$$

is

$$\rho(t)=\pm[Ay_1^2+2By_1y_2+Cy_2^2]^{1/2},\quad B^2-AC=-\lambda/W^2p^2, \tag{5}$$

where A, B, C are constants. This fact was also pointed out earlier by Pinney[8] [they treat only the case $p(t)\equiv 1$, but the extension is immediate]. Of course, this solution may take on complex values.

The case $\lambda=0$ is uninteresting, since in this case ρ solves (1) if and only if ρ solves (2) on J. This can be argued as follows. Suppose $\rho(t)\in C^2(J)$ solves (4) with $\lambda=0$, and assume that $L[\rho]\neq 0$ at some point $t_0\in J$. Then $L[\rho]\neq 0$ in a neighborhood of t_0 (a one-sided neighborhood if t_0 is an endpoint), and we must have $\rho(t)\equiv 0$ on this neighborhood, in order that (1) be satisfied. But then ρ' and ρ'' are identically zero on the (relative) interior of this neighborhood (including t_0 as well), which implies $L[\rho]=0$ at t_0. Thus $L[\rho]\neq 0$ at t_0 leads to a contradiction.

If $\lambda\neq 0$, then for each $t_0\in J$ and any real initial conditions $\rho(t_0)=a$, $\rho'(t_0)=b$, a unique real solution of the corresponding initial value problem for (4) exists on *some* interval containing t_0. For $\rho_0\neq 0$, this follows immediately from the standard existence theory for ordinary differential equations, writing (4) as $L[\rho]=\lambda/p(t)\rho^3(t)$ and noting that the right side satisfies a Lipschitz condition at (t_0,ρ_0). When $\rho_0=0$, we must have $\lambda=0$, and the preceding paragraph applies.

Lemma: There exists a *local real-valued* solution of any initial value problem for (4), for any real λ. This solution is extendable as long as $\rho(t)\neq 0$.

All but the extendability assertion has been proved in the preceding discussion. To settle the extendability

question, let $\tilde{\rho}(t)$ be a local solution of (1), existing and nonzero in a neighborhood of $t_0 \in J$. Now choose the (arbitrary) independent solutions $y_1(t)$, $y_2(t)$ of (2) that occur in (5) so that the function $\rho(t)$ defined by (5) satisfies $\rho(t_0) = \tilde{\rho}(t_0)$, $\rho'(t_0) = \tilde{\rho}'(t_0)$ (an easy computation shows that this is always possible). Then by uniqueness, $\rho(t) = \tilde{\rho}(t)$ on their common interval of existence. Clearly $\tilde{\rho}(t)$ must extend to J if $\rho(t)$ does. The only way that $\rho(t)$ could fail to extend would be for the quadratic form under the radical in (5), $Q(t) = Ay_1^2 + 2By_1y_2 + Cy_2^2$, to become negative. But $\tilde{\rho}(t)$ is never zero, hence $\rho(t) = \tilde{\rho}(t)$ never vanishes, hence they both extend. On the other hand, if $\tilde{\rho}(t_1) = 0$ for some $t_1 \in J$, then of course $Q(t_1) = 0$. Then $\rho(t)$ (hence $\tilde{\rho}$) cannot be extended beyond t_1, because $Q(t)$ changes sign when it vanishes. Suppose (say) that $y_1(t_1) \neq 0$, then $Q(t) = y_1^2(t)[A + 2Bx(t) + Cx^2(t)]$ near t_1, where $x(t) = y_2(t)/y_1(t)$. Now $x'(t) = W(y_1, y_2)/y_1^2 \neq 0$, and since $B^2 - AC \neq 0$ [from (5)], the form $A + 2Bx + Cx^2 = R(x)$ changes sign as x increases (or decreases) through the zero $x_1 = x(t_1)$. Thus $Q(t)$ changes sign at t_1.

An example of the extendability problem for (4) is given by the equation $L[y] = y'' + y = 0$ on $[0, \infty)$, with solutions $y_1(t) = \cos t$, $y_2(t) = \sin t$. If $\lambda = -1$ in (4), then the form (5) is just $A\cos^2 t + 2B\cos t \sin t + C\sin t$, with $B^2 - AC = 1$. A particular case is $A = 1$, $C = -1$, $B = 0$, which gives $\rho(t) = [\cos 2t]^{1/2}$, valid on (for example) $[0, \pi/4)$. One can continue $\rho(t)$ as $[|\cos 2t|]^{1/2}$, but it is nondifferentiable at odd multiples of $\pi/4$, so it is not a solution of (4) on $J = [0, \infty)$. Note that if we use $\beta = \rho = [|\cos 2t|]^{1/2}$ in the change of variable (3), then $s = \int_0^t (1/\rho^2)\, d\xi$ so the t-interval $[0, \pi/4)$ maps into the s-interval $[0, \infty)$.

The reason for this extendability problem for (4) can be indicated by examining the above example more closely. In that example, we had (*) $u(s) = \rho(t)y(t)$, where $y(t)$ solved $y'' + y = 0$ and $u(s)$ solved $d^2u/ds^2 - u(s) = 0$. Thus $y(t)$ vanished infinitely often on $[0, \infty)$, while $u(s)$ could vanish at most once, so (*) could make sense at only one (at most) zero of $y(t)$ [$\rho(t)$ cannot vanish for $\lambda \neq 0$]. The extendability problem for (4) is of interest because of the following result.

Theorem 2: Equation (2), $t \in J$, can be reduced by a change of variable of the form (3) to the equation $d^2u/ds^2 + \lambda u(s) = 0$, with λ constant, if and only if there exists a positive solution $\rho(t)$ of (4) that extends to J.

Proof: If $\rho(t) > 0$ solves (4) on J, then we use (3) with $\beta = \rho$, $\gamma = 1/p\rho^2$, and the conclusion follows.

If the equation $L[y] = 0$ is mapped into $d^2u/ds^2 + \lambda u = 0$ for $t \in J$ by the change of variable (3), then $\gamma\beta^2 p \equiv 1$, which implies $\beta(t) \neq 0$ on J [without loss of generality $\beta(t) > 0$ on J]. Finally $p(t)\beta^3 L[\beta] = \lambda$ for $t \in J$, so $\beta(t)$ is the desired solution of (4).

Before stating a theorem that partially resolves the extendability problem, we need some preparatory remarks and a lemma. Note that if $\rho_+(t)$, $\rho_-(t)$ solve, respectively, (4) with $\lambda = +1$, $\lambda = -1$, then $\rho(t) = h\rho_+(t)$, $\rho(t) = h\rho_-(t)$ (h constant), respectively, solve (4) with $\lambda = +h^4$, $\lambda = -h^4$. Therefore, we need only discuss (4) with $\lambda = \pm 1$, which we label (4+), (4−), respectively.

Lemma: The general solution of (2) may be written

$$y(t) = K[p(t)\alpha'(t)]^{-1/2} \sin[\alpha(t) + \alpha_0], \quad t \in J, \tag{6}$$

where K and α_0 are constants, $\alpha(t) = \arg(y_1 + iy_2)$ for any (fixed) linearly independent solutions $y_1(t)$, $y_2(t)$ of $(p, q)(t)$ satisfying $W(y_1, y_2) > 0$. Furthermore, $\alpha \in C^3(J)$, $\alpha'(t) > 0$ on J.

Proof: According to Boruvka,[9] the general solution of any equation of the form $d^2w/ds^2 + Q(s)W(s) = 0$ on an interval $\mathcal{J}$ is $w(s) = K[da/ds]^{-1/2} \sin[a(s) + a_0]$, with K, a_0 constants, $a(s) = \arg(u + iv)$ where $u(s)$, $v(s)$ solve this same equation with $W(u, v) > 0$ on $\mathcal{J}$; also $a(s) \in C^3(\mathcal{J})$, $da/ds > 0$ on $\mathcal{J}$.

We now choose any fixed positive differentiable $\beta(t)$, $t \in J$, and define $\gamma(t) = 1/p\beta^2$, thus transforming (2) into a new equation of the form given above. By the above quoted result from Boruvka, the general solution of (2) is

$$y(t) = \beta(t)u(s) = \beta(t)K[da/ds]^{-1/2} \sin[a(s) + a_0],$$

where $a(s) = \arctan(v(s)/u(s)) = \arctan(y_2(t)/y_1(t))$, and $y_1(t) = \beta(t)u(s)$, $y_2(t) = \beta(t)v(s)$ are linearly independent solutions of (2) with $W(y_1, y_2) = \beta^2 W(u, v) > 0$. If we define $\alpha(t) = a(s) = \arctan(y_1(t)/y_2(t))$, and use the fact that $ds = (1/\beta^2 p)\, dt$, we have

$$\frac{d}{dt} = \frac{W(y_1, y_2)}{y_1^2 + y_2^2} = \frac{W(u, v)}{p^2(u^2 + v^2)} = \frac{1}{p^2}\frac{da}{ds},$$

which yields the desired result (6). The fact that $\alpha(t)$ is C^3 follows in a straightforward manner from the fact that $a(s)$ is C^3.

Theorem 3: (a) Every solution of (4+) extends to J.

(b) If all real-valued solutions of (4−) extend to J, then (2) is disconjugate on J^0, the interior of J.

Proof: (a) We first establish a subset of solutions of (4+), each of which extends to all of J. We will then show that every solution of (4+) lies in this subset. Let $y_1(t)$, $y_2(t)$ be any independent solutions of (2) [defined on all of J, since (2) is linear]. We define

$$\alpha(t) = \arg(y_1 + iy_2), \quad \beta(t) = [p(t)\, d\alpha/dt]^{-1/2},$$

$$s = \int_{t_0}^{t} (1/p\beta^2)\, d\xi = \alpha(t) - \alpha(t_0) \quad \text{for some } t_0 \in J.$$

Applying the change of variables (3) we see from the lemma above that $u(s) = y(t)/\beta(t) = K\sin(s + s_0)$, and this family solves $d^2u/ds^2 + p(t)\beta^3(t)L[\beta]u(s) = 0$. This two-parameter family corresponds uniquely to the second order linear equation $d^2u/ds^2 + u(s) = 0$, therefore, $p(t)\beta^3 L[\beta] = 1$ on J, that is $\beta(t)$ is an extendable solution of (4+).

Now suppose that $\rho(t)$ is any (local) solution of (4+), existing in some neighborhood of a point $t_0 \in J$. It is easy to show, as mentioned before, that we can choose solutions y_1, y_2 of (2) in such a way [by specifying $y_1(t_0)$, $y_1'(t_0)$, $y_2(t_0)$, $y_2'(t_0)$] that $\beta(t)$ as defined in the paragraph above satisfies $\beta(t_0) = \rho(t_0)$, $\beta'(t_0) = \rho'(t_0)$. By uniqueness, $\beta(t) \equiv \rho(t)$ on their common interval of existence. But a real solution $\rho(t)$ of (4+) can fail to extend only by having $\rho(t)$ vanish. This would make $\beta(t)$ vanish, a contradiction.

(b) We prove the contrapositive. Suppose (2) is not disconjugate on J^0. Then there is a solution $y_0(t)$ with two consecutive zeros $t_1 < t_2$ in J^0. Now $y_0(t)$ can be written in the form (6), and since $y_0(t_1) = y_0(t_2) = 0$, we have $\alpha(t_2) - \alpha(t_1) = \pi$ [recall $\alpha'(t) > 0$]. Now suppose $\rho(t)$ solves (4–) near t_1, with $\rho(t_1) > 0$. The change of variable (3), with $y(t) = \rho(t)u(s)$, $s = \int_{t_1}^{t} (1/p\rho^2)\, d\xi$ reduces (2) to $d^2u/ds^2 - u(s) = 0$ near t_1. Since $y_0(t_1) = 0$, and $\rho(t_1) > 0$, the corresponding function $u_0(s)$ must satisfy $u_0(0) = 0$, so $u_0(s) = A \sinh s$. Since $u_0(s)$ never vanishes again, yet $y_0(t_2) = 0$, we conclude that $\rho(t)$ must approach zero as $t \uparrow t_2$, so $\rho(t)$ does not extend to J.

Theorem 4: Let $p(t) > 0$, $q(t)$ be continuous for $t \in J$, and suppose that $\rho(t)$ solves $p(t)\rho^3 L[\rho] = -k^2$ on J. Then the function

$$H(t) = \tfrac{1}{2}[(k^2y^2/\rho^2) - p^2(\rho y' - y\rho')^2]$$

is constant in time for each solution $y(t)$ of (2).

The *proof* is an exact analog of Theorem 2, using the identity $\cosh^2 Z - \sinh^2 Z = 1$ in place of $\sin^2\theta + \cos^2\theta = 1$. Note that Theorem 4 is the same as the invariant of Lewis with $k = ih$.

Finally, we remark that Theorem 3(b) has a partial converse: If (2) is disconjugate on J, then there exists a solution of (1) that extends to the interior of J. To prove this, we let $y_0(t)$ be a solution on J, positive on the interior of J (such a solution exists for any disconjugate equation). If ρ is any local solution of (4–), then we can use $\beta = \rho$ in the change of variable (3) to transform (2) into $d^2u/ds^2 - u(s) = 0$, and the correspondence between solutions of (2) and (1) is $y(t) = \rho(t)u(s)$. In particular, using $y_0(t)$ in this relation, we see that $\rho(t)$ can never vanish on the interior of J, and ρ is thus extendable to the interior. There are a great many concrete criteria for disconjugacy, and if one of these is satisfied, then Theorem 4 applies.

ACKNOWLEDGMENTS

The author expresses his appreciation to the referee for many helpful suggestions, and to Professor Leach and Professor Günther for providing preprints of their work. Research for this paper was supported, in part, by the National Research Council of Canada under Grant number NRC A-3053.

*Research supported, in part by the National Research Council of Canada under grant NRC A-3053.

[1]H.R. Lewis, Jr., J. Math. Phys. **9**, 1976–86 (1968).
[2]H.R. Lewis Jr. and W.B. Riesenfeld, J. Math. Phys. **10**, 1458–73 (1969).
[3]C.J. Eliezer and A. Gray, SIAM J. Appl. Math. **30**, 463–8 (1976).
[4]P.G.L. Leach, J. Math. Phys. "On the theory of time-dependent linear canonical transformations as applied to Hamiltonians of the harmonic oscillator type," **18** (1977) (to be published).
[5]N.J. Günther and P.G.L. Leach, J. Math. Phys. **18**, 572 (1977).
[6]J.W. Macki, SIAM Rev. **18**, 269–74 (1976).
[7]H.R. Lewis, Jr., Phys. Rev. **172**, 1313 (1968).
[8]E. Pinney, Proc. A.M.S. **1**, 681 (1950).
[9]O. Boruvka, *Linear Differential Transformations of the Second Order*, transl. by F.M. Arscott (The English Universities Press, London, 1971), pp. 38–43.

Localized solutions of a nonlinear electromagnetic field*

Luis Vazquez†

Department of Mathematics, Brown University, Providence, Rhode Island 02912
(Received 16 June 1976)

The electronlike localized solutions, for a nonlinear electromagnetic field obtained from the Born–Infeld Lagrangian, are studied.

I. INTRODUCTION

About sixty years ago Mie[1] studied a nonlinear electromagnetic field in order to explain the structure of the charges with the very idea of the electromagnetic field. Thus they are regions of space–time where the field is very concentrated and the electromagnetic energy and momentum can be made finite. This attempt breaks down because Mie's field equations have the unacceptable property that their solutions depend on the absolute value of the potentials. Later Born and Infeld[2] proposed a nonlinear electromagnetic field, gauge invariant, postulating that there exists an absolute field b which is the natural unit for all field components and the upper limit of a purely electric field. In particular they studied the static and spherically symmetric solutions without a magnetic field.

We present in this paper the study of some solutions for the nonlinear electromagnetic field which is obtained using an expansion of the Born–Infeld Lagrangian. If we consider the two first terms in the expansion, for a suitable coupling constant, we find static localized solutions with a purely electric field. We mainly study the electronlike localized solutions corresponding to the first term in the above expansion.

The localized solutions are without radiation, bounded in the whole space and such that the physical quantities—energy, charge, spin and magnetic moment—associated with the field are finite. These quantities must be calculated as volume integrals of the corresponding densities which, in general, depend on the fields and their first derivatives. Thus for the electronlike solutions we demand that the potential A_μ belongs to $L^\infty(\mathbb{R}^3)$ and its first derivatives to $L^2(\mathbb{R}^3)$ and also it must be a function of class $C^2(\mathbb{R}^3)$. For the static localized solutions with a purely electric field we find it to be piecewise $C^2(\mathbb{R}^3)$.

In Sec. II we give the general description of the model while in Sec. III we study some properties of the electronlike localized solutions. In Sec. IV we present the numerical results, and its significance is discussed in Sec. V.

II. GENERAL DESCRIPTION OF THE MODEL

The Born–Infeld Lagrangian is

$$\mathcal{L}_{BI}=\left\{1+\frac{1}{2b^2}F_{\mu\nu}F^{\mu\nu}-\frac{1}{16b^4}(\tilde{F}_{\mu\nu}F^{\mu\nu})^2\right\}^{1/2}-1\,, \tag{1}$$

where

$$F_{\mu\nu}=\partial_\nu A_\mu-\partial_\mu A_\nu,\quad \tilde{F}_{\mu\nu}=\tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}F^{\alpha\beta}A^\mu=(A^0,\mathbf{A}),$$

and b is a constant.

Expanding $\mathcal{L}_{BI}$ and considering the two first terms, we get, apart from a factor, the following Lagrangian:

$$\mathcal{L}=-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}+(\beta/64)(\tilde{F}_{\mu\nu}F^{\mu\nu})^2+(\beta/64)\{F_{\mu\nu}F^{\mu\nu}-$$
$$-(\beta/16)(\tilde{F}_{\mu\nu}F^{\mu\nu})^2\}, \tag{2}$$

where $\beta=2/b^2$.

A. Localized solutions with a purely electric field

We consider in (2) the electrostatic case where the magnetic field is zero ($\tilde{F}_{\mu\nu}F^{\mu\nu}=0$). Then the corresponding field equation is

$$\partial_\nu F^{\mu\nu}=(\beta/8)\partial_\nu[(F_{\alpha\beta}F^{\alpha\beta})F^{\mu\nu}]. \tag{3}$$

We solve this equation for the case of central symmetry

$$A_\mu=(A(r),0,0,0), \tag{4}$$

where $F_{\alpha\beta}F^{\alpha\beta}=-2E_r^2$ and $E_r=-dA/dr\equiv A'$. Then (3) is simply

$$A''+\frac{2}{r}A'=-\frac{\beta}{2}\frac{A'^3}{r}-\frac{\beta}{4}\frac{d}{dr}(A'^3). \tag{5}$$

We can obtain the same equation (5) if we substitute A_μ, given by (4), in the Lagrangian (2), integrate over the angles, and make variations of the radial function $A(r)$. This method was advocated by Finkelstein, *et al.*[3] and it will be used in Sec. IIB.

The electric induction is given by[2,4]

$$D_r=\frac{\partial\mathcal{L}}{\partial E_r}=-\left(A'+\frac{\beta}{4}A'^3\right). \tag{6}$$

If we consider $\beta<0$, then $D_r=-A'[1-(|\beta|/4)A'^2]$, and we can find a nontrivial localized solution for Eq. (5) in the following way:

$$A'=(2\epsilon/\sqrt{|\beta|})\,\theta(R-r), \tag{7a}$$

$$A=\begin{cases}(2\epsilon/\sqrt{|\beta|})\,r, & r\le R,\\ (2\epsilon/\sqrt{|\beta|})R, & r>R,\end{cases} \tag{7b}$$

where $\epsilon=\pm1$, R is arbitrary, and θ is the step function. A is a function of class C^0 and piecewise C^2.

From Eqs. (6) and (7) we obtain $D_r=0$ and so the associated charge is $Q=0$.[5] Following Born and Infeld, we can interpret the above solution corresponding to a static distribution of the space charge which is the source of the **E** field. It consists in taking $\mathrm{div}\mathbf{E}=\rho$ (we use Heaviside–Lorentz rationalized units) as definition of charge density ρ. So from (7) and the second member of (5) we get

$$\rho=\frac{|\beta|}{2}\left(\frac{A'^3}{r}+\frac{1}{2}\frac{d}{dr}(A'^3)\right),$$
$$\rho=\frac{2\epsilon}{\sqrt{|\beta|}}\left(2\frac{\theta(R-r)}{r}-\delta(r-R)\right), \tag{8}$$

where δ is the Dirac function. Thus in our localized solution we have a spherical distribution of charge such that the charge in the surface confines another equal charge with opposite sign, inside of the sphere.

The energy–momentum tensor in the present case is $T^{\mu\nu}=-g^{\mu\nu}\mathcal{L}$; thus the energy associated with the solution (7) is

$$E=-(4\pi/3)(R^3/|\beta|). \tag{9}$$

B. Electronlike localized solutions

Considering in (2) the two first terms, we have the Lagrangian

$$\mathcal{L}=-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}+(\beta/64)(\tilde{F}_{\mu\nu}F^{\mu\nu})^2. \tag{10}$$

The field equations are

$$\partial_\nu F^{\mu\nu}=(\beta/8)\tilde{F}^{\mu\nu}\partial_\nu(\tilde{F}_{\chi R}F^{\chi R}). \tag{11}$$

These equations admit no stationary solutions which are separable in spherical coordinates, when the system is at rest at the origin. So we have to make a multipole expansion and write down equations for each partial wave. It is, however, much simpler and equivalent to substitute a multipole approximation in the Lagrangian, integrate over the angles and make variations of the radial functions.

Taking the first term in the expansion of A^μ corresponding to an electronlike solution,

$$A^\mu=(A(r),\ -V(r)\sin\theta\sin\phi,\ V(r)\cos\theta\cos\phi,\ 0), \tag{12}$$

we thus obtain

$$\mathcal{L}=\tfrac{1}{2}A'^2-\tfrac{1}{2}[(V'^2+2(VV'/r))\sin^2\theta + (V^2/r^2)(1+3\cos^2\theta)]+\beta(A'^2V^2/r^2)\cos^2\theta, \tag{13a}$$

$$L=\int\mathcal{L}\,d^3r=2\pi\int dr\, r^2[A'^2-2(V^2/r^2)-\tfrac{2}{3}(V'^2+2VV'/r)+\tfrac{2}{3}\beta A'^2V^2/r^2], \tag{13b}$$

from which we get the radial equations

$$\frac{dA}{dr}=\frac{-Q/4\pi}{r^2+\frac{2}{3}\beta V^2}, \tag{14a}$$

$$\frac{d}{dr}\left(r^2\frac{dV}{dr}\right)=V\left(2-\left(\frac{dA}{dr}\right)^2\right), \tag{14b}$$

where Q is a constant of integration and must be interpreted as a charge. Thus, if we compute it from the charge density given by (11)

$$j^0=(\beta/8)\tilde{F}^{0\nu}\partial_\nu(\tilde{F}_{\chi R}F^{\chi R}),$$

we obtain

$$\int j^0 d^3r=Q. \tag{15a}$$

Also we can use the electric induction D_r to obtain the charge

$$D_r=\frac{\partial\mathcal{L}}{\partial E_r}=-\frac{\partial\mathcal{L}}{\partial A'}=-[A'+2\beta A'(V^2/r^2)\cos^2\theta],$$

and, using (14a), we get

$$\int D_r\,d\sigma=Q. \tag{15b}$$

Using the following convenient changes in functions and variables

$$A=(|Q|/4\pi)^{1/2}\beta^{-1/4}\mathcal{A},\quad r=(|Q|/4\pi)^{1/2}\beta^{1/4}\rho,$$
$$V=(\tfrac{3}{2}|Q|/4\pi)^{1/2}\beta^{-1/4}\mathcal{V},\quad \epsilon=Q/|Q|, \tag{16}$$

we get

$$\frac{d\mathcal{A}}{d\rho}=\frac{-\epsilon}{\rho^2+\mathcal{V}^2}, \tag{17a}$$

$$\frac{d}{d\rho}\left(\rho^2\frac{d\mathcal{V}}{d\rho}\right)=\mathcal{V}\left(2-\left(\frac{d\mathcal{A}}{d\rho}\right)^2\right). \tag{17b}$$

Equations (17) can be reduced to the equation

$$\frac{d}{d\rho}\left(\rho^2\frac{d\mathcal{V}}{d\rho}\right)=\mathcal{V}\left(2-\frac{1}{(\rho^2+\mathcal{V}^2)^2}\right). \tag{18}$$

Thus in our approach the solutions are independent of the charge Q and coupling constant β. For the localized solutions the regularity at the origin implies

$$\mathcal{V}(0)=\pm(2)^{-1/4}\left(\frac{d\mathcal{A}}{d\rho}\right)_{\rho=0}=-\epsilon\sqrt{2}\,.$$

The energy–momentum tensor is

$$T^{\alpha\beta}=F^{\rho\alpha}F^{\beta}{}_{\rho}+(\beta/8)(\tilde{F}_{\mu\nu}F^{\mu\nu})\tilde{F}^{\mu\alpha}F^{\beta}_{\mu}-g^{\alpha\beta}\mathcal{L}\,, \tag{19}$$

and we get for the energy

$$E=\int T^{00}d^3r=\tfrac{1}{2}(1/\sqrt{4\pi})|Q|^{3/2}\beta^{-1/4}I_E, \tag{20}$$
$$I_E=\int_0^\infty\rho^2 d\rho(\mathcal{A}'^2+\mathcal{V}'^2+2\mathcal{V}\mathcal{V}'/\rho+3\mathcal{V}^2/\rho^2+\mathcal{A}'^2\mathcal{V}^2/\rho^2),$$

and the energy density is positive.

The definition of spin comes from the spin vector:

$$S_k=\tfrac{1}{2}\epsilon_{ijk}\mathcal{T}^{ij},$$

where

$$\mathcal{T}^{ij}=\int d^3r(x^iT^{j0}-x^jT^{i0}).$$

We find in our case

$$S_1=S_2=0,$$
$$S\equiv S_3=\frac{2}{\sqrt{6}}\frac{Q^2}{4\pi}I_S, \tag{21}$$
$$I_S=\int_0^\infty\rho^2\mathcal{A}'(\rho\mathcal{V}'+\mathcal{V})d\rho.$$

The magnetic moment is given by

$$\mathcal{M}=\tfrac{1}{2}\int\mathbf{r}\times\mathbf{j}\,d^3r,$$

where $\mathbf{j}$, the current density, is

$$j^k=(\beta/8)\tilde{F}^{k\nu}\partial_\nu(\tilde{F}_{\chi R}F^{\chi R});$$

so we have

$$\mathcal{M}=\frac{1}{\sqrt{6}\sqrt{4\pi}}\beta^{1/4}|Q|^{3/2}I_M, \tag{22}$$
$$I_M=\int_0^\infty\mathcal{A}'^2\mathcal{V}\rho d\rho,$$

and, using the field equations (14), we can also prove

$$\mathcal{M}=4\pi\lim_{r\to\infty}r^2V.$$

The mean square charge radius for the charge density j^0 is

$$\langle r^2\rangle = -\frac{Q}{2\pi}\beta^{1/2} I_R, \qquad (23)$$

$$I_R = \int_0^\infty \rho \mathcal{A}' V^2 d\rho.$$

If we fit the model to the physical constants of the electron, we have the following result:

$$Q = -e, \quad E = m_e;$$

from (20),

$$\beta^{1/4} = (e^{3/2}/2\sqrt{4\pi}\, m_e) I_E,$$

and so we get $S = (2\sqrt{6})\alpha\, I_S$, $\mathcal{M} = (\sqrt{6})\ \alpha\, I_E I_M \mu_B$ and, $\langle r^2\rangle^{1/2} = (1/\sqrt{2})\ r_e I_E \sqrt{I_R}$, where $\alpha = e^2/4\pi$ is the fine structure constant, μ_B is a Bohr magneton, and $r_e = \alpha/m_e$ is the classical radius of electron.

III. SOME ESTIMATES FOR ELECTRONLIKE SOLUTIONS

From Eq. (17a) we have

$$|\mathcal{A}'(\rho)| \leq 1/\rho^2 \quad \text{as } \rho \to \infty, \qquad (24)$$

and, since $V^2(0) = 1/\sqrt{2}$, then $\mathcal{A}'(\rho) \equiv \mathcal{A}'(|\mathbf{x}|) \in L^2(\mathbb{R}^3)$. Also we have $\mathcal{A}(\rho) = \mathcal{A}(0) - \int_0^\rho [\epsilon/(\rho^2 + V^2)]\, d\rho$, where $\mathcal{A}(0) = \epsilon \int_0^\infty d\rho/(\rho^2 + V^2)$ and so $\mathcal{A} \in L^\infty(\mathbb{R}^3)$ but $\mathcal{A} \notin L^2(\mathbb{R}^3)$.

On the other hand, we have reduced Eqs. (17) to Eq. (18),

$$\rho^2 V'' + 2\rho V' - 2V = -\frac{V}{(\rho^2 + V^2)^2} \equiv -f(\rho, V),$$

and for their localized solutions we have the following properties:

Lemma 1: Any solution $V(\rho)$ of Eq. (18), which is $C^2(\mathbb{R}^3)$, $L^\infty(\mathbb{R}^3)$ and such that $V^2(0) = 1/\sqrt{2}$, satisfies:

(1) The *a priori* bound

$$|V(\rho)| \leq D/\rho^2, \quad \rho \to \infty, \qquad (25)$$

where D is a constant depending on the

$$\| f(S, V)/S \|_{L^1(\mathbb{R}^3)}.$$

Actually

$$|V(\rho)| = O(1/\rho^2) \quad \text{as } \rho \to \infty. \qquad (26)$$

(2) $V(\rho) \in H^1(\mathbb{R}^3)$, where $H^1(\mathbb{R}^3)$ is the Sobolev space of L^2-functions whose generalized first-order derivatives belong to L^2. For a review of the Sobolev spaces and Sobolev inequalities, we refer the reader to Friedman.[6]

Proof: (1) The Green's function $g(\rho, s)$ for the operator $LV = \rho^2 V_{\rho\rho} + 2\rho V_\rho - 2V$ can be written

$$g(\rho, s) = \begin{cases} -\frac{1}{3}(1/\rho^2)s & \text{for } \rho \geq s, \\ -\frac{1}{3}\rho(1/s^2) & \text{for } \rho \leq s; \end{cases}$$

thus we have as $\rho \to \infty$

$$|V(\rho)| \leq \frac{1}{3\rho^2}\int_0^\rho \frac{s|V|}{(s^2 + V^2)^2}\, ds + \frac{\rho}{3}\int_\rho^\infty \frac{|V|\, ds}{s^2(s^2 + V^2)^2} \leq \frac{1}{3\rho^2}\int_0^\infty \frac{s|V|\, ds}{(s^2 + V^2)^2} = \frac{D}{\rho^2},$$

where $D = (1/\sqrt{2\pi})\|f(s, V)/s\|_{L^1(\mathbb{R}^3)}$, which is finite by hypothesis. Now we can prove that $|V(\rho)| = O(1/\rho^2)$ as $\rho \to \infty$ because in the above expression we have

$$\lim_{\rho\to\infty} \frac{\rho}{3}\int_0^\infty \frac{|V|\, ds}{s^2(s^2 + V^2)^2} = 0.$$

(2) By the above result we have $V(\rho) = V(|\mathbf{x}|) \in L^2(\mathbb{R}^3)$ and also $V(\rho) \in L^2(0, \infty)$ looking at V as a function in the variable ρ.

Using the Green's function, we get

$$|V'(\rho)| \leq \frac{2}{3\rho^3}\int_0^\rho \frac{s|V|}{(s^2 + V^2)^2}\, ds + \tfrac{1}{3}\int_\rho^\infty \frac{|V|\, ds}{s^2(s^2 + V^2)^2} \leq \frac{2}{3\rho^3}\int_0^\infty \frac{s|V|}{(s^2 + V^2)^2} ds = 2D/\rho^3 \text{ as } \rho \to \infty \qquad (27)$$

since $V'(\rho)$ is bounded. In particular $\lim_{\rho\to 0} V'(\rho) < \infty$. Then we have $V'(\rho) \in L^2(\mathbb{R}^3)$ and also $V'(\rho) \in L^2(0, \infty)$ and thus we prove $V(\rho) \in H^1(\mathbb{R}^3)$. In particular $V(\rho) \in H^1(0, \infty)$ too. But by the Sobolev inequality in one dimension $H^1(0, \infty) \subset L^\infty(0, \infty)$, which implies that $V(\rho)$ is bounded, which is the hypothesis of the lemma.

In the following we use the same argument considered by Strauss[7] for a nonlinear elliptic equation.

Lemma 2: Any localized solution $V(\rho)$ of Eq. (18) satisfies the identity

$$\int_0^\infty \frac{2V^2 - \rho^2}{(\rho^2 + V^2)^2}\, d\rho = 0. \qquad (28)$$

Proof: We can regard $V(\rho)$ satisfying

$$\int_0^\infty (-\Delta V + F(\rho, V))(VV')\rho^2 d\rho = 0, \qquad (29)$$

where

$$F(\rho, V) = \frac{1}{\rho^2}\left(2V - \frac{V}{(\rho^2 + V^2)^2}\right);$$

thus

$$\int_0^\infty \left(V'^2 + 6G + \rho\frac{dG}{d\rho}\right)\rho^2\, d\rho = 0, \qquad (30)$$

where

$$G(\rho, V) = \int_0^V F(\rho, u)\, du = \frac{1}{\rho^2}\left(V^2 + \frac{1}{2}\,\frac{1}{(\rho^2 + V^2)}\right).$$

On the other hand from Eq. (18) we get

$$\int_0^\infty (\rho^2 V'^2 + 2V^2)\, d\rho = \int_0^\infty \frac{V^2}{(\rho^2 + V^2)}\, d\rho.$$

Hence by (30) we find

$$\int_0^\infty \frac{2V^2 - \rho^2}{(\rho^2 + V^2)^2}\, d\rho = 0;$$

this condition implies there exist values of V such that $V^2(\rho) > \rho^2/2$, and we can see that is satisfied by the solutions obtained numerically (Figs. 1 and 2) because $V^2(0) = 1/\sqrt{2}$.

IV. NUMERICAL RESULTS

One solves Eq. (18) with boundary values such that $\mathcal{A} \sim \epsilon/\rho$ and $V \sim \mu/\rho^2$ as $\rho \to \infty$. On the other hand, these equations are invariant under the change $V \to -V$ and both solutions correspond to opposite magnetic moments. If we fit the solutions to negative charges we must

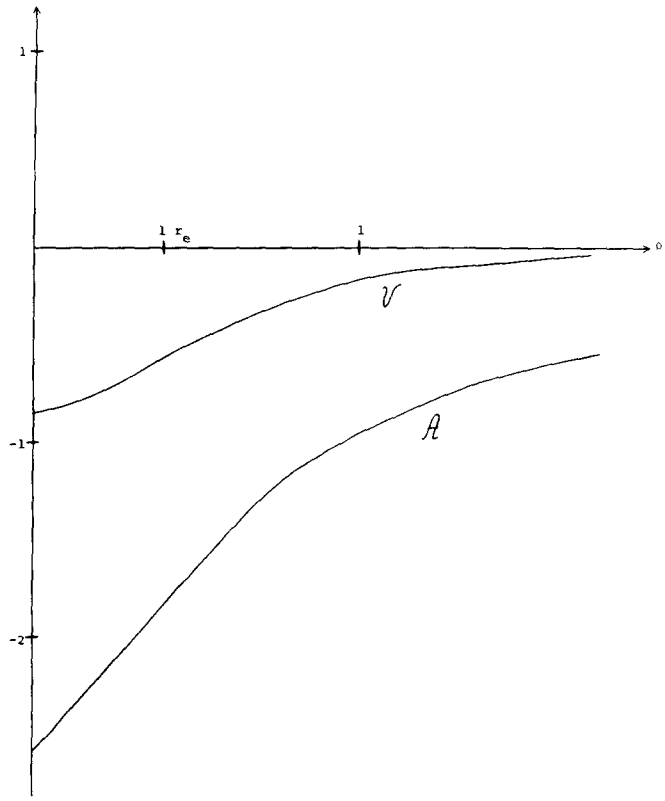


FIG. 1. The first lowest-order solution to (18). A and V are plotted against ρ.

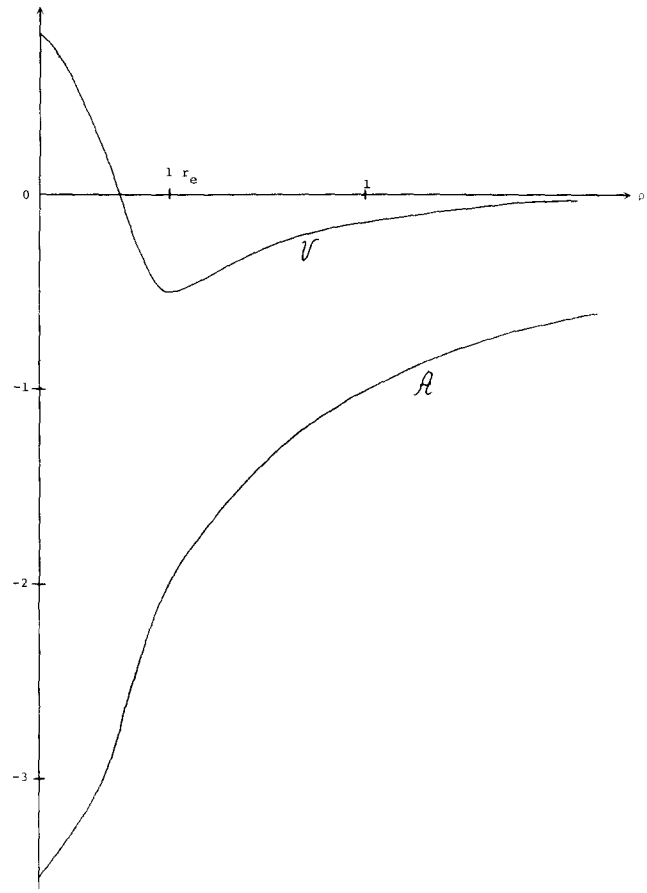


FIG. 2. The second lowest-order solution to (18).

choose these such that $\epsilon = -1$, $\mu < 0$, and the opposite for positive charges.

Looking for solutions with negative charge, we have the following numerical results for the two lowest-order localized states:

$$A(0) = -2.57930, \quad V(0) = -1/(2)^{1/4}, \quad V'(0) = 0.010465,$$

$$A(0) = -3.5240, \quad V(0) = 1/(2)^{1/4}, \quad V'(0) = -0.017966.$$

In Figs. 1, 2 A and V are plotted against ρ, having A 0 and 1 nodes respectively. In Table I we fit the ground state to the physical constants of the electron, as we indicated in Sec. IIB, also, we give the corresponding values for the first excited state. We can see that the spin and magnetic moment are of the order of fine structure constant α, and their sizes are of the order of the classical radius of the electron r_e. From β we obtain $b = 25.25$ (Mev)^2, which is less than the value $b_{BI} = 2268.07$ (Mev)^2 obtained in Refs. 2. For the numerical integration we used Hamming's predictorcorrector method of fourth order with an interval $\Delta\rho = 0.01$.

TABLE I.

	Ground state	First excited state
Charge	$-e$	$-e$
Mass	m_e	$1.39\, m_e$
Magnetic moment	$-1.21\, \alpha\mu_B$	$-1.09\, \alpha\mu_B$
Spin	$0.18\, \alpha$	$0.04\, \alpha$
Radius	$1.48\, r_e$	$1.35\, r_e$
β	$0.31 \cdot 10^{-2}$(Mev)$^{-4}$	$0.31 \cdot 10^{-2}$(Mev)$^{-4}$

V. CONCLUSIONS

We have obtained electronlike solutions for a nonlinear electromagnetic field, numerically, and they are independent of the charge Q and coupling constant β which appears in the Lagrangian obtained by expansion from the Born–Infeld Lagrangian.

Although the electromagnetic field is too simple to be considered as a model of an electron, the values exhibited in the ground state suggest extending the model by introducing another field, like the Dirac field, in order to have a more realistic model. So, for instance, with the spinor field we may expect a better value for the magnetic moment, and the value obtained here will be like an electromagnetic correction as that considered in Quantum Electrodynamics.

On the other hand, we found localized solutions with a purely electric field, its energy negative, the total charge zero, but with a distribution of charge such that a charge in the surface of a sphere confines another equal charge with opposite sign inside of the sphere.

ACKNOWLEDGMENT

I should like to express my sincere thanks to Profes-

sor W. A. Strauss for his very useful suggestions and for reading the manuscript.

*Research supported by a fellowship of the Subdireccion General de Promocion de la Investigacion of Spain.
†On leave from Departamento de Fisica Teorica, Universidad de Zaragoza, Spain.
[1]G. Mie, Ann. Physik 37, 511 (1912); 39, 1 (1912); 40, 1 (1913).
[2]M. Born, Proc. Roy. Soc. A 143, 410 (1934); M. Born and L. Infeld, Proc. Roy. Soc. A 144, 425 (1934).
[3]R. Finkelstein, C. Fronsdal, and P. Kaus, Phys. Rev. 103, 1571 (1956).
[4]A.I. Akhiezer and V.B. Berestetskii, *Quantum Electrodynamics* (English transl.) (Interscience, New York, 1965), p. 781.
[5]The author is grateful to the referee for pointing out and helping to clarify the determination of the charge density and charge.
[6]A. Friedman, *Partial Differential Equations* (Holt, New York, 1969).
[7]W.A. Strauss (unpublished), for the nonlinear equation $-\Delta u + F(u) = 0$ defined in $\mathbb{R}^n$, obtains that localized solutions of the above equation satisfy

$$\int_{\mathbb{R}^n} [2nG(u) - (n-2)uF(u)]\, d^n x = 0, \quad G(u) = \int_0^u F(s)\, ds.$$

Pseudoparticle configurations in two-dimensional ferromagnets*

G. Woo†

Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138
(Received 20 January 1977)

It is proved that all the finite energy solutions to the field equations of the two-dimensional Heisenberg ferromagnet theory are topologically stable.

The motivation for the study of solutions to classical field equations has broadened from an interest in heavy particle spectra to a desire to better understand vacuum phenomena. Although a number of theories could be quoted as examples of the latter, two stand out on geometrical grounds. They are the continuum Heisenberg ferromagnet theory in two space dimensions,[1] and the Euclidean SU(2) Yang–Mills theory in four space–time dimensions.[2] A particle physicist might refer to the ferromagnetic theory as the Euclidean O(3) nonlinear σ-model.

There is more to the mathematical kinship between these two theories than the existence of nontrival mappings of spheres in real space onto spheres in field space. For in both, the topologically stable field configurations can be found by saturating a positivity condition involving the field energy or Euclidean action and the topological index.[1,2] The equations so determined have the great merit of having only first-order derivatives, which aids considerably the construction of exact solutions.

The first-order equations for the two-dimensional theory have been found by Belavin and Polyakov[1] to be completely solvable. Given their work, it is immediately possible to write down for any positive value of the topological index, the general stable multi-pseudoparticle solution, and for any negative value, the corresponding general stable anti-pseudoparticle solution.

However, solutions describing other configurations, in particular those with both pseudoparticles and anti-pseudoparticles, cannot be obtained this way; a return must be made to the original second-order equations.

These are studied in this paper and searched for finite energy solutions. We find that they imply a set of first-order equations which encompass those previously mentioned. We further discover that these general first-order equations only admit finite energy solutions which are topologically stable. There are thus no finite energy solutions to the full field equations other than those representing pure pseudoparticle or pure anti-pseudoparticle configurations. The derivation of this result is presented below.

We consider a two-dimensional theory of a three-component spin field $s^a(x_1, x_2)$ of unit modulus, defined by the Hamiltonian density

$$H = \frac{\partial s^a}{\partial x_i}\,\frac{\partial s^a}{\partial x_i}\,. \tag{1}$$

It has been observed[1] that finite energy field configurations can be classified according to the integer-valued topological index n, where

$$n = \frac{1}{8\pi}\int d^2x\,\epsilon_{abc}\epsilon_{ij}\,s^a\frac{\partial s^b}{\partial x_i}\,\frac{\partial s^c}{\partial x_j}\,. \tag{2}$$

With the angle parametrization

$$s^1 = \cos\theta,\quad s^2 = \sin\theta\cos\phi,\quad s^3 = \sin\theta\sin\phi, \tag{3}$$

new field variables w_1 and w_2 are brought in,[1]

$$w_1 = \cot\frac{\theta}{2}\cos\phi,\quad w_2 = \cot\frac{\theta}{2}\sin\phi. \tag{4}$$

$w_1 + iw_2$ is written as w. In terms of these new variables,

$$H = 4\frac{\nabla w\cdot\nabla w^*}{(1+ww^*)^2}\,. \tag{5}$$

The field equations determining the stationary points of the Hamiltonian are compactly given by the following equation, together with its complex conjugate,

$$\nabla^2 w = \frac{2w^*}{(1+ww^*)}\,(\nabla w)^2. \tag{6}$$

It proves to be convenient to use complex variable techniques to solve these equations. We shall write $w = w(z, z^*)$ where

$$z = x_1 + ix_2 \quad\text{and}\quad z^* = x_1 - ix_2. \tag{7}$$

w is not assumed to be a complex analytic function of these variables, although of course its real and imaginary parts are assumed to be differentiable functions of x_1 and x_2.

Using complex variables then, the first integral of the field equations can be deduced,

$$\left(\frac{\partial w}{\partial z}\,\frac{\partial w^*}{\partial z}\right)/(1+ww^*)^2 = f(z), \tag{8}$$

where $f(z)$ is an analytic function of z, with real and imaginary parts, of course, satisfying the Cauchy–Riemann equations.

Consider first the case where $f(z)$ is not only analytic, but entire. Because the energy density must vanish at infinity, we know that

$$\left|\frac{\partial w/\partial z}{(1+ww^*)}\right| \to 0 \quad\text{as}\quad |z| \to \infty \tag{9a}$$

and

$$\left|\frac{\partial w/\partial z^*}{(1+ww^*)}\right| \to 0 \quad\text{as}\quad |z| \to \infty. \tag{9b}$$

Hence it follows that

$$f(z)\to 0 \quad \text{as} \quad |z|\to\infty. \tag{10}$$

The maximum modulus theorem then implies that

$$f(z)\equiv 0. \tag{11}$$

The field equations then reduce to the two alternatives

$$\frac{\partial w}{\partial z}=0 \quad \text{or} \quad \frac{\partial w}{\partial z^*}=0. \tag{12}$$

The first gives $w=w(z^*)$, an analytic function of z^*, while the second gives $w=w(z)$, an analytic function of z.

In either case, w must be continuous and single-valued everywhere, so it cannot have branch points or essential singularities, hence it must be meromorphic.

The first class of solutions accounts for every field configuration with positive topological index n and energy $E=8\pi n$. These are the stable n-pseudoparticle configurations. The second class of solutions describes the corresponding stable anti-pseudoparticle configurations. The symmetry between particle and antiparticle arises from the freedom in defining the direction of the azimuthal angle of the coordinate system for the sphere S^2 obtained from the two-dimensional plane by identifying the points at infinity.

It is instructive to examine the behavior of the energy density for these solutions. Let us write $w=g(z)/h(z)$ where $g(z)$ and $h(z)$ are polynomials with no common factors. The energy density for this is

$$\mathcal{H}=8\left|\frac{h(z)g'(z)-g(z)h'(z)}{(|h(z)|^2+|g(z)|^2)}\right|^2. \tag{13}$$

Since $h(z)$ and $g(z)$ can never simultaneously vanish, $\mathcal{H}$ is finite everywhere. In particular for the simplest nontrivial topological configuration with $w=z$,

$$\mathcal{H}=8/(1+|z|^2)^2. \tag{14}$$

We note however that the singular energy density

$$\mathcal{H}=2/[|z|(1+|z|)^2] \tag{15}$$

can be obtained with the double-valued choice $w=z^{1/2}$. Even though the energy density has a singularity, the total energy is finite. In fact it is precisely one-half of the energy for $w=z$; a consequence of the behavior of the Hamiltonian under conformal transformations.

We now return to Eq. (8). We have just considered the case where $f(z)$ is entire. Since w is continuous and single-valued, the only singularities allowed to $f(z)$ are poles. The existence of a pole in f signals a singularity in the energy density. In view of our demonstration that pure pseudoparticle and pure anti-pseudoparticle solutions do not have any such singularities, it would be difficult to interpret a solution with one solely in terms of pseudoparticles and anti-pseudoparticles.

To pursue the possibility further, we note that f can only have single poles since the presence of higher-order poles would imply a locally nonintegrable energy density, and hence be in conflict with the finite energy condition.

We shall now study Eq. (8) close to a pole in $f(z)$. Without loss of generality, we may assume that the pole is at the origin and has unit residue. In the neighborhood between the pole considered at $z=0$ and the adjacent one, $f(z)$ has the expansion

$$f(z)=\sum_{n=-1}^{\infty} f_n\left(\frac{z}{z^*}\right)(zz^*)^{n/2} \tag{16}$$

with $f_{-1}=(z/z^*)^{-1/2}$. We solve the equation

$$\frac{\partial w}{\partial z}\frac{\partial w^*}{\partial z}=(1+ww^*)^2 f(z) \tag{17}$$

in the neighborhood of the origin, by making an appropriate series expansion for $w(z,z^*)$,

$$w(z,z^*)=\sum_{n=0}^{\infty} a_n\left(\frac{z}{z^*}\right)(zz^*)^{n/4}. \tag{18}$$

The continuity of w implies that a_0 is a constant. It may be taken to be finite, since solutions of Eq. (8) come in pairs w and $1/w$, which do not both go to infinity as z approaches zero.

To lowest order,

$$\frac{\partial w}{\partial z}=(zz^*)^{-1/4}\left\{\frac{1}{4}\left(\frac{z}{z^*}\right)^{-1/2}a_1+\left(\frac{z}{z^*}\right)^{1/2}a_1'\right\}, \tag{19a}$$

$$\frac{\partial w^*}{\partial z}=(zz^*)^{-1/4}\left\{\frac{1}{4}\left(\frac{z}{z^*}\right)^{-1/2}a_1^*+\left(\frac{z}{z^*}\right)^{1/2}a_1^{*\prime}\right\}, \tag{19b}$$

and

$$f(z)=(zz^*)^{-1/2}\left(\frac{z}{z^*}\right)^{-1/2} \tag{19c}$$

Introducing the polar coordinate angle α by

$$z/z^*=\exp\{2i\alpha\}, \tag{20}$$

we have the equation

$$\left(\frac{a_1(\alpha)}{4}+\frac{1}{2i}a_1'(\alpha)\right)\left(\frac{a_1^*(\alpha)}{4}+\frac{1}{2i}a_1^{*\prime}(\alpha)\right)=(1+|a_0|^2)\exp\{i\alpha\}. \tag{21}$$

From this we deduce that

$$a_1(\alpha)a_1^*(\alpha)=(1+|a_0|^2)(4c+8\cos\alpha), \tag{22}$$

where c is a real integration constant ≥ 2. The solutions for each possible value of c are

$$a_1(\alpha)=4(1+|a_0|^2)^{1/2}\cos\alpha/2 \quad \text{for} \quad c=2 \tag{23a}$$

and

$$a_1(\alpha)=2(1+|a_0|^2)^{1/2}(c+2\cos\alpha)^{1/2}\times\exp\left(\pm i\tan^{-1}\frac{(c^2-4)^{1/2}\tan\alpha/2}{c+2}\right) \quad \text{for} \quad c>2. \tag{23b}$$

In either case, $a_1(\alpha)$ is double-valued, making w locally double-valued at the pole. This situation is somewhat similar to that described earlier in the discussion of pure pseudoparticle solutions. There too, a singularity in the energy density arose from local double-valuedness.

The result then is that singularities in $f(z)$ are not permitted, and the only finite energy solutions to the field equations are those representing stable configurations of pseudoparticles alone, or anti-pseudoparticles alone.

In the stationary phase approximation to the partition function, configurations with both pseudoparticles and anti-pseudoparticles are not selected, and hence are not dominant. However there is the compensating circumstance that certainly many of these configurations only fail to be stationary points of the Hamiltonian by very small margins.

In any sector with a definite value n of the topological index, the difference between the number of pseudoparticles in a configuration and the number of anti-pseudoparticles is, of course, n. So the existence of a finite correlation length can only come from the presence of sufficient numbers of both. It is crucial then that the mean distance between a pseudoparticle and an anti-pseudoparticle be finite. This must be checked by a calculation more sophisticated than the stationary phase approximation for an infinite volume system.

Migdal[3] has suggested that the phase transition characteristics of the lattice version of the theory discussed are related by recursion equations to those of the lattice SU(2) Yang—Mills theory in four space—time dimensions. This invites the speculation that there might be a connection between the possible pseudoparticle configurations which the solutions to the classical equations can describe in each theory.

*Work supported in part by the National Science Foundation under Grant No. MPS75-20427.
†Junior Fellow, Harvard University Society of Fellows.
[1]A.A. Belavin and A.M. Polyakov, Pis'ma Zh. Eksp. Teor. Fiz. **22**, 503 (1975) [JETP Lett. **22**, 245 (1975)].
[2]A.M. Polyakov, Phys. Lett. B **59**, 82 (1975); A.A. Belavin, A.M. Polyakov, A.S. Schwartz, and Yu.S. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[3]A.A. Migdal, Zh. Eksp. Teor. Fiz. **69**, 810, 1457 (1975) [Sov. Phys. JETP **42**, 413, 743 (1975)].

Time-dependent dynamical symmetries and constants of motion. III. Time-dependent harmonic oscillator

Gerald H. Katzin* and Jack Levine

Department of Physics and Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27607*
(Received 20 December 1976)

This paper is a continuation of previous Papers I and II [J. Math. Phys. **17**, 1345 (1976); **18**, 424 (1977)]. In the present paper we apply the theory (based upon Lagrangian dynamics) developed in I and II to obtain the dynamical symmetries and concomitant constants of motion admitted by the time-dependent n-dimensional oscillator (a) $E^i \equiv \ddot{x}^i + 2\omega(t)x^i = 0$. The dynamical symmetries are based upon infinitesimal transformations of the form (b) $\bar{x}^i = x^i + \delta x^i$, $\delta x^i \equiv \xi^i(x,t)\delta a$; $\bar{t} = t + \delta t$, $\delta t \equiv \xi^0(x,t)\delta a$ which satisfy the condition (c) $\delta E^i = 0$, whenever $E^i = 0$. It is shown that such symmetries of the oscillator (a) will be time-dependent projective collineations. For such symmetries which satisfy the R_1 restriction (defined in I) it is shown there exist concomitant constants of motion C_1 of the oscillator, which for $n = 1$ are time-dependent cubic polynomials in the $\dot{x}$ variable, and for $n \geq 2$ are time-dependent quadratic polynomials in the $\dot{x}^i$ variables. It is shown that those symmetries which satisfy the R_2 restriction (Noether symmetry condition discussed in I) are time-dependent homothetic mappings consisting of time-dependent scale changes, time-dependent translations, and rotations. The concomitant Noether constants of motion C_2 are time-dependent quadratic polynomials in the $\dot{x}^i$ variables for all n. The Noether constant of motion C_2 [referred to as $C_2(B)$] for which the associated underlying symmetry mapping is the time-dependent scale change is shown to include as a special case when $n = 1$ a class of invariants formulated by Lewis [Phys. Rev. Lett. **18**, 510 1967)] (by means of a phase space analysis which applies Kruskal's theory in closed form). For the case of general n it is shown that the time-dependent symmetric tensor constant of motion I_{ij} constructed by Günther and Leach [J. Math. Phys. **18**, 572 (1977)] is included as a special case of a time-dependent symmetric tensor constant of motion K_{ij}, where K_{ij} is obtained by use of a time-dependent related integral theorem by means of the symmetry deformation of the constant of motion $C_2(B)$ with respect to the affine collineations; such collineations are a subset of the projective collineation symmetries mentioned above. The symmetries and their concomitant constants of motion of the oscillator (a) with $\omega(t)$ of the form $\omega(t) = a + be^{ct}$ are obtained.

1. INTRODUCTION

In several recent papers[1] we considered the general problem of formulating time-dependent constants of motion as concomitants of infinitesimal time-dependent dynamical symmetry mappings of time-dependent classical particle systems. The conditions for such symmetries were formulated directly at the level of the dynamical equations for both Hamiltonian [Ref. 1(a)] and Lagrangian mechanics [Ref. 1(b), (c)]. In this paper we shall apply the above-mentioned symmetry theory based upon Lagrange's equations to obtain the dynamical symmetries and associated constants of motion of the n-dimensional time-dependent isotropic harmonic oscillator[2]

$$\ddot{x}^i + 2\omega(t)x^i = 0. \tag{1.1}$$

In terms of a Hamiltonian formulation of the one-dimensional time-dependent oscillator (1.1) Lewis[3] has obtained by an application of Kruskal's theory[4] in closed form a class of time-dependent constants of motion which in configuration space notation takes the form

$$I(\rho) \equiv \tfrac{1}{2}\rho^2\mathring{x}^2 - \rho\rho' x\dot{x} + \tfrac{1}{2}(\rho^{-2} + \rho'^2)x^2, \tag{1.2}$$

where $\rho(t)$ satisfies

$$\rho'' + 2\omega\rho - \rho^{-3} = 0. \tag{1.3}$$

Also, in the Hamiltonian formulation Günther and Leach[5] have obtained a generalization of the Lewis constant of motion (1.2) for the n-dimensional oscillator (1.1). This generalization takes the form of a time-dependent symmetric tensor (analogous to the time-independent symmetric tensor discussed by Fradkin[6] for the time-independent oscillator) which we write in configuration space notation as

$$I_{ij}(\rho) \equiv \tfrac{1}{2}[\rho^{-2}x^i x^j + (\rho\mathring{x}^i - \rho' x^i)(\rho\mathring{x}^j - \rho' x^j)], \tag{1.4}$$

where $\rho(t)$ satisfies (1.3). [Note that (1.4) reduces to (1.2) if $n = 1$.]

As part of the work in this paper we show how the above-mentioned constants of motion (1.2) and (1.4) can be obtained in a straightforward manner as concomitants of configuration space symmetries [of the oscillator (1.1)] which have simple geometric interpretation.

In Sec. 2 we give, based upon a Lagrange formulation, a summary of the formulas needed to obtain dynamical symmetry mappings of the form (2.1), (2.2), and their concomitant constants of motion.

In Sec. 3 the symmetry equations are obtained for the general class of time-dependent Lagrangians (3.1). The resulting symmetry equations are found to have solutions in the form of time-dependent projective collineations (3.14) and (3.16).

In Sec. 4 these symmetries are specialized to apply to the case of the oscillator (1.1), and the forms of the symmetry mapping functions $\xi^i(x,t)$, $\xi^0(x,t)$ are obtained.

In Sec. 5 we impose the R_1 restriction (2.16) on the symmetry mapping functions ξ^i, ξ^0 of Sec. 4 in order that the oscillator admit the constant of motion C_1 given

by (2.14) and obtain a general form of the C_1 as given by (5.22) and (5.23).

In Sec. 6 we evaluate the Noether symmetry condition R_2 (2.13) to obtain the Noether symmetry mapping equations for the general class of time-dependent Lagrangians (3.1). The resulting symmetry equations are found to have solutions in the form of time-dependent homothetic motions (6.18), (6.9) [which consist of rotations, time-dependent scale changes, and time-dependent translations]. The concomitant Noether constants of motion C_2 (2.17) associated with these symmetries are obtained in the form (6.26), (6.24), (6.25), respectively.

In Sec. 7 it is shown that the Noether constant of motion $C_2(B)$ (6.24) [which is associated with the time-dependent scale change portion of the general homothetic mapping (6.18), (6.9)] includes as a special case the Lewis constant of motion (1.2).

In Sec. 8 we show the time-dependent symmetric tensor constant of motion (1.4) may be obtained by a simple application of a time-dependent related integral theorem. This application is based upon the symmetry deformation of the constant of motion $C_2(B)$ (6.24) with respect to the affine collineation part (8.3) of the projective collineation symmetry mapping (4.14), (4.15) admitted by the oscillator (1.1).

In Sec. 9 we discuss a specific example of an oscillator (1.1) with $\omega(t)$ given by (9.4).

2. BASIC THEORY

In this section we give a brief review of the procedure for obtaining dynamical symmetry mappings and also give a summary of the equations [derived in Ref. 1(b), (c)] which will be needed to formulate constants of motion associated with these symmetry mappings.[7]

We consider infinitesimal mappings of the form

$$\bar{x}^i = x^i + \delta x^i, \quad \delta x^i \equiv \xi^i(x,t)\delta a, \tag{2.1}$$

$$\bar{t} = t + \delta t, \quad \delta t \equiv \xi^0(x,t)\delta a. \tag{2.2}$$

To within first order in δa it follows that

$$\delta \dot{x}^i \equiv \frac{d\bar{x}^i}{d\bar{t}} - \frac{dx^i}{dt} = (\dot{\xi}^i - \dot{x}^i\dot{\xi}^0)\delta a, \tag{2.3}$$

$$\delta \ddot{x}^i \equiv \frac{d^2\bar{x}^i}{d\bar{t}^2} - \frac{d^2x^i}{dt^2} = (\ddot{\xi}^i - \dot{x}^i\ddot{\xi}^0 - 2\ddot{x}^i\dot{\xi}^0)\delta a. \tag{2.4}$$

For any function

$$G(\ddot{x}, \dot{x}, x, t) \equiv G(\ddot{x}^1, \ldots, \ddot{x}^n; \dot{x}^1, \ldots, \dot{x}^n; x^1, \ldots, x^n, t), \tag{2.5}$$

δG is defined by

$$\delta G \equiv \frac{\partial G}{\partial \ddot{x}^i}\delta \ddot{x}^i + \frac{\partial G}{\partial \dot{x}^i}\delta \dot{x}^i + \frac{\partial G}{\partial x^i}\delta x^i + \frac{\partial G}{\partial t}\delta t. \tag{2.6}$$

A mapping (2.1), (2.2) is said to define a dynamical symmetry of the system described by the differential equations

$$E^i(\ddot{x}, \dot{x}, x, t) = 0, \tag{2.7}$$

provided

$$\delta E^i = 0 \text{ whenever } E^i = 0. \tag{2.8}$$

To obtain the explicit conditions on the functions $\xi^\alpha(x,t)$ $(\equiv \xi^0, \xi^i)$, in order that (2.8) be satisfied, we expand (2.8) by use of (2.1)–(2.6) and then eliminate the $\ddot{x}^i$ terms in the resulting equation by means of (2.7) [which we assume to be solvable for the $\ddot{x}^i$]. The resulting equation will be of the form

$$G^i_\alpha(\dot{x}, x, t)\xi^\alpha + G^{\beta i}_\alpha(\dot{x}, x, t)\xi^\alpha_{,\beta} + G^{\beta\gamma i}_\alpha(\dot{x}, x, t)\xi^\alpha_{,\beta\gamma} = 0. \tag{2.9}$$

The explicit symmetry equations for $\xi^\alpha(x,t)$ are obtained by considering (2.9) as identically zero in the $\dot{x}$ variables.

If the dynamical equations (2.7) are taken to be Lagrange's equations,

$$\Lambda_i(L) \equiv \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i} - \frac{\partial L}{\partial x^i} = 0, \tag{2.10}$$

where the Lagrangian $L = L(\dot{x}, x, t)$, it can be shown [Ref. 1(b)] that the symmetry condition

$$\delta\Lambda_i(L) = 0 \quad [\text{whenever } \Lambda_i(L) = 0], \tag{2.11}$$

may be expressed in the equivalent form

$$\Lambda_i\left(\delta L + L\frac{d}{dt}\delta t\right) = 0 \quad [\text{whenever } \Lambda_i(L) = 0]. \tag{2.12}$$

Noether symmetries [based upon mappings (2.1), (2.2)] are defined by the condition

$$R_2 \equiv \delta L + L\frac{d}{dt}\delta t + \frac{d\gamma_2}{dt} = 0, \tag{2.13}$$

for some $\gamma_2(x,t)$. From (2.12) it follows that all Noether symmetries are solutions of (2.12) [or equivalently (2.11)] but not conversely.

The symmetries defined by (2.12) have as concomitants the constants of motion

$$C_1 \equiv \left(\frac{\partial N}{\partial \dot{x}^i}\dot{x}^i - N\right) + \frac{\partial}{\partial t}\left(\frac{\partial L}{\partial \dot{x}^i}\xi^i - E\xi^0\right) + \gamma_1(x,t), \tag{2.14}$$

$$N\delta a \equiv \delta L + L\frac{d}{dt}\delta t, \quad E \equiv \frac{\partial L}{\partial \dot{x}^i}\dot{x}^i - L, \tag{2.15}$$

provided the restriction

$$R_1 \equiv \frac{\partial \Lambda_i(L)}{\partial t}(\xi^i - \dot{x}\xi^0) + \frac{d\gamma_1}{dt} = 0, \tag{2.16}$$

is satisfied for some $\gamma_1(x,t)$. [We note for all $L = L(\dot{x}, x)$, with $\gamma_1 = 0$, that $R_1 \equiv 0$.]

The Noether symmetries defined by (2.13) have the well-known associated constants of motion [Ref. 1(b)]

$$C_2 \equiv \frac{\partial L}{\partial \dot{x}^i}\xi^i - E\xi^0 + \gamma_2(x,t). \tag{2.17}$$

We shall make use of a related integral theorem [See Ref. 1(b)] which states that if a dynamical system (2.10) admits a symmetry mapping (2.1), (2.2) as described above and also a constant of motion $K(\dot{x}, x, t)$, then the system will also admit the constant of motion

$$\frac{\delta K}{\delta a} \equiv \frac{\partial K}{\partial \dot{x}^i}\frac{\delta \dot{x}^i}{\delta a} + \frac{\partial K}{\partial x^i}\frac{\delta x^i}{\delta a} + \frac{\partial K}{\partial t}\frac{\delta t}{\delta a}. \tag{2.18}$$

In the sections to follow we shall apply the procedures and formulas given above to the specific case of a time-dependent harmonic oscillator (1.1).

3. SYMMETRIES OF TIME-DEPENDENT LAGRANGIAN SYSTEMS

In this section we derive the explicit time-dependent symmetry equations for the dynamical system defined by the Lagrangian

$$L(\overset{\circ}{x},x,t)\equiv\tfrac{1}{2}g_{ij}(x)\overset{\circ}{x}{}^i\overset{\circ}{x}{}^j-V(x,t), \tag{3.1}$$

where g_{ij} defines the metric of a Riemannian (configuration) space V_n.

From (3.1) Lagrange's equations take the form

$$\Lambda_i(L)=g_{ij}(\overset{\circ\circ}{x}{}^j+\Gamma^j_{ab}\overset{\circ}{x}{}^a\overset{\circ}{x}{}^b+g^{ja}V_{,a})=0, \tag{3.2}$$

where $\Gamma^j_{ab}(x)$ is the Christoffel symbol based on the g_{ij}. By inspection (3.2) may be expressed in the equivalent form

$$E^i=\overset{\circ\circ}{x}{}^i+\Gamma^i_{ab}\overset{\circ}{x}{}^a\overset{\circ}{x}{}^b+g^{ia}V_{,a}=0. \tag{3.3}$$

To obtain the symmetry equations by use of (2.8) we calculate δE^i from (3.3) and, as described in Sec. 2, use (2.1)–(2.4) to eliminate the δx^i, δt, $\delta\overset{\circ}{x}{}^i$, $\delta\overset{\circ\circ}{x}{}^i$ terms respectively. In the resulting equation we use (3.3) to eliminate the $\overset{\circ\circ}{x}{}^i$ terms. As a result we obtain

$$\begin{aligned}&-(\xi^0_{,jk}-\xi^0_{,m}\Gamma^m_{jk})\overset{\circ}{x}{}^j\overset{\circ}{x}{}^k\overset{\circ}{x}{}^i+(\xi^i_{,jk}+\xi^m_{,k}\Gamma^i_{jm}+\xi^m_{,j}\Gamma^i_{km}-\xi^i_{,m}\Gamma^m_{jk}\\&\quad+\xi^m\Gamma^i_{jk,m}-\delta^i_j\xi^0_{,kt}-\delta^i_k\xi^0_{,jt})\overset{\circ}{x}{}^j\overset{\circ}{x}{}^k+[\xi^0_{,j}g^{jk}V_{,k}\delta^i_m\\&\quad+2\xi^0_{,m}g^{ij}V_{,j}-\xi^0_{,tt}\delta^i_m+2(\xi^i_{,mt}+\Gamma^i_{mk}\xi^k_{,t})]\overset{\circ}{x}{}^m\\&\quad+(\xi^i_{,tt}-\xi^i_{,j}g^{jm}V_{,m}+2\xi^0_{,t}g^{ij}V_{,j}+\xi^kg^{ij}_{,k}V_{,j}\\&\quad+\xi^kg^{ij}V_{,jk}+\xi^0g^{ij}V_{,jt})=0.\end{aligned} \tag{3.4}$$

The symmetry equations in ξ^0, ξ^i are obtained by requiring that (3.4) be identically zero in the $\overset{\circ}{x}{}^i$. As a consequence we obtain

$$\xi^0_{;jk}=0, \tag{3.5}$$

$$\pounds\Gamma^i_{jk}-\delta^i_j\xi^0_{,tk}-\delta^i_k\xi^0_{,tj}=0, \tag{3.6}$$

$$\xi^0_{,j}g^{jk}V_{,k}\delta^i_m+2\xi^0_{,m}g^{ij}V_{,j}-\xi^0_{,tt}\delta^i_m+2(\xi^i_{,t})_{;m}=0, \tag{3.7}$$

$$\xi^i_{,tt}+g^{ij}V_{;jk}\xi^k-g^{jk}V_{,j}\xi^i_{;k}+2\xi^0_{,t}g^{ij}V_{,j}+\xi^0g^{ij}V_{,jt}=0, \tag{3.8}$$

where[8]

$$\pounds\Gamma^i_{jk}\equiv\xi^i_{,jk}+\xi^m_{,k}\Gamma^i_{jm}+\xi^m_{,j}\Gamma^i_{km}-\xi^i_{,m}\Gamma^m_{jk}+\xi^m\Gamma^i_{jk,m}. \tag{3.9}$$

It is of interest to note that if we define $\phi\equiv\xi^0_{,t}$, then (3.6) is recognized as the condition for a (time-dependent) projective collineation[9] in the configuration space V_n. In addition (3.5) implies that $\phi_{;jk}=0$, which in turn shows the projective collineation (3.6) is a curvature collineation.[10]

We now assume the configuration space to be Euclidean referred to rectangular coordinates x^i. It then follows that $g_{ij}=g^{ij}=\delta_{ij}$, $\Gamma^i_{jk}=0$, and the symmetry equations (3.5)–(3.8) reduce respectively to

$$\xi^0_{,jk}=0, \tag{3.10}$$

$$\xi^i_{,jk}-\delta^i_j\xi^0_{,tk}-\delta^i_k\xi^0_{,tj}=0, \tag{3.11}$$

$$\xi^0_{,k}V_{,k}\delta^i_m+2\xi^0_{,m}V_{,i}-\xi^0_{,tt}\delta^i_m+2\xi^i_{,tm}=0, \tag{3.12}$$

$$\xi^i_{,tt}+V_{,ik}\xi^k-V_{,k}\xi^i_{,k}+2\xi^0_{,t}V_{,i}+\xi^0V_{,tt}=0. \tag{3.13}$$

From (3.10) we immediately obtain

$$\xi^0(x,t)=A_m(t)x^m+B(t). \tag{3.14}$$

The projective collineation equations [in the form (3.11)] have the known solutions[9]

$$\xi^i(x,t)=a_j(t)x^jx^i+B^i_j(t)x^j+C^i(t). \tag{3.15}$$

Use of (3.15) and (3.14) in (3.11) implies $a_j=A'_j$ and hence (3.15) may be expressed in the form

$$\xi^i(x,t)=A'_j(t)x^jx^i+B^i_j(t)x^j+C^i(t). \tag{3.16}$$

We now substitute ξ^i and ξ^0 as given by (3.16) and (3.14) into the two remaining equations (3.12) and (3.13) to obtain, respectively

$$\begin{aligned}&A_kV_{,k}\delta^i_m+2A_mV_{,i}-(A''_kx^k+B'')\delta^i_m\\&\quad+2(A''_mx^i+A''_kx^k\delta^i_m+B^{i\prime}_m)=0,\end{aligned} \tag{3.17}$$

$$\begin{aligned}&A'''_mx^mx^i+B^{i\prime\prime}_mx^m+C^{i\prime\prime}-(A'_jx^i+A'_mx^m\delta^i_j+B^i_j)V_{,j}\\&\quad+2(A'_mx^m+B')V_{,i}+V_{,ik}(A'_mx^mx^k+B^k_mx^m+C^k)\\&\quad+V_{,it}(A_mx^m+B)=0.\end{aligned} \tag{3.18}$$

We may now state the following theorem.

Theorem 3.1: *For a dynamical system defined by Lagrangian* (3.1) *the associated symmetry equations derived from condition* (2.8) *are given by* (3.5)–(3.8). *When the configuration space V_n is a flat space referred to rectangular coordinates (x^i) these symmetry equations reduce to* (3.17) and (3.18) *in the unknowns $A_i(t)$, $B^i_j(t)$, $C^i(t)$. The symmetry mapping* (2.1), (2.2) *is then determined by* (3.14), (3.16).

4. SYMMETRIES OF THE TIME-DEPENDENT HARMONIC OSCILLATOR (1.1)

We now specify the potential in (3.1) to be[2]

$$V(x,t)\equiv\omega(t)r^2,\quad r^2\equiv\textstyle\sum(x^i)^2 \tag{4.1}$$

and obtain from (3.3) (with $g^{ij}=\delta_{ij}$, $\Gamma^i_{jk}=0$) the dynamical equations for the n-dimensional time-dependent harmonic oscillator

$$E^i\equiv\ddot{x}^i+2\omega(t)x^i=0. \tag{4.2}$$

To determine the symmetry mapping (ξ^i,ξ^0) for this dynamical system we use (4.1) in (3.17) and (3.18) and obtain, respectively

$$(2\omega A_k\delta^i_m+4\omega A_m\delta^i_k+2A''_m\delta^i_k+A''_k\delta^i_m)x^k+2B^{i\prime}_m-B''\delta^i_m=0, \tag{4.3}$$

$$\begin{aligned}&(A'''_m+2\omega A'_m+2\omega'A_m)x^mx^i\\&\quad+(B^{i\prime\prime}_m+4\omega B'\delta^i_m+2\omega'B\delta^i_m)x^m+C^{i\prime\prime}+2\omega C^i=0.\end{aligned} \tag{4.4}$$

It follows from (4.3) that we must have

$$2\omega A_k\delta^i_m+4\omega A_m\delta^i_k+2A''_m\delta^i_k+A''_k\delta^i_m=0, \tag{4.5}$$

$$2B^{i\prime}_m-B''\delta^i_m=0. \tag{4.6}$$

It is easily shown that (4.5) can be written in the equivalent form

$$A''_k+2\omega A_k=0, \tag{4.7}$$

and (4.6) can be integrated to obtain

$$B^i_m=\tfrac{1}{2}\delta^i_mB'+\alpha^i_m,\quad \alpha^i_m\equiv\text{consts}. \tag{4.8}$$

We next obtain from (4.4) the conditions

$$C^{i\prime\prime}+2\omega C^i=0, \tag{4.9}$$

$$B^{i\prime\prime}_m+(4\omega B'+2\omega' B)\delta^i_m=0, \tag{4.10}$$

$$A'''_m+2\omega A'_m+2\omega' A_m=0. \tag{4.11}$$

By inspection, solutions to (4.7) will satisfy (4.11). Hence we dispense with (4.11).

If (4.8) is used in (4.10) we obtain

$$B'''+8\omega B'+4\omega' B=0. \tag{4.12}$$

It is noted that if (4.12) is multiplied by B the resulting equation can be immediately integrated to give

$$B''B-\tfrac{1}{2}B'^2+4\omega B^2=c_0,\quad c_0=\text{const.} \tag{4.13}$$

We summarize the above results in the theorem to follow.

Theorem 4.1: *An n-dimensional time-dependent harmonic oscillator (in a Euclidean configuration space referred to rectangular coordinates x^i) (4.2) will admit a time-dependent symmetry mapping* (2.1), (2.2) [*as determined by the condition* (2.8)] *provided*

$$\xi^i(x,t)=A'_j x^j x^i+\tfrac{1}{2}B' x^i+\alpha^i_j x^j+C^i,\quad \alpha^i_j=\text{consts}, \tag{4.14}$$

$$\xi^0(x,t)=A_j x^j+B, \tag{4.15}$$

where $A_j(t)$, $B(t)$, and $C^i(t)$ satisfy (4.7), (4.13), *and* (4.9), *respectively*.

5. TIME-DEPENDENT HARMONIC OSCILLATOR CONSTANTS OF MOTION DERIVED FROM SYMMETRIES SATISFYING THE R_1 RESTRICTION

We now determine the conditions imposed on the symmetries (given in Theorem 4.1) of the time-dependent harmonic oscillator by the R_1 restriction (2.16). For a Euclidean configuration space referred to rectangular coordinates we note from (2.10) and (4.2) that $\Lambda_i(L)=g_{ij}E^j=E^i$. Hence

$$\frac{\partial\Lambda_i(L)}{\partial t}=\frac{\partial E^i}{\partial t}=2\omega' x^i. \tag{5.1}$$

By use of (4.14), (4.15), and (5.1) in the R_1 restriction (2.16) we obtain

$$2\omega'(A'_j x^j r^2+B^i_j x^j x^i+C^i x^i)+\gamma_{1,t}-[2\omega'(A_j x^j+B)x^i-\gamma_{1,i}]\dot{x}^i=0, \tag{5.2}$$

where for compactness of notation we have reintroduced B^i_j as defined by (4.8). Since (5.2) must hold identically in the $\dot{x}^i$ we obtain

$$\gamma_{1,t}=-2\omega'(A'_j x^j r^2+B^i_j x^j x^i+C^i x^i),\qquad \gamma_{1,i}=2\omega'(A_j x^j+B)x^i. \tag{5.3}$$

From (5.3) the integrability conditions

$$\gamma_{1,ik}=\gamma_{1,ki},\quad \gamma_{1,tk}=\gamma_{1,kt}, \tag{5.4}$$

require respectively that

$$(A_k\delta^i_j-A_i\delta^k_j)x^j=0, \tag{5.5a}$$

$$[\omega'(A'_k\delta^j_m+3A'_j\delta^k_m)+\omega''A_j\delta^k_m]x^j x^m+[\omega'(B^j_k+B^k_j)+(\omega' B)_{,t}\delta^k_j]x^j+\omega' C^k=0. \tag{5.5b}$$

It follows from (5.5a) and (5.5b) that (assuming $\omega'\neq 0$)

$$A_k\delta^i_j-A_i\delta^k_j=0, \tag{5.6}$$

$$\omega'[A'_k\delta^j_m+\tfrac{3}{2}(A'_j\delta^k_m+A'_m\delta^k_j)]+\tfrac{1}{2}\omega''(A_j\delta^k_m+A_m\delta^k_j)=0, \tag{5.7}$$

$$\omega'(B^j_k+B^k_j)+(\omega' B)'\delta^k_j=0, \tag{5.8}$$

$$C^k=0. \tag{5.9}$$

From (5.8) with the use of (4.8) we obtain

$$(\omega''B+2\omega' B')\delta^k_j+\omega'(\alpha^j_k+\alpha^k_j)=0. \tag{5.10}$$

In (5.10) we find for the case $j\neq k$ that

$$\alpha^j_k+\alpha^k_j=0,\quad j\neq k, \tag{5.11}$$

and for the case $k=j$ it follows that

$$\alpha^1_1=\alpha^2_2=\cdots=\alpha^n_n\equiv\alpha. \tag{5.12}$$

We combine (5.11) and (5.12) into the condition

$$\alpha^j_k+\alpha^k_j=2\alpha\delta^j_k. \tag{5.13}$$

By use of (5.13) we find (5.10) is identically satisfied when $j\neq k$, and when $j=k$ (5.10) reduces to the single equation

$$\omega''B+2\omega' B'+2\alpha\omega'=0. \tag{5.14}$$

From (5.6) we find by contraction on i and k that

$$(n-1)A_j=0. \tag{5.15}$$

We thus have two cases to consider.

Case $n=1$: If $n=1$ (5.15) and (5.6) are satisfied identically and (5.7) reduces to the single equation

$$\omega''A+4\omega' A'=0\quad (A_1\equiv A). \tag{5.16}$$

From (5.16) we find that

$$A^4=c_1/\omega'\quad (c_1\equiv\text{const.}) \tag{5.17}$$

Case $n\geq 2$: If $n\geq 2$, then from (5.15) we have

$$A_j=0,\quad j=1,\ldots,n, \tag{5.18}$$

and (5.6), (5.7) are identically satisfied.

Collecting together the above results we have that the integrability conditions (5.4) can be reduced to the three equations (5.9), (5.13), (5.14) [which hold for all n], to which must be added (5.17) when $n=1$ or (5.18) when $n\geq 2$.

Hence the integration of (5.3) results in

$$(n=1),\quad \gamma_1(x,t)=\omega'\left(\frac{2A}{3}x^3+Bx^2\right)+\nu_0,\quad \nu_0=\text{const}, \tag{5.19}$$

$$(n\geq 2),\quad \gamma_1(x,t)=\omega' Br^2+\nu_0. \tag{5.20}$$

The above results are incorporated in the following theorem.

Theorem 5.1: *For the n-dimensional time-dependent harmonic oscillator a necessary and sufficient condition that the mapping functions $\xi^i(x,t)$ and $\xi^0(x,t)$ [given by* (4.14), (4.15), *respectively] will define a symmetry which in addition will also satisfy the R_1 restriction* (2.16) *is that the following equations hold*:

$$\alpha^i_j+\alpha^j_i=2\alpha\delta^i_j,\quad \alpha^i_j=\text{const}\quad (\text{all } n), \tag{5.13}$$

$$B''B-\tfrac{1}{2}(B')^2+4\omega B^2=c_0,\quad c_0=\text{const}\quad(\text{all } n),\tag{4.13}$$

$$2\omega'B'+\omega''B+2\alpha\omega'=0\quad(\text{all } n),\tag{5.14}$$

$$C^i=0\quad(\text{all } n),\tag{5.9}$$

$$A''+2\omega A=0\quad(n=1),\quad A\equiv A_1,\tag{4.7}$$

$$A^4=c_1/\omega',\quad c_1=\text{const}\quad(n=1),\tag{5.17}$$

$$A_i=0\quad(n\geq 2).\tag{5.18}$$

The function $\gamma_1(x,t)$ appearing in (2.16) *is given by* (5.19) and (5.20) *for the cases* $n=1$ and $n\geq 2$, *respectively.*

We now assume the conditions of Theorems 4.1 and 5.1 are satisfied and determine the C_1 constant of motion (2.14) associated with the existence of symmetry mappings which satisfy the R_1 restriction.

Based upon the oscillator Lagrangian

$$L\equiv\tfrac{1}{2}\delta_{ij}\dot{x}^i\dot{x}^j-\omega(t)r^2,\quad i,j=1,\ldots,n,\tag{5.21}$$

we evaluate (2.14) by the use of (2.15), (4.14), (4.15), (5.19), and (5.20) to obtain

$$(n=1),\quad C_1=C_1(A)+C_1(B),\tag{5.22}$$

$$(n\geq 2),\quad C_1=C_1(B),\tag{5.23}$$

where (the additive constant ν_0 has been dropped)

$$C_1(A)\equiv -A\dot{x}^3+A'x\dot{x}^2+A''x^2\dot{x}+2(\omega A'+\tfrac{1}{3}\omega'A)x^3,\tag{5.24}$$

$$C_1(B)\equiv(-\tfrac{1}{2}B'+\alpha)\dot{x}^i\dot{x}^i+\tfrac{1}{2}B''x^i\dot{x}^i+(\omega B'+\omega'B+2\alpha\omega)x^ix^i,\tag{5.25}$$

and where the functions ω, A, and B and the constant α satisfy the conditions of Theorem 5.1.

By use of (1.1), (4.7), and (5.16) it is easily verified that $dC_1(A)/dt$ vanishes along a dynamical path and hence $C_1(A)$ [for $n=1$] is itself a constant of motion. In a similar manner by use of (1.1), (4.12), and (5.14), $C_1(B)$ (for all n) is also a constant of motion.

We may thus state the following theorem.

Theorem 5.2: *The n-dimensional time-dependent harmonic oscillator (in a Euclidean configuration space referred to rectangular coordinates) with Lagrangian* (5.21) *will admit the constant of motion* C_1 [*based on* (2.14)] *given by* (5.22) *when* $n=1$, *provided that* (4.7), (5.17), (4.13), *and* (5.14) *are satisfied; and will admit the constant of motion* (5.23) *when* $n\geq 2$, *provided that* (4.13) *and* (5.14) *are satisfied. In* (5.22) $C_1(A)$ *and* $C_1(B)$ *are individually constants of motion.*

6. NOETHER SYMMETRY MAPPINGS FOR THE TIME-DEPENDENT HARMONIC OSCILLATOR

In this section we first derive the explicit time-dependent Noether symmetry equations for the general class of dynamical systems defined by the Lagrangian (3.1). These equations will be obtained by use of the Noether symmetry condition (2.13).

For any Lagrangian $L=L(\dot{x},x,t)$ we find by the use of (2.6), (2.1), (2.2), and (2.3) in (2.13) that

$$\frac{\partial L}{\partial\dot{x}^i}\dot{x}^j\xi^i_{,j}+\frac{\partial L}{\partial\dot{x}^i}\xi^i_{,t}-\frac{\partial L}{\partial\dot{x}^i}\ddot{x}^i\xi^0_{,j}-\frac{\partial L}{\partial\dot{x}^i}\dot{x}^i\xi^0_{,t}+\frac{\partial L}{\partial x^i}\xi^i+\frac{\partial L}{\partial t}\xi^0+L\dot{x}^j\xi^0_{,j}+L\xi^0_{,t}+\gamma_{2,j}\dot{x}^j+\gamma_{2,t}=0.\tag{6.1}$$

If (3.1) is used in (6.1) we obtain

$$-\tfrac{1}{2}g_{ab}\xi^0_{,i}\dot{x}^a\dot{x}^b\dot{x}^j+\tfrac{1}{2}(g_{ab,i}\xi^i+g_{ib}\xi^i_{,a}+g_{ia}\xi^i_{,b}-g_{ab}\xi^0_{,t})\dot{x}^a\dot{x}^b+(g_{ij}\xi^i_{,t}-V\xi^0_{,j}+\gamma_{2,j})\dot{x}^j-(V_{,i}\xi^i+V_{,t}\xi^0+V\xi^0_{,t}-\gamma_{2,t})=0.\tag{6.2}$$

We require that (6.2) hold identically in the $\dot{x}^i$ and thus obtain the explicit Noether symmetry equations

$$g_{ij}\xi^0_{,k}+g_{kj}\xi^0_{,i}+g_{ik}\xi^0_{,j}=0,\tag{6.3}$$

$$\pounds g_{ij}-\xi^0_{,t}g_{ij}=0,\tag{6.4}$$

$$g_{ij}\xi^i_{,t}-V\xi^0_{,j}+\gamma_{2,j}=0,\tag{6.5}$$

$$V_{,i}\xi^i+V_{,t}\xi^0+V\xi^0_{,t}-\gamma_{2,t}=0,\tag{6.6}$$

where the Lie derivative of the metric tensor g_{ij} is defined by

$$\pounds g_{ij}\equiv g_{ij,k}\xi^k+g_{kj}\xi^k_{,i}+g_{ik}\xi^k_{,j}.\tag{6.7}$$

Equation (6.3) may be expressed in the equivalent form

$$\xi^0_{,k}=0.\tag{6.8}$$

Hence

$$\xi^0=B(t),\tag{6.9}$$

and (6.4) then defines time-dependent homothetic motions.[11]

We now assume the space to be Euclidean referred to rectangular coordinates. By this assumption and use of (6.9) the remaining Noether symmetry equations (6.4), (6.5), and (6.6) reduce respectively to the equations (since now $g_{ij}=\delta_{ij}$, $\xi^i=\xi_i$)

$$\xi_{i,j}+\xi_{j,i}-B'\delta_{ij}=0,\tag{6.10}$$

$$\xi_{i,t}+\gamma_{2,i}=0,\tag{6.11}$$

$$V_{,i}\xi^i+V_{,t}B+VB'-\gamma_{2,t}=0.\tag{6.12}$$

The time-dependent homothetic motion equations (6.10) have the known solution[12]

$$\xi^i(x,t)=\tfrac{1}{2}B'x^i+B^i_j(t)x^j+C^i(t),\quad B^i_j+B^j_i=0.\tag{6.13}$$

If ξ^i from (6.13) is used in (6.11) we obtain

$$\tfrac{1}{2}B''x^i+B^i_j{}'x^j+C^{i\prime}+\gamma_{2,i}=0.\tag{6.14}$$

From the integrability conditions $\gamma_{2,ik}=\gamma_{2,ki}$ we obtain by use of (6.14) that $B^i_j=\omega^i_j=$ consts and hence (6.14) takes the form

$$\tfrac{1}{2}B''x^i+C^{i\prime}+\gamma_{2,i}=0,\tag{6.15}$$

which integrates to give

$$\gamma_2=-\tfrac{1}{4}B''r^2-C^{i\prime}x^i+D(t),\quad r^2=\textstyle\sum(x^i)^2.\tag{6.16}$$

If (6.16) and (6.13) [with $B^i_j=\omega^i_j$] are used in (6.12) we obtain as the remaining condition on ω^i_j, $B(t)$, $C^i(t)$, $D(t)$ the equation

$$V_{,i}(\tfrac{1}{2}B'x^i+\omega^i_jx^j+C^i)+V_{,t}B+VB'+\tfrac{1}{4}B'''r^2+C^{i\prime\prime}x^i-D'=0.\tag{6.17}$$

Hence for a given $V(x,t)$ a Noether symmetry will be defined by (6.9) and

$$\xi^i = \tfrac{1}{2}B'x^i + \omega^i_j x^j + C^i(t), \qquad \omega^i_j + \omega^j_i = 0, \quad \omega^i_j = \text{consts}, \tag{6.18}$$

provided (6.17) is satisfied.

We now obtain the Noether symmetries for the time-dependent harmonic oscillator by taking the potential to be of the form (4.1).

Use of (4.1) in (6.17) gives

$$\tfrac{1}{4}(B''' + 8\omega B' + 4\omega' B)r^2 + (C^{i\prime\prime} + 2\omega C^i)x^i - D' = 0. \tag{6.19}$$

From (6.19) we immediately obtain

$$B''' + 8\omega B' + 4\omega' B = 0, \tag{6.20}$$

$$C^{i\prime\prime} + 2\omega C^i = 0, \tag{6.21}$$

$$D' = 0. \tag{6.22}$$

[Note that (6.20) is the same as (4.12) and hence can be integrated to yield (4.13).]

We summarize the above in the following theorem.

Theorem 6.1: *An n-dimensional time-dependent harmonic oscillator (in a Euclidean configuration space referred to rectangular coordinates x^i) with Lagrangian defined by* (5.21) *will admit a time-dependent Noether symmetry mapping* (2.1), (2.2) [*as determined by the condition* (2.13)] *where ξ^0, ξ^i are given by* (6.9), (6.18) *respectively, provided $B(t)$ satisfies* (6.20) [*or equivalently* (4.13)] *and $C^i(t)$ satisfies* (6.21). *Such a symmetry mapping is a time-dependent homothetic motion (a time-dependent scale change). The associated function $\gamma_2(x,t)$ is given by* (6.16) *with $D(t) = d_0 = \text{const}$.*

By comparison of the forms of ξ^0, ξ^i given in Theorem 6.1 with the corresponding forms given in Theorem 4.1 it is easily verified (as mentioned in Sec. 2) that the Noether symmetry mapping defined by (2.13) is a special case of the general symmetry mapping defined by (2.11).

For the oscillator referred to in Theorem 6.1 the Noether constant of motion C_2 given by (2.17) is now easily constructed by use of (2.17), (2.15), (5.21), and the functions ξ^0, ξ^i, γ_2 defined in Theorem 6.1. The C_2 so obtained may be expressed in the form (the additive constant d_0 referred to in Theorem 6.1 is taken to be zero)

$$C_2 \equiv C_2(B, C^i, \omega^i_j) = C_2(B) + \sum_{i=1}^{n} C_2(C^i) + \sum_{i<j} C_2(\omega^i_j), \tag{6.23}$$

where

$$C_2(B) \equiv -\tfrac{1}{2}B\mathring{x}^i\mathring{x}^i + \tfrac{1}{2}B'x^i\mathring{x}^i - \tfrac{1}{4}(B'' + 4\omega B)x^ix^i, \tag{6.24}$$

$$C_2(C^i) \equiv C^i\mathring{x}^i - C^{i\prime}x^i, \tag{6.25}$$

$$C_2(\omega^i_j) \equiv \omega^i_j x^j\mathring{x}^i, \quad \omega^i_j + \omega^j_i = 0, \quad \omega^i_j = \text{const}. \tag{6.26}$$

By use of (1.1), (6.20), (6.21) and the skew-symmetry of ω^i_j it is easily verified that $C_2(B)$, $C_2(C^i)$, and $C_2(\omega^i_j)$ are individually constants of motion.

We thus have the theorem.

Theorem 6.2: *The n-dimensional time-dependent harmonic oscillator defined in Theorem* 6.1 *will admit the Noether constants of motion* (6.24), (6.25), (6.26), *and hence* (6.23), *provided* (6.20) *and* (6.21) *are satisfied and the constants ω^i_j are skew-symmetric.*

7. THE LEWIS CONSTANT OF MOTION AS A CONCOMITANT OF A NOETHER SYMMETRY

In this section we show how the Lewis constant of motion I (1.2) for the one-dimensional oscillator (1.1) can be expressed as a particular case of the Noether constant of motion C_2 (6.23) which was derived in Sec. 6 from a Noether symmetry.

To show this we first make a change in variable in the function $B(t)$ of Theorem 6.1 by putting

$$B = -\rho^2. \tag{7.1}$$

From (7.1) it follows that

$$B' = -2\rho\rho', \quad B'' = -2(\rho')^2 - 2\rho''. \tag{7.2}$$

We now express the integrated form (4.13) of (6.20) in terms of ρ by means of (7.1) and (7.2) and obtain

$$\rho'' + 2\omega\rho - \frac{c_0}{2}\rho^{-3} = 0. \tag{7.3}$$

Next we use (7.1) and (7.2) in (6.23) to express C_2 in the form

$$\bar{C}_2 \equiv C_2[B(\rho), C^i, \omega^i_j] \equiv \bar{C}_2(\rho) + \sum_{i=1}^{n} C_2(C^i) + \sum_{i<j} C_2(\omega^i_j), \tag{7.4}$$

where

$$\bar{C}_2(\rho) \equiv C_2[B(\rho)] = \frac{\rho^2}{2}\mathring{x}^i\mathring{x}^i - \rho\rho' x^i\mathring{x}^i + \{\tfrac{1}{2}[(\rho')^2 + \rho\rho''] + \omega\rho^2\}x^ix^i. \tag{7.5}$$

If ρ'' is eliminated from (7.5) by means of (7.3) and the result used in (7.4), then we obtain

$$\bar{C}_2 = \frac{\rho^2}{2}\mathring{x}^i\mathring{x}^i - \rho\rho' x^i\mathring{x}^i + \tfrac{1}{2}\left[(\rho')^2 + \frac{c_0}{2}\rho^{-2}\right]x^ix^i + \sum_{i=1}^{n} C_2(C^i) + \sum_{i<j} C_2(\omega^i_j). \tag{7.6}$$

Hence the Noether constant of motion C_2 of (6.23) is expressed in the form $\bar{C}_2$ of (7.6) where ρ satisfies (7.3) [and $C^i(t)$ and ω^i_j satisfy the conditions of Theorem 6.1].

If now we consider the one-dimensional oscillator ($n = 1$) and choose $C^i = 0$ in (7.6) and make use of (6.25) and (6.26), $\bar{C}_2$ reduces to the form

$$\bar{C}_2 = \bar{C}_2(\rho) = \frac{\rho^2}{2}\mathring{x}^2 - \rho\rho' x\mathring{x} + \tfrac{1}{2}\left[(\rho')^2 + \frac{c_0}{2}\rho^{-2}\right]x^2. \tag{7.7}$$

Finally if in (7.7) we choose $c_0 = 2$, then $\bar{C}_2$ of (7.7) reduces to the Lewis constant of motion I (1.2), and in addition the condition (7.3) reduces to the Lewis condition (1.3).

The Noether symmetry mapping underlying the constant of motion $\bar{C}_2(\rho)$ of (7.7) takes the form

$$\xi = -\rho\rho' x, \quad \xi^0 = -\rho^2, \tag{7.8}$$

as is evident from (6.9), (6.18), and (7.1).

From the above results and the remarks in Theorem 6.1 we may now state the following theorem.

Theorem 7.1: *The Lewis constant of motion* (1.2) [*where* ρ *satisfies* (1.3)] *of the one-dimensional oscillator* (1.1) *is a Noether constant of motion based upon the time-dependent homothetic symmetry mapping* (7.8) (*a time-dependent scale change*).

8. RELATED INTEGRAL THEOREM DERIVATION OF TIME-DEPENDENT SYMMETRIC TENSOR CONSTANTS OF MOTION

It is well known[6] that for the time-independent harmonic oscillator [of the form (1.1) with $\omega = \omega_0 = \text{const}$] there exists a symmetric tensor whose components S_{ij} are constants of motion

$$S_{ij} \equiv \dot{x}^i \dot{x}^j + 2\omega_0 x^i x^j. \tag{8.1}$$

Using a Lagrangian formulation we have previously shown[13] how such constants of motion arise in a straightforward manner as concomitants of infinitesimal symmetry mappings (in the form of affine collineations) admitted by the oscillator. As also shown in our previous work this symmetric tensor constant of motion could be reformulated as the symmetry deformation of the energy integral. Such a reformulation is consistent with the alternative derivation of these constants of motion by means of a related integral theorem[13]—a method for deriving additional constants of motion by means of symmetry deformation of given constants of motion [see (2.18)]. Similar remarks apply to the phase space formulation of the oscillator problem in which these constants of motion were expressible as symmetry deformations of the Hamilton.[14]

For the n-dimensional time-dependent oscillator Günther and Leach[5(a)] have constructed a time-dependent symmetric tensor by generalizing the Lewis constant of motion (1.2) in a manner analogous to Fradkin's generalization which led to (8.1). The Günther—Leach generalization (in our notation) is given by (1.4).

We now show how the constants of motion I_{ij} of (1.4) may be obtained as a simple application of a time-dependent related integral theorem (see Sec. 2) based upon the symmetry deformation of the constant of motion $C_2(B)$ (6.24) with respect to the affine collineation

$$\xi^i = \alpha^i_j x^j, \quad \xi^0 = 0, \quad \alpha^i_j = \text{consts}. \tag{8.2}$$

[Note that (8.2) is a special case of the general symmetry solution (4.14), (4.15) of the time-dependent oscillator; note also that both the time-independent and dependent oscillators admit this affine collineation.]

Consider then the Noether constant of motion $C_2(B)$ (6.24) and form the derived constant of motion by means of (2.18) to obtain

$$\frac{\delta C_2(B)}{\delta a} = \frac{\partial C_2(B)}{\partial x^k}\frac{\delta x^k}{\delta a} + \frac{\partial C_2(B)}{\partial \dot{x}^k}\frac{\delta \dot{x}^k}{\delta a} + \frac{\partial C_2(B)}{\partial t}\frac{\delta t}{\delta a}. \tag{8.3}$$

By (8.2) and (2.1), (2.2), (2.3) we find

$$\frac{\delta x^k}{\delta a} = \alpha^k_j x^j, \quad \frac{\delta t}{\delta a} = 0, \quad \frac{\delta \dot{x}^k}{\delta a} = \alpha^k_j \dot{x}^j. \tag{8.4}$$

By use of (8.4) and (6.24) in (8.3) we obtain

$$\frac{\delta C_2(B)}{\delta a} = \alpha^k_j K_{jk}(B), \tag{8.5}$$

where

$$K_{jk}(B) \equiv -B\dot{x}^j\dot{x}^k + \tfrac{1}{2}B'(\dot{x}^k x^j + \dot{x}^j x^k) - \tfrac{1}{2}(B'' + 4\omega B)x^j x^k. \tag{8.6}$$

Since the left-hand side of (8.5) is a constant of motion and the α^k_j are arbitrary constants it follows that the $K_{jk}(B)$ are components of a symmetric tensor constant of motion.

Recall that $B(t)$ must satisfy (4.13). If in (4.13) we assume $B = B_0 = \text{const}$, then $\omega = \omega_0 = \text{const}$ and (8.6) reduces, to within a constant factor, to the time-independent symmetric tensor constant of motion (8.1).

To obtain from (8.6) the Günther—Leach time-dependent symmetric tensor constant of motion I_{ij} (1.4) we make the substitution (7.1) in (8.6). This gives

$$K_{jk}(B) \equiv \bar{K}_{jk}(\rho) = \rho^2\dot{x}^j\dot{x}^k - \rho\rho'(\dot{x}^k x^j + \dot{x}^j x^k) + [\rho\rho'' + (\rho')^2 + 2\omega\rho^2]x^j x^k. \tag{8.7}$$

From the calculation of Sec. 7 ρ must satisfy (7.3). Hence by use of (7.3) in (8.7) we obtain

$$\bar{K}_{jk}(\rho) = \rho^2\dot{x}^j\dot{x}^k - \rho\rho'(\dot{x}^k x^j + \dot{x}^j x^k) + \left((\rho')^2 + \frac{c_0}{2}\rho^{-2}\right)x^j x^k. \tag{8.8}$$

By the choice of the arbitrary constant $c_0 = 2$ we obtain

$$\bar{K}_{jk}(\rho) = 2I_{jk}(\rho). \tag{8.9}$$

Hence for $c_0 = 2$, $B = -\rho^2$ and δ based upon ξ^i, ξ^0 of (8.2) we may write

$$\frac{\delta C_2(B)}{\delta a} = \alpha^k_j K_{jk}(B) = \alpha^k_j \bar{K}_{jk}(\rho) = 2\alpha^k_j I_{jk}(\rho). \tag{8.10}$$

We summarize the above in the following theorem.

Theorem 8.1: *The time-dependent harmonic oscillator* (1.1) *admits a time-dependent symmetric tensor constant of motion* $K_{ij}(B)$ (8.6) *obtainable as the* [*affine collineation* (8.2)] *symmetry deformation of the Noether constant of motion* $C_2(B)$, (6.24), *by means of a related integral theorem in the form* (8.5), (8.6). *By the change in variable* (7.1) *in* (8.6) *we obtain* (8.8) *which for the choice* $c_0 = 2$ *reduces to* (8.9) *in which* $I_{jk}(\rho)$ *is the Günther—Leach time-dependent symmetric tensor* (1.4).

Remark: It is of interest to examine (8.10) for the case $n = 1$. The last term in (8.10) reduces to

$$2\alpha^1_1 I_{11}(\rho) = 2\alpha_0 I(\rho), \quad \alpha_0 \equiv \alpha^1_1, \tag{8.11}$$

where $I(\rho)$ is the Lewis constant of motion (1.2).

If $n = 1$ (8.2) reduces to (the homothetic motion)

$$\xi = \alpha_0 x, \quad \xi^0 = 0. \tag{8.12}$$

Since $C_2(B)$ is homogeneous of degree two in x and $\dot{x}$ the first term of (8.10) reduces to

$$\frac{\delta C_2(B)}{\delta a} = 2\alpha_0 C_2(B) \tag{8.13}$$

[where the δ operator is based on (8.12)].

Hence from (8.11) and (8.13) we have for $n=1$ that (8.10) reduces to the statement $C_2(B)=I(\rho)$ which agrees with a result obtained in Sec. 7.

9. EXAMPLE OF A TIME-DEPENDENT HARMONIC OSCILLATOR

In this section we present an example of a time-dependent oscillator which admits symmetry mappings which lead to constants of motion C_1, C_2 as described in Theorems 5.2 and 6.2, respectively.

From Theorem 4.1 we have that the functions ξ^i, ξ^0 given by (4.14), (4.15), respectively, will define a symmetry mapping if (4.7), (4.9), and (4.13) are satisfied. For simplicity we choose

$$A_i = 0, \quad C^i = 0, \tag{9.1}$$

thus satisfying (4.7) and (4.9). We write the remaining equation (4.13) in the form (assuming $B \neq 0$)

$$\omega = \frac{c_0}{4B^2} + \frac{(B')^2}{8B^2} - \frac{B''}{4B}. \tag{9.2}$$

Hence any nonzero B determines an $\omega(t)$ which defines a time-dependent oscillator (1.1) which will admit a symmetry mapping.

The choice

$$B \equiv \mu \exp(kt), \quad \mu, k, \equiv \text{consts} \tag{9.3}$$

gives, by (9.2),

$$\omega = \frac{-k^2}{8} + \frac{c_0}{4\mu^2}\exp(-2kt). \tag{9.4}$$

Hence for the oscillator determined by (9.4) it follows by use of (9.1), (9.3) in (4.14), (4.15) that

$$\xi^i = \frac{\mu k \exp(kt)}{2} x^i + \alpha^i_j x^j, \tag{9.5}$$

$$\xi^0 = \mu \exp(kt) \tag{9.6}$$

[where the constants α^i_j, μ, k are arbitrary ($\mu \neq 0$)] will define a symmetry mapping as described in Theorem 4.1.

We now determine if the symmetry mapping given by (9.5), (9.6) will satisfy the R_1 restriction (2.16) which is necessary and sufficient for the existence of the constant of motion C_1 given by (2.14). For the oscillator (1.1) the conditions on ξ^i and ξ^0 in order that the infinitesimal mapping (2.1), (2.2) be a symmetry mapping which satisfies the R_1 restriction are summarized in Theorem 5.1. For the values of A_i, C^i, B, ω given by (9.1), (9.3), (9.4), the only conditions listed in Theorem 5.1 which are not already satisfied are (5.13) and (5.14). [The constant c_1 in (5.17) is chosen to be zero.] Hence we choose the so-far arbitrary α^i_j in (9.5) to satisfy (5.13), where α is an arbitrary constant. Use of B and ω as given by (9.2), (9.3), respectively, reduces the remaining condition (5.14) to

$$\alpha c_0 k = 0. \tag{9.7}$$

The choice $c_0 = 0$ or $k = 0$ implies $\omega = \text{const}$ which we exclude. Hence (9.6) requires that

$$\alpha = 0. \tag{9.8}$$

This implies by (5.13) and (9.8) that α^i_j be skew-symmetric in (9.5).

Since we have taken $A^i = 0$ (for all n) it follows from (5.22)–(5.25) that $C_1 = C_1(B)$. Hence the oscillator defined by ω of (9.4) will admit (for all n) the constant of motion (5.25),

$$C_1 = C_1(B) = -\frac{\mu k}{2}\exp(kt)\sum_{i=1}^{n}\left(\dot{x}^i - \frac{k}{2}x^i\right)^2 - \frac{kc_0}{4\mu}\exp(-kt)x^i x^i. \tag{9.9}$$

Remark 1: It can be shown when $n=1$ that the constant of motion C_1 of (9.9) will reduce to the Lewis constant of motion (1.2) based upon the ω given by (9.4) if we choose $c_0 = 2/k^2$ and $\rho^2 = -B'$, where B is given by (9.3).

Remark 2: The B and ω used in the above example satisfy the conditions of Theorem 6.2. A comparison of $C_1(B)$ and $C_2(B)$ given respectively by (9.9) and (6.24) shows that, for this example,

$$C_1(B) = kC_2(B). \tag{9.10}$$

[1] G.H. Katzin and J. Levine, (a) J. Math. Phys. 16, 548 (1975); (b) 17, 1345 (1976); (c) with R.N. Sane, 18, 424 (1977).

[2] For the case of a harmonic oscillator the function $\omega(t)$ occurring in (1.1) is assumed to be positive. However this assumption is not essential to the mathematical techniques we employ to obtain the symmetries and associated constants of motion of the dynamical system (1.1).

[3] H.R. Lewis, Jr., (a) Phys. Rev. Lett. 18, 510 (1967); (b) J. Math. Phys. 9, 1976 (1968); (c) Phys. Rev. 172, 1313 (1968). Also see H.R. Lewis, Jr. and W.B. Riesenfeld, J. Math. Phys. 10, 1458 (1968) with regard to quantum mechanical applications.

[4] M. Kruskal, J. Math. Phys. 3, 806 (1962).

[5] (a) N.J. Günther and P.G.L. Leach, J. Math. Phys. 18, 572 (1977); (b) P.G.L. Leach, "On the Theory of Time-Dependent Linear Canonical Transformations as Applied to Hamiltonians of the Harmonic Oscillator Type," to appear in J. Math. Phys.

[6] D.M. Fradkin, Am. J. Phys. 33, 207 (1965).

[7] Unless otherwise indicated lower case Latin indices will have the range $1,\ldots,n$, Greek indices have the range $0,1,\ldots,n$. A repeated index denotes summation. A dot (·) indicates total time derivative d/dt. A comma (,) indicates partial differentiation. A prime (′) indicates differentiation with respect to the indicated argument. A semicolon (;) indicates covariant differentiation.

[8] The symbol $\pounds$ denotes the Lie derivative with respect to the vector ξ^i.

[9] L.P. Eisenhart, *Non-Riemannian Geometry* (Amer. Math. Soc. Colloquium Publications, New York, 1927), Vol. 8, p. 127.

[10] For a discussion of projective collineations and their relationship to curvature collineations see G.H. Katzin, J. Levine, and W.R. Davis, J. Math. Phys. 10, 617 (1969); G.H. Katzin and J. Levine, Colloq. Math. 26, 21 (1972).

[11] K. Yano, *The Theory of Lie Derivatives and Its Applications* (North-Holland, Amsterdam, 1957), p. 166.

[12] The solution (6.13) to the homothetic motion equations (6.10) is essentially contained in J. Levine, Bull. Amer. Math. Soc. 42, 418 (1936).

[13] (a) G.H. Katzin, J. Math. Phys. 14, 1213 (1973); (b) G.H. Katzin and J. Levine, J. Math. Phys. 15, 1460 (1974), Secs. 6 and 7.

[14] (a) G.H. Katzin, Lett. Nuovo Cimento 7, 213 (1973); (b) See Ref. 13(b); (c) G.H. Katzin and J. Levine, J. Math. Phys. 16, 548 (1975).

Eigenfrequency density oscillations and Walfisz lattice sums

H. P. Baltes and B. Steinle

Zentrale Forschung und Entwicklung, LGZ Landis & Gyr Zug AG, CH-6301 Zug, Switzerland
(Received 7 February 1977)

We show that recent results on the distributions of eigenfrequencies for the scalar and electromagnetic wave equations in a cube-shaped domain were anticipated by the work of Walfisz on the number of lattice points in a sphere.

As a by-product of our recent coherence studies[1,2] we obtained the eigenfrequency density for the *electromagnetic* wave equation in a cube-shaped empty lossless cavity of edge length L, viz.,

$$D_{\mathrm{elmag}}(k)=\frac{L^3k^2}{\pi^2}\sum_{n_1,n_2,n_3=-\infty}^{+\infty}\frac{\sin 2kLn^{1/2}}{2kLn^{1/2}}-\frac{3L}{2\pi}\sum_{\substack{n_1=-\infty\\ n_2\equiv n_3\equiv 0}}^{+\infty}\cos 2kLn^{1/2}+\tfrac{1}{2}\delta(k), \tag{1}$$

with $n=n_1^2+n_2^2+n_3^2$ and with k denoting the wavenumber. The underlying eigenvalues are $k_n^2=(\pi/L)^2n$. Balian and Bloch[3] derived the related mode density of the *scalar* wave equation for Dirichlet ($\epsilon=-1$) and Neumann ($\epsilon=1$) boundary conditions, viz.,

$$D_{\mathrm{scalar}}(k)=\frac{L^3k^2}{2\pi^2}\sum_{n_1,n_2,n_3=-\infty}^{+\infty}\frac{\sin 2kLn^{1/2}}{2kLn^{1/2}}+\epsilon\frac{3L^2k}{4\pi}\sum_{\substack{n_1,n_2=-\infty\\ n_3\equiv 0}}^{+\infty}J_0(2kLn^{1/2})+\frac{3L}{4\pi}\sum_{\substack{n_1=-\infty\\ n_2\equiv n_3\equiv 0}}^{+\infty}\cos 2kLn^{1/2}+\epsilon\tfrac{1}{8}\delta(k), \tag{2}$$

where J_0 denotes the Bessel function of order zero. The terms corresponding to $n=0$ are known from previous improvements of the scalar[4-10] and the electromagnetic[11,12] versions of Weyl's theorem. The terms with $n>0$ describe the oscillations of the eigenfrequency density. The implications of (1) and (2) for the thermal radiation laws and the physics of small solid particles are discussed in recent reviews.[13,14]

In the present note we wish to add that (1) and (2) as well as the above mentioned previous work were anticipated about 50 years ago by the results of Walfisz,[15,16] Oppenheim,[17] and Wilton[18] on the number of lattice points $N_p(x)$ in a hypersphere in p dimensions of radius $x^{1/2}$. The general relationship between the eigenvalues of the wave equation in a cube-shaped domain and the lattice points is well known and was earlier exploited for the calculation of the terms corresponding to $n=0$ (see, e.g., Ref. 19, Chap. V). The complete results of Walfisz etc. include the oscillations of the lattice-point number in terms of a series of Bessel functions,[20] but unfortunately were overlooked by the physicists up to this time. The only exception seems to be Ref. 9 where $N_3(x)$ from Ref. 17 is used for a rough remainder estimate. The classical result is summarized as

$$N_p(x)=\frac{\pi^{p/2}}{\Gamma(1+p/2)}x^{p/2}+x^{p/4}\sum_{\substack{n_1,n_2,n_3=-\infty\\ n>0}}^{+\infty}n^{-p/4}J_{p/2}[2\pi(xn)^{1/2}] \tag{3}$$

with $J_{p/2}$ denoting the Bessel function of order $p/2$. Oppenheim[17] proves that formal differentiation of (3) leading to the lattice-point density $D_p(x)$ is a meaningful procedure for noninteger x. For the eigenvalue density problem we need

$$D_1(x)=x^{-1/2}\sum_{\substack{n_1=-\infty\\ n_2\equiv n_3\equiv 0}}^{+\infty}\cos 2\pi(nx)^{1/2}, \tag{4}$$

$$D_2(x)=\pi\sum_{\substack{n_1,n_2=-\infty\\ n_3\equiv 0}}^{+\infty}J_0[2\pi(nx)^{1/2}], \tag{5}$$

and

$$D_3(x)=2\pi x^{1/2}\sum_{n_1,n_2,n_3=-\infty}^{+\infty}\frac{\sin 2\pi(nx)^{1/2}}{2\pi(nx)^{1/2}}. \tag{6}$$

The boundary conditions of the wave equation lead to the relationship[5,6,10,11,19]

$$D_{\mathrm{elmag}}(\lambda)=\tfrac{1}{4}[D_3(\lambda)-3D_1(\lambda)+2\delta(\lambda)] \tag{7}$$

and

$$D_{\mathrm{scalar}}(\lambda)=\tfrac{1}{8}[D_3(\lambda)+\epsilon 3D_2(\lambda)+3D_1(\lambda)+\epsilon\delta(\lambda)], \tag{8}$$

where $\lambda=\pi x/L=k^2$. Combining (7) and (8) with (4)—(6) and allowing for $dk^2=2k\,dk$, the results (1) and (2) are immediately reproduced. The analogous mode densities for cuboidal domains as derived recently[2,3] are as well-anticipated by the more general lattice-point relations for hyperellipsoids in p dimensions.[15,16,21] An extensive bibliography is offered by Ref. 22. We emphasize that the main objective of Refs. 15—18 is not the formal Bessel sum for $N_p(x)$ or $D_p(x)$, but rather the proof of the pertinent convergence or summability[22,23] properties.

[1] B. Steinle, H. P. Baltes, and M. Pabst, Phys. Rev. A **12**, 1519 (1975).
[2] H. P. Baltes, B. Steinle, and M. Pabst. Phys. Rev. A **13**, 1866 (1976).
[3] R. Balian and C. Bloch, Ann. Phys. (N.Y.) **69**, 76 (1972).
[4] E. W. Montroll, J. Chem. Phys. **18**, 183 (1950).
[5] T. L. Hill and J. A. Wheeler, Phys. Rev. **89**, 1102 (1953).
[6] F. H. Brownell, J. Math. Mech. **6**, 119 (1957).
[7] B. V. Fedosov, Sov. Math. (Doklady) **5**, 988 (1964).
[8] R. K. Pathria, Suppl. Nuovo Cimento **4**, 276 (1966).
[9] R. Ebert and E. Hilf, J. Phys. Soc. Jpn. Suppl. **26**, 307 (1969).

[10]H. P. Baltes, P. Draxl, and E.R. Hilf, J. Reine Angew. Math. 268/269, 410 (1974).
[11]H. P. Baltes and F.K. Kneubühl, Helv. Phys. Acta 45, 481 (1972).
[12]H. P. Baltes, Phys. Rev. A 6, 2225 (1972).
[13]H. P. Baltes, Infrared Phys. 16, 1 (1976).
[14]H. P. Baltes, "Phonons in small particles," J. Phys. Paris, (to be published).
[15]A. Walfisz, thesis (Göttingen, 1922).
[16]A. Walfisz, Math. Z. 19, 300 (1924).
[17]A. Oppenheim, Proc. London Math. Soc. 26, 295 (1926).
[18]J. R. Wilton, Proc. London Math. Soc. 29, 168 (1928).
[19]H. P. Baltes and E. R. Hilf, *Spectra of Finite Systems* (Bibliographisches Institut, Zürich, 1976).
[20]E. Landau, Sitzungsber, K. Pressu. Akad. Wiss., 458 (1915).
[21]J. R. Wilton, Proc. Roy. Soc. A 120, 358 (1928).
[22]A. Walfisz, *Gitterpunkte in mehrdimensionalen Kugeln*, Monografie Matematyczne 33 (Pánstwowe Wydawnictwo Naukowe, Warsaw, 1957), Chap. X.
[23]G.H. Hardy, *Divergent Series* (Clarendon, Oxford, 1949).

$-\Delta$ plus a bad potential

H. P. McKean

Courant Institute of Mathematical Sciences, New York University, New York, New York 10012
(Received 6 October 1976)

The purpose of this paper is to associate with $-\Delta+V(x)$ a self-adjoint operator in case V is nonnegative. This has been done in many ways by previous authors under a variety of conditions. The point of the present work is to do this under conditions applying to, e.g., *any* many-body potential, the only requirements being that $V: R^{3n}\to[0,\infty]$ be continuous and that $V^{-1}(\infty)$ be not too large in the sense of volume and/or capacity; especially, no conditions of growth are imposed upon V. The method of proof is probabilistic, being based upon familiar properties of the Brownian motion.

1. INTRODUCTION

Let Δ be the d-dimensional Laplace operator $\partial^2/\partial x_1^2 + \cdots + \partial^2/\partial x_d^2$. The question of whether $H=-\Delta+V$ is self-adjoint or essentially so has occupied many mathematicians; recent information on the subject can be found in Faris[1] and Simon[2] to which the reader is referred. The purpose of the present note is to prove that $H=-\Delta+V$ *always defines a self-adjoint operator if* $0\le V\le\infty$. The important point is that it is unnecessary to place any restrictions whatsoever upon the local or global growth of V; however, in the interests of simplicity, let us require[3] $e(-V)\in C(R^d)$. The plan is to write down the operator $e(-tH)$ directly by means of the well known recipe[4]

$$\exp(-tH)f(x)=E_x[f\circ\mathfrak{X}(t)\exp(-\mathfrak{B}(t))]\quad(x\in R^d,\ t>0),$$

in which $\mathfrak{X}(t)$: $0\le t\le\infty$ is the d-dimensional Brownian motion associated with Δ and $\mathfrak{B}$ is the additive functional

$$\mathfrak{B}(t)=\int_0^t V\circ\mathfrak{X}(s)\,ds\le\infty.$$

Then it will be easy to verify that the infinitessimal operator of $e(-tH)$ is a self-adjoint candidate for $\Delta-V$ and it will be possible to describe its domain quite explicitly. The proof is carried out under two (realistic) technical assumptions to be explained below; they will always be satisfied for $d=3n$ if $V(x)$ is a sum $\sum_{i<j}v(x_{ij})$ of pair potentials between n three-dimensional particles with (a) $0\le v\le\infty$, (b) $v\in C(R^3-0)$, and (c) $\lim v=\infty$ at the origin of R^3. This paper originated ten years ago in conversations with Goodman, Kinney, and Segal at M.I.T. The version produced at that time was top-heavy owing to excessive generality and remained unpublished. However, as the subject still seems to be of some interest, it is hoped that the present more humane version may be of some use.

2. DISCUSSION OF e(-*tH*)

The discussion is broken into a number of easy steps.

Step 1. $e(-tH)f(x)$ makes sense for any nonnegative $f\in L^2(R^d)$ and $x\in R^d$ since the distribution[5] $P_x[\mathfrak{X}(t)\in dy]$ of the Brownian particle has a density $(4\pi t)^{-d/2}\times e(-|x-y|^2/4t)$ relative to the volume element dy; moreover, for such f,

$$e(-tH)f(x)\le[(4\pi t)^{-d}\int e(-|x-y|^2/2t)\,dy]^{1/2}\|f\|_2\le(8\pi t)^{-d/4}\|f\|_2<\infty,$$

so that $e(-tH)f$ is defined *pointwise* for any $f\in L^2(R^d)$ of one sign or not.

Step 2: $\|e(-tH)\|_2\le 1$.

Proof: $|e(-tH)f|\le e(t\Delta)|f|$. The rest is plain.

Step 3: $e(-tH)$ *is self-adjoint*.

Proof: It suffices to write $e(-tH)f(x)=\int p(t,x,y)f(y)\,dy$ with

$$p(t,x,y)=E_x[\exp(-\mathfrak{B}(t))\,|\,\mathfrak{X}(t)=y]\times(4\pi t)^{-d/2}e(-|x-y|^2/4t)$$

and to check that the conditional expectation is a symmetrical function of x and y. Let $D_{xy}=[\mathfrak{X}(t')$: $0\le t'\le t,\ P_{xy}]$ be the tied Brownian motion starting at x at time $t=0$ and conditioned to end at y at time $t'=t$. Then D_{yx} is simply D_{xy} run backwards, i.e., $D_{yx}=[\mathfrak{X}(t-t')$: $0\le t'\le t,\ P_{xy}]$, and $E_{xy}[e(-\mathfrak{B}(t))]$ is symmetrical in x and y since $\mathfrak{B}(t)$ is insensitive to time reversal.[6]

Step 4: $e(-t_1H)e(-t_2H)=e(-tH)$ for $t=t_1+t_2$.

Proof: This is automatic from the Markovian property of the Brownian motion and from the additive property of $\mathfrak{B}$, to wit, $\mathfrak{B}(t_1+t_2)=\mathfrak{B}(t_1)+\mathfrak{B}^+(t_2)$, in which $\mathfrak{B}^+$ is the same functional $\mathfrak{B}$ recomputed for the shifted path $\mathfrak{X}^+(t)=\mathfrak{X}(t+t_1)$.

Step 5: $0\le e(-tH)\le 1$, i.e., $0\le(f,e(-tH)f)\le\|f\|_2^2$.

Proof: $e(-2tH)=[e(-tH)]^2=e(-tH)\,e(-tH)^\dagger$.

Step 6: $\lim_{t\downarrow0}e(-tH)$ *is a projection* $\mathfrak{E}$; *moreover, for fixed* $x\in R^d$ *either* $P_x[\mathfrak{B}(0+)=0]=1$ *or else* $P_x[\mathfrak{B}(0+)=\infty]=1$. $\mathfrak{E}$ *is the projection onto functions vanishing off the set* Z *where the second alternative prevails.*

Proof: The fact that $\lim e(-tH)=\mathfrak{E}$ exists and is a projection is clear. Now $P_x[\mathfrak{B}(0+)<\infty]=0$ or 1 by the 0:1 law of the Brownian motion since the event $\mathfrak{B}(0+)<\infty$ concerns only the germ of $\mathfrak{X}$. The disjunction $P_x[\mathfrak{B}(0+)=0]=1$ or $P_x[\mathfrak{B}(0+)=\infty]=1$ is now clear, and the evaluation of $\mathfrak{E}$ is self-evident from that.

Technical assumption: Z *is assumed henceforth to be of capacity* 0; *in particular, the volume of* Z *is* 0. It is not necessary to do this, but it simplifies life, and it makes no practical difference. The most interesting case for applications is when $d=3n$ and V is a sum $\sum_{i<j}v(x_{ij})$ of pair potentials between n three-dimensional particles with (a) $0\le v\le\infty$, (b) $v\in C(R^3-0)$, and (c) $\lim v=\infty$ at the origin of R^3. Then Z is included in $\cup_{i<j}(x_i=x_j)$, and the latter is of capacity 0 being of codimension 3. The advantage of the assumption is that it makes

$$P_\cdot[\mathfrak{x}(t)\in Z \text{ at some time } t\neq 0]=0$$

in view of the fact that the Brownian traveller does not meet sets of capacity 0; also, the projection $\mathfrak{E}$ is just the identity, $\mathfrak{E}=1$.

Step 7: $e(-tH)$ may now be expressed as $\int_{0-}^{\infty}\exp(-t\lambda)\,d\mathfrak{E}(\lambda)$ with a resolution of the identity $\mathfrak{F}(\lambda)$ $(0\leq\lambda<\infty)$. The purpose of Step 7 is to evaluate the (self-adjoint) infinitessimal operator $H=\int_{0-}^{\infty}\lambda\,d\mathfrak{E}(\lambda)$ on its domain $D(H)$. Let D be any bounded region of R^d, let G be its Green function $[\Delta\int Gf\,dy=-f \text{ in } D]$, and let h be a harmonic function. *The statement is threefold*: (a) $u\in D(H)$ *is locally of the form* $\int G\,de^u+h$, (b) $Hu=de^u/dx+uV$, *and* (c) $H[u]\equiv(u,Hu)=\int_{R^d}|\mathrm{grad}u|^2+\int_{R^d}u^2V<\infty$. Note that $de^u=-\Delta u\,dx$ for nice u and that, in any case, de^u is independent of the domain D. The meaning of the statement will be clarified in the proof.

Proof: $(1+H)^{-1}$ maps $L^2(R^d)$ 1 : 1 onto $D(H)$, with the implication that it is enough to identify Hu for $u=(1+H)^{-1}f$ with $f\geq 0$. Now

$$\begin{aligned}u(x)&=\int_0^\infty\exp(-t)\exp(-tH)f(x)\,dt\\&=\int_0^\infty\exp(-t)\,E_x[f\circ\mathfrak{x}\,\exp(-\mathfrak{B})]\,dt\\&=E_x\int_0^\infty\exp(-t)f\circ\mathfrak{x}\exp(-\mathfrak{B})\,dt]\ \text{a.e.},\end{aligned}$$

in view of $\int_0^\infty\exp(-t)E[f\circ\mathfrak{x}]\,dt<\infty$ a.e. The right-hand expectation is now declared to be the *preferred version* of $u(x)$ for every $x\in R^d$. Let D and G be as above and let T be the exit time $\min[t:\mathfrak{x}(t)\notin D]$, noting for future use the connection of G to the Brownian motion,

$$Gf(x)=\int_D G(x,y)f(y)\,dy=E_x[\int_0^T f\circ\mathfrak{x}(t)\,dt]$$

for $f\geq 0$ and $x\in D$. For $x\in D-Z$, $P_x(T<\infty)=1$ and

$$\begin{aligned}&u(x)+Gu(1+V)(x)\\&\quad=u(x)+E_x[\int_0^T[1+V\circ\mathfrak{x}(t)]\,dt\\&\qquad E_{\mathfrak{x}(t)}\,[\int_0^\infty\exp(-s)f\circ\mathfrak{x}(s)\exp[-\mathfrak{B}(s)]\,ds]]\\&\quad=u(x)+E_x[\int_0^T[1+V\circ\mathfrak{x}(t)]\,dt\exp[t+\mathfrak{B}(t)]\\&\qquad\times\int_t^\infty\exp(-s)f\circ\mathfrak{x}(s)\exp[-\mathfrak{B}(s)]\,ds]\\&\quad=u(x)+E_x[\int_0^T[1+V\circ\mathfrak{x}(t)]\exp[t+\mathfrak{B}(t)]\,dt\\&\qquad\times\int_T^\infty\exp(-s)f\circ\mathfrak{x}(s)\exp[-\mathfrak{B}(s)]\,ds]\\&\qquad+E_x[\int_0^T\exp(-s)f\circ\mathfrak{x}(s)\\&\qquad\times\exp(-\mathfrak{B}(s)\,ds\int_0^s[1+V\circ\mathfrak{x}(t)]\exp[t+\mathfrak{B}(t)]\,dt]\\&\quad=E_x[\int_0^T f\circ\mathfrak{x}(t)\,dt]+E_x[\exp[T+\mathfrak{B}(T)]\\&\qquad\times\int_T^\infty\exp[-t-\mathfrak{B}(t)]f\circ\mathfrak{x}(t)\,dt]\\&\quad=Gf+E_x[u\circ\mathfrak{x}(T)]\end{aligned}$$

whether the integrals are finite or not, the integrands being nonnegative. The final term $h=E[u\circ\mathfrak{x}(T)]$ is *either* $\equiv\infty$, *or else* it is $<\infty$ and harmonic in D; it is to be proved that *the second alternative prevails*. The point is that

$$u\leq V=E_x[\int_0^\infty\exp(-t)f\circ\mathfrak{x}(t)\,dt]$$

and the latter being square-summable, $E_x[V\circ\mathfrak{x}(T)]<\infty$ a.e. in D, provided D is, e.g., a ball with center at the origin and almost any radius. The special choice of D makes no difference. The upshot is that

$$u=G(f-u)-GuV+h=G(H-V)u+h\ \text{ a.e.}$$

with a *bona fide* harmonic function h, i.e., $de^u=(H-V)u\,dx$, as required in (b). The formula (c) for the quadratic form $H[u]$ is immediate: in detail, $de^u\leq f\,dx$, so

$$\infty>H[u]=\int uHu\,dx=\int u\,de^u+\int u^2V\,dx,$$

the integrals being extended over the whole d-dimensional space, provided either $\int u\,de^u>-\infty$ or $\int u^2V\,dx<\infty$. But if[7] $0\leq f\in C_1^\infty(R^d)$ and likewise V, then $u\in C_1^\infty(R^d)$ also, so that

$$\int u\ de^u=-\int u\,\Delta u=\int|\mathrm{grad}u|^2,$$

and

$$\int|\mathrm{grad}u|^2+\int u^2V=H[u]=\int u(f-u);$$

in particular,

$$\int|\mathrm{grad}u|^2\leq\int uf\leq\int f^2,$$

and now it is easy to see that even in the general case[8] $u\in H^1(R^d)$ and

$$\int|\mathrm{grad}u|^2+\int u^2V=H[u]<\infty.$$

The proof is finished.

Step 8 is to prove the converse: *if* $u\in H^1(R^d)$ *is locally of the form* $\int G\,de^u+h$, if $de^u+uV\,dx=v\,dx$ *with* $v\in L^2(R^d)$, *and if* $\int|\mathrm{grad}u|^2+\int u^2V<\infty$, *then* $u\in D(H)$ *and* $Hu=v$.

Proof: Define $f=u+v$. This function should be $(1+H)u$, so you expect $u=(1+H)^{-1}f$. Let $w=(1+H)^{-1}f-u$. Then

$$\frac{de^w}{dx}+wV=f-(1+H)^{-1}f-v=-w.$$

It is required to prove that $w=0$ a.e.

Technical assumption: *The closure of* Z *is now assumed to be of volume* 0; as for the first technical assumption, the present condition is satisfied by any realistic V.

To continue the proof, fix $x\notin Z$ and look at $\mathfrak{Z}(t)=\exp[-t-\mathfrak{B}(t)]w\circ\mathfrak{x}(t)$. The identity $de^w/dx=-w(1+V)$ implies that $w\in H^1(R^d)$ is a classical solution of $\Delta w=w(1+V)$ off Z. The moral is that $\mathfrak{Z}$ is a martingale for almost every choice of $\mathfrak{x}(0)=x$: In detail, $\mathfrak{x}$ does not meet Z at positive times, so $\mathfrak{Z}(t)=\mathfrak{Y}(T)$ with a one-dimensional Brownian motion $\mathfrak{Y}$,

$$T=\int_0^T\exp(-2s)\exp[-2\mathfrak{B}(s)]\,|\mathrm{grad}w\circ\mathfrak{x}(s)|^2\,ds,$$

and $E_\cdot[\mathfrak{Z}]=0$ a.e. since

$$E_\cdot[T]\leq E_\cdot[\int_0^\infty\exp(-2s)|\mathrm{grad}w\circ\mathfrak{x}(s)|^2\,ds]<\infty\ \text{ a.e.}$$

The upshot is that $\mathfrak{Z}^2$ is a submartingale for almost every choice of $\mathfrak{X}(0)=x$, whence

$$w^2 \leq E[\mathfrak{Z}^2(t)] \leq \exp(-t)\, E_x[w^2 \circ \mathfrak{X}]$$
$$\leq \exp(-t)(4\pi t)^{-d/2}\|w\|_2^2 \text{ a.e.}$$

and $w=0$ a.e., as required.

To sum up:

Theorem: *Under the present assumptions* (a) $0 \leq V \leq \infty$, (b) $e(-V) \in C(R^d)$, (c) $\operatorname{cap} Z = 0$, *and* (d) $\operatorname{vol} \bar{Z} = 0$, *Hf* $= de^f/dx + fV$ *defines a nonnegative self-adjoint operator with quadratic form* $H[f] = \int |\operatorname{grad} f|^2 + \int f^2 V$.

3. EXTENSIONS

The technical conditions (b), (c), and (d) can all be dropped: It is enough that $0 \leq V \leq \infty$ be measurable. The proof becomes a little more intricate, but the only real difference is that Z and /or $\bar{Z}$ may have positive capacity and/or volume and $\mathfrak{E} = \lim_{t \downarrow 0} e(-tH)$ may be <1. Now $H = \int_{0-}^{\infty} \lambda\, d\mathfrak{E}(\lambda)$ *with* $\mathfrak{E}(\infty) = \mathfrak{E}$, *the vanishing* of $f \in D(H)$ *on* Z *appears as a boundary condition,* $H[f] = \int |\operatorname{grad} f|^2 + \int f^2 V$, *the integrals being extended over* $R^d - Z$, *and* $H: f \to de^f/dx + fV$, *acting on functions that live off* Z, *is the natural candidate for* $-\Delta + V$. Δ could also be replaced by any reasonable elliptic operator on any reasonable manifold M, provided only that the lifetime of the associated diffusion be infinite, and even that is not really necessary.

*Research was done at the Courant Institute of Mathematical Sciences and supported by the National Science Foundation under Grant NSF MCS 76-07039.

[1] W. Faris, *Self-adjoint Operators*, Lecture Notes in Mathematics, no. 433 (Springer-Verlag, Berlin, 1975).

[2] B. Simon, Math. Ann. **201**, 211—20 (1973).

[3] $e(x)$ stands for $\exp(x)$.

[4] $E_x(F)$ is the expectation of the Brownian functional F for paths starting at $\mathfrak{X}(0) = x$.

[5] $P_x(E)$ is the probability of the event E for Brownian paths starting at $\mathfrak{X}(0) = x$.

[6] For details about the tied Brownian motion see G. Hunt, Trans. Amer. Math. Soc. **81**, 294—319 (1956).

[7] $C_1^\infty(R^d)$ is the class of infinitely differentiable rapidly decreasing functions.

[8] $H^1(R^d)$ is the class of functions f with $\int f^2 + \int |\operatorname{grad} f|^2 < \infty$.

Relativistically rotating dust cylinders*

C. V. Vishveshwara† and J. Winicour

Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260
(Received 22 October 1976)

A previously derived method for obtaining the general differentially rotating dust metric is applied to the cylindrically symmetric case. The dust metric, the exterior vacuum metric, the matching conditions at the boundary, and the asymptotic behavior at infinity are worked out explicitly. As a special case, the solutions include van Stockum's rigidly rotating models. Inertial dragging effects are found to be enhanced when the angular velocity decreases outward so that the angular momentum is concentrated near the axis. In the extreme relativistic case, the null cones emanating from points on the axis develop a caustic surface at which they turn around and eventually refocus at the axis. However, this is associated with acausal behavior rather than horizons.

1. INTRODUCTION

This paper describes an application of the method developed in Ref. 1 for obtaining the general stationary, axisymmetric, rotating dust metric. According to that work, the general solution depends upon one arbitrary axisymmetric solution of the flat three-dimensional Laplace equation and one arbitrary function of one variable which plays the role of a state function for the differential rotation. Once these functions are prescribed, all other metric and matter variables are determined by elementary operations. However, for most choices of the differential rotation state function, functional relationships occur which are noninvertible in terms of simple analytic expressions. In such cases, various quantities, such as the density, are only implicitly determined.

In Sec. 2, we present a differential rotation state function for which all results may be expressed in explicit analytic form. We then specialize to the case of cylindrical symmetry in order to simplify the problem of matching the dust region to an exterior vacuum. The solutions thus obtained contain as special cases van Stockum's rigidly rotating cylindrical dust solutions matched to Lewis'[3] cylindrical vacuum solutions. A fresh discussion of these vacuum solutions in terms of the present formalism is given in Sec. 3. In contrast to the global mathematical difficulties involved in the usual asymptotically flat case, in the cylindrical case the matching problem is reduced to a simple set of algebraic conditions. Although no known strongly gravitating systems are even approximately cylindrically symmetric, the present model allows the effects of differential rotation to be isolated in a simple way. The global behavior and physical properties of the solutions are discussed in Sec. 4. Many interesting features are uncovered which deserve further investigation.

We adhere (with only minor changes as noted) to the formalism and notation of Ref. 1, which in turn was developed in Refs. 4 and 5. We assume enough familarity with Ref. 1 to understand the notation of this paper. For convenience, we have summarized the relevant ideas of those papers in an Appendix.

2. THE DUST SOLUTIONS

In Ref. 1, the relevant field equations are reduced to

$$\tau^* D_m\omega = D_m(\tfrac{1}{2}\beta\tau^2+\alpha), \tag{2.1}$$

where β and α are related to the more physical quantities η and ψ by

$$D_m\beta = D_m\psi/\eta\psi \tag{2.2}$$

and

$$D_m\alpha = 2D_m\eta - (\eta/\psi)D_m\psi, \tag{2.3}$$

and where α, β, η, and ψ are all mutually dependent functionally. A solution to Eq. (1) depends upon an axisymmetric solution ω of the flat three-dimensional Laplace equation and a differential rotation state function $\eta=\eta(\psi)$, which gives $\alpha=\alpha(\beta)$ through Eqs. (2.2) and (2.3). Given these, Eq. (2.1) may be integrated along an arbitrary curve to find β in terms of the conjugate harmonic coordinates τ and σ for the 2-space of trajectories. If the integration curve is in the τ direction, we have

$$\tfrac{1}{2}\beta\tau^2+\alpha(\beta) = \int \tau\,\frac{\partial\omega}{\partial\sigma}\,d\tau. \tag{2.4}$$

All other relevant physical quantities can then be obtained by elementary operations. In particular,

$$D_m\psi = 2\eta D_m\Omega \tag{2.5}$$

and

$$16\pi\tau^2\eta^2\psi^{-2}\mu = \left[D_m\left(\frac{\eta^2}{\psi}\right)\right]D_m\left(\frac{\eta^2}{\psi}\right) - \frac{\tau^4}{4\psi^4}(D^m\psi)D_m\psi \tag{2.6}$$

determine the angular velocity Ω and the dust density μ.

In this paper we concentrate on the particular state function

$$\psi = -(1+p^2\eta^2). \tag{2.7}$$

The procedure outlined in the preceding paragraph then leads to the following simple results:

$$\alpha = (\beta+k)p^{-2}, \tag{2.8}$$

$$\beta(\tfrac{1}{2}\tau^2+p^{-2}) = \int_0^\tau \tau\,\frac{\partial\omega}{\partial\sigma}\,d\tau - kp^{-2}, \tag{2.9}$$

$$\eta = p^{-1}\tan\theta, \tag{2.10}$$

$$\psi = -\sec^2\theta, \tag{2.11}$$

$$\Omega = \Omega_0 - p^2\eta, \tag{2.12}$$

where k and Ω_0 are integration constants and where

$$\theta=(\beta+k)/2p. \tag{2.13}$$

The constant k is determined by the boundary condition $\eta=0$ on the axis $\tau=0$. The rigidly rotating solutions of van Stockum[2] are obtained by taking the limit $p\to 0$. Both real and pure imaginary values of p are allowed corresponding to the respective cases for which Ω decreases or increases for increasing η. For imaginary p, all measurable quantities are real when the trigonmetric functions are replaced by hyberbolic functions, i.e.,

$$\eta=ip^{-1}\tanh(-i\theta)=q^{-1}\tanh\chi, \tag{2.14}$$

where $p=iq$ and $\theta=i\chi$. The global behavior of the two cases, real or imaginary p, is markedly different as described in Sec. 4. Also note that Eq. (2.7) has been chosen so that $\psi=-1$ on the axis. This could be easily generalized but that would not lead to any essential change.

We now restrict ourselves to the special case of cylindrical symmetry. The only Laplace solution ω consistent with such symmetry is $\omega=c\sigma$, where c is a constant. However, it is easy to verify that the constant c does not play any measurable role, so without loss of generality we set $c=0$. Equation (2.9) then leads to

$$\beta=-k/(1+\tfrac{1}{2}p^2\tau^2). \tag{2.15}$$

The Killing scalars λ_α are given by

$$\lambda_{11}=\tfrac{1}{2}\tau^2\cos^2\theta-p^{-2}\sin^2\theta, \tag{2.16}$$

$$\lambda_{01}=\tfrac{1}{2}p^{-1}(1+\tfrac{1}{2}p^2\tau^2)\sin 2\theta-\Omega_0\lambda_{11}, \tag{2.17}$$

$$\lambda_{00}=(\lambda_{01}^2+\tfrac{1}{2}\tau^2)/\lambda_{11}. \tag{2.18}$$

The expression (2.6) for the dust density μ involves the metric for the space of trajectories, which in harmonic coordinates has the form

$$h_{mn}=e^{2\phi}\delta_{mn}.$$

Equation (2.6) directly determines the scalar density

$$\hat{\mu}=h^{1/2}\mu=e^{2\phi}\mu.$$

We find

$$8\pi\hat{\mu}=\tfrac{1}{2}\tau^{-1}\frac{\partial}{\partial\tau}(\beta^2\tau^2)=\frac{k^2(1-\frac{1}{2}p^2\tau^2)}{(1+\frac{1}{2}p^2\tau^2)^3}. \tag{2.19}$$

The integration procedure for determining ϕ given in Ref. 1 leads to

$$8\phi=-k^2\tau^2/(1+\tfrac{1}{2}p^2\tau^2). \tag{2.20}$$

In the usual stationary, axisymmetric, asymptotically flat case, the axial Killing vector is determined by the closed orbit condition and the timelike Killing vector T^a is determined by the conditions $T^aT_a=-1$ and $T^a\Phi_a/\Phi^b\Phi_b=0$ at infinity. In the present case of stationary, cylindrical symmetry, Φ^a is determined by the usual condition of closed orbits. However, it is more convenient to fix T^a by the conditions $T^aT_a=-1$ and $T^a\Phi_a/\Phi^b\Phi_b=0$ at the axis, in the sense of a limit. These conditions remove any transformation freedom of the form

$$T^a\to AT^a+B\Phi^a, \tag{2.21}$$

where A and B are constants, and ensure that observers on the axis who follow the T^a trajectories are locally nonrotating. The axis condition $\lambda_{00}=-1$ has already been incorporated in choosing the form of Eq. (2.7). The axis condition $\lambda_{01}/\lambda_{11}=0$ implies that

$$\Omega_0=\tfrac{1}{2}k. \tag{2.22}$$

It is interesting to compare our model with its Newtonian analog. The relevant Newtonian equations are (setting $G=c=1$):

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\lambda}{\partial r}\right)=4\pi\mu$$

and

$$\frac{\partial\lambda}{\partial r}=\Omega^2 r,$$

where λ is the Newtonian potential, r is the distance from the rotation axis (in the Newtonian limit $\tau^2\to 2r^2$), and Ω is the angular velocity, which according to (2.12) has the radial dependence

$$\Omega=\Omega_0/(1+p^2r^2).$$

These equations combine to give

$$4\pi\mu=\frac{1}{r}\frac{\partial}{\partial r}(\Omega^2r^2)=\frac{2\Omega_0^2}{(1+p^2r^2)^2},$$

which is the Newtonian analog of Eq. (2.19). However, we see that in the relativistic case it is $-\frac{1}{2}\beta$, not Ω, which determines the mass density. Only in the rigidly rotating case does $\Omega=-\frac{1}{2}\beta$. In the presence of differential rotation, the relationship between β and Ω is quite complicated.

3. THE EXTERIOR

The vacuum equations for the Killing scalars are[4]

$$D^m(\tau^{-1}D_m\lambda_\alpha)=\tau^{-3}\lambda_\alpha(D^m\lambda^\beta)D_m\lambda_\beta. \tag{3.1}$$

An equivalent set of equations are

$$D^mD_m\tau=0 \tag{3.2}$$

and

$$D_m(\tau\gamma_{[\alpha}D^m\gamma_{\beta]})=0, \tag{3.3}$$

where $\gamma_\alpha=\tau^{-1}\lambda_\alpha$, so that

$$\gamma^\alpha\gamma_\alpha=-1. \tag{3.4}$$

We take advantage of Eq. (3.2), as in the case of the dust solutions, to introduce τ as a harmonic coordinate. For cylindrical symmetry, Eq. (3.3) becomes

$$\gamma_{[\alpha}\ddot{\gamma}_{\beta]}=0, \tag{3.5}$$

where a dot indicates a derivative with respect to $\log\tau$. Hence, we may set

$$\ddot{\gamma}_\alpha=b^2\gamma_\alpha. \tag{3.6}$$

Here b must be constant due to the normalization condition (3.4).

For $b^2>0$, the solutions are

$$\gamma_\alpha=A_\alpha\tau^b+B_\alpha\tau^{-b}, \tag{3.7}$$

where A_α and B_α are constants satisfying

$$A_\alpha A^\alpha=B_\alpha B^\alpha=0,\quad A_\alpha B^\alpha=-\tfrac{1}{2}. \tag{3.8}$$

In terms of an integration constant C, the harmonic form of the 2-metric is determined by

$$e^{2\Phi} = C\tau^{(b^2-1)}. \tag{3.9}$$

The corresponding curvature scalar $\mathcal{R}$ for the 2-manifold of trajectories is given by

$$\mathcal{R} = -\tfrac{1}{2}(1-b^2)C^{-1}\tau^{-(1+b^2)}. \tag{3.10}$$

In the asymptotically flat vacuum case, the Killing vectors T^a and Φ^a are associated with surface independent integrals which define, respectively, the total mass and angular momentum of the interior. Detailed formulas are given in Ref. 5. In the present cylindrically symmetric case, the relevant quantities are the mass and angular momentum per unit value of z ($\sigma = \sqrt{2}\,z$), obtained by factoring out the integration along the direction of cylindrical symmetry σ. They are given by

$$m = -\tfrac{1}{2}\tau^{-1}(\lambda_{11}\lambda_{00,\tau} - \lambda_{01}\lambda_{01,\tau}) = \tfrac{1}{4} + \tfrac{1}{2}b(A_{11}B_{00} - A_{00}B_{11}) \tag{3.11}$$

and

$$j = \tfrac{1}{4}\tau^{-1}(\lambda_{11}\lambda_{01,\tau} - \lambda_{01}\lambda_{11,\tau}) = \tfrac{1}{2}b(A_{01}B_{11} - A_{11}B_{01}). \tag{3.12}$$

The solutions (3.7) and the analogous solutions for $b=0$ and $b^2<0$ (b pure imaginary) are the stationary cylindrical vacuum solutions found by Lewis.[3] They form a 4-parametric set of solutions. In matching them to the dust interiors, γ_α and $\gamma_{\alpha,\tau}$ must be continuous across the interface. Since γ_α has only two independent components, this constitutes four matching conditions so that the exterior is uniquely determined by the dust interior. Of course, we must also require that τ be smoothly continued across the interface to ensure the smoothness of the harmonic coordinate system. This fixes the constant C in Eq. (3.9).

Flat space corresponds to

$$b=1, \quad A_\alpha = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}, \quad \text{and} \quad B_\alpha = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{3.13}$$

Another simple example is the static vacuum solution

$$A_\alpha = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B_\alpha = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}, \tag{3.14}$$

with b arbitrary. It corresponds to the Weyl, Levi-Civita solution whose source is an infinite rod of linear density $\frac{1}{2}b$. It thus corresponds to the infinite mass cylindrical limit of the augmented Schwarzschild solution discussed in Ref. 6. The special case $b=1$ corresponds to a cylindrical limit of the Schwarzschild solution.

The two cases, (3.13) and (3.14), are representative of the two general categories in which vacuum solutions with $b^2>0$ fall. These two categories arise from the $(+--)$ signature of the metric $G_{\alpha\beta}$ for the space of Killing scalars. The null vector P^α, such that $P^\alpha\lambda_\alpha$ gives the norm of the rotational Killing vector, is invariant under the freedom indicated in (2.21). Using the conventions of Ref. 1, with the notational change of R^α to P^α, we have

$$P^\alpha = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad P_\alpha = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

We let P^α define the *future* null cone of $G_{\alpha\beta}$. According to (3.8), A_α and B_α are null vectors belonging to opposite null cones. Hence, the solutions are characterized by whether A_α is future directed as in (3.13) or past directed as in (3.14).

The future directed cases with $b^2>0$ are clearly of most immediate physical importance because they are continuously related to the flat space case. Therefore, they typify the exterior when the dust interior is not extremely relativistic. In the cases $b^2<0$, the λ_α have sinusoidal dependence on $\log\tau$. Consequently, as τ increases λ_{11} goes through zero and negative values corresponding to closed null and timelike curves. This suggests that these solutions are unphysical in the sense that such acausal behavior would not be expected if the extension of the dust region in the σ direction were large but finite.

In the transitional case $b=0$, the description in terms of (3.7) is again pathological. In that case, we have

$$\gamma_\alpha = C_\alpha + D_\alpha \log\tau,$$

where

$$1 + C^\alpha C_\alpha = C^\alpha D_\alpha = D^\alpha D_\alpha = 0.$$

4. GLOBAL PROPERTIES

We now consider the match of a dust interior to a vacuum exterior across a given boundary $\tau = 2^{1/2}R$ (where the factor $2^{1/2}$ is included to facilitate comparison with the Newtonian limit). For definiteness, we assume that Ω_0 is nonnegative and that either p is nonnegative or $p=iq$ where q is positive. We also assume the locally nonrotating axis condition (2.22). Our results are written for the case of real p but they may be translated to the imaginary case by direct substitution.

The matching procedure described in the last section goes through in a straightforward way to determine the exterior parameters directly in terms of the three dust parameters Ω_0, p, and R. For the case $b^2>0$, this is accomplished by inserting (2.16)–(2.18) into the vacuum identities

$$A_\alpha = \frac{\tau^{-b}}{2b}\left[\lambda_{\alpha,\tau} - (1-b)\tau^{-1}\lambda_\alpha\right], \tag{4.1}$$

$$B_\alpha = -\frac{\tau^{b}}{2b}\left[\lambda_{\alpha,\tau} - (1+b)\tau^{-1}\lambda_\alpha\right], \tag{4.2}$$

and

$$b^2 = 1 + \lambda^\alpha{}_{,\tau}\lambda_{\alpha,\tau}. \tag{4.3}$$

Two particularly important results are

$$A_{11} = \frac{(2^{1/2}R)^{1-b}}{4b}\left((1+b)^{1/2}\cos\Theta - \frac{(1-b)^{1/2}\sin\Theta}{pR}\right)^2 \tag{4.4}$$

and

$$b^2 = 1 - \left(\frac{2\Omega_0 R}{1+p^2R^2}\right)^2 = 1 - \frac{4\Theta^2}{p^2R^2}, \tag{4.5}$$

where

$$\Theta = \frac{\Omega_0 pR^2}{1+p^2R^2}. \tag{4.6}$$

Expressions for some of the other exterior parameters become quite lengthy and are not of much interest. We will give such expression only as the need arises in the following discussion.

In the limit $R \to 0$, $\Theta \to 0$ and Eqs. (4.1)—(4.3) give $b \to 1$, and A^α and B^α approach their flat space values (3.13), as expected. The leading terms of order R^2 describe Newtonian behavior. We are interested in relativistic effects occurring for large R.

We must first ascertain the constraints on R imposed by the physical conditions $\psi<0$ and $0 \leq \hat{\mu} < \infty$. Equation (2.11) guarantees that the ψ condition is always satisfied. With the use of (2.19), the $\hat{\mu}$ condition becomes

$$-1 < p^2R^2 \leq 1. \tag{4.7}$$

This includes both the real p and imaginary p cases. Also note that $\hat{\mu}=0$ if $\Omega_0=0$ so we can assume that Ω_0 is positive.

We now consider the three related questions: Is A_α future null directed? Is λ_{11} positive off the axis? Must b^2 be positive? Let us first assume that b^2 is positive so that (by convention) b is positive. According to (4.5), we must always have $b^2 \leq 1$ so the present assumptions imply that

$$0 < b^2 \leq 1. \tag{4.8}$$

Then (4.1) immediately gives $A_{11} \leq 0$, so that A_α is either a future directed null vector or zero. The latter possibility can be ruled out by the inequalities (4.7) and (4.8). This result is easy to understand by using (4.1) and (4.3) to interpret A_α and b as functions of $\tau, A_\alpha(\tau)$, and $b(\tau)$, throughout the interior. At $\tau=0$, $A_\alpha(0)$ is a future directed null vector and $b(0)=1$. By the form of its definition, $A_\alpha(\tau)$ is automatically a null vector if $b(\tau)>0$, so that by continuity $A_\alpha(\tau)$ must remain future directed as long as $b(\tau)$ remains positive. But according to (4.5), $b(\tau)$ is a decreasing function of τ, as long as the positive density condition (4.7) is satisfied. Therefore, if b is positive on the boundary then $b(\tau)$ cannot vanish in the interior and by continuity $A_\alpha(\tau)$ must be future directed in the interior and on the boundary.

Next we consider λ_{11} under the same assumption $b^2>0$. In the neighborhood of the axis, $\lambda_{11} \approx A_{11}\tau^2 \geq 0$. Inside the dust, we have

$$\lambda_{11,\tau} = \tau \cos^2\theta \left(1 - \frac{(1-b^2)\tan\theta}{2\theta}\right), \tag{4.9}$$

where b is again to be interpreted as a function of τ. The inequality (4.6) now ensures that $\lambda_{11,\tau}>0$ for $\tau>0$. Consequently, not only must λ_{11} be positive, but in addition the circumference of the orbits of Φ^a must increase monotonically with τ, inside the dust. When $b^2>0$, there are no timelike or null orbits either inside or outside the dust. Similar considerations hold in the limiting case $b=0$.

It is clear that there are no physical constraints on the dust which rule out the possibilities $b^2=0$ or $b^2<0$. Van Stockum[2] already recognized this surprising circumstance in the case of rigid rotation. For the $b^2<0$ solutions, λ_{11} initially increases as τ increases outward from the axis. However, as explained in Sec. 3, λ_{11} must eventually return to zero at some positive value $\tau=\tau_c$. Consider the null hypersurface generated by null geodesics, with tangent vectors l^a, which emanate from a point on the axis. At the axis, $\Phi^a=0$ so that $l^a\Phi_a=0$. Since $l^a\Phi_a$ is constant along the geodesics, l^a and Φ^a must be parallel at τ_c where Φ^a is null. (The possibility $\Phi^a=0$ at τ_c is ruled out because $\Phi^a=0$ implies $\tau=0$.) The points $\tau=\tau_c$ are caustic points of this null hypersurface. They are also turning points for the null geodesics. After reaching τ_c, the null geodesics continue in the direction of *decreasing* τ and eventually refocus at the axis. (Van Stockum[2] identified all the caustic points $\tau=\tau_c$ on a Φ-orbit as a single point antipodal to the axis. However, that construction is improper since $\Phi^a \neq 0$ at τ_c.) The hypersurface $\tau=\tau_c$ is timelike and therefore not a horizon. An analogous situation occurs in the Gödel[7] universe, which has a higher degree of symmetry than being considered here.

The case $b=0$ can be characterized by an infinite relative angular velocity between locally nonrotating observers at $\tau=0$ and at $\tau=\tau_c$. Although such drastic behavior is not expected in the case of bounded sources, analagous dragging-of-inertial-frame effects do occur. It is useful, therefore, in our present model to ask whether differential rotation might enhance such effects. To approach this question in an unambiguous way, we consider j, R, and p to be the three parameters determining a dust model with $b^2>0$. Except for constant scale changes, these parameters, as well as b, are independent of transformations of the type (2.21). A straightforward calculation based upon (3.12) gives

$$\frac{4j}{R} = \cos^2\Theta\left[\frac{\Theta}{pR}\left(1 + \frac{\tan^2\Theta}{p^2R^2}\right) - \frac{\tan\Theta}{pR}\right], \tag{4.10}$$

where, according to (4.5)

$$\Theta = \tfrac{1}{2}pR(1-b^2)^{1/2}.$$

Thus, (4.10) implicitly determines the function $b(j,R,p)$. For fixed angular momentum per unit length j and boundary R, we now ask whether b is an increasing or decreasing function of p^2. To simplify this question, we expand (4.10) about the rigid case $p=0$,

$$\frac{4j}{R} = (1-b^2)^{3/2}[1 - p^2R^2(5-b^2)/12] + O(p^4). \tag{4.11}$$

For fixed j and R, it is clear from (4.11) that b^2 decreases as p^2 increases from zero. Correspondingly, in the case of imaginary p, b^2 increases as p^2 decreases. Positive p^2 represents differential rotation for which the angular velocity decreases outwards from the axis. Hence relativistic effects are enhanced (smaller b^2) as the angular momentum is concentrated close to the axis.

This result takes on a more natural form when the parameter b is related to the mass per unit length. Up to now we have adopted the convention that T^a coincides with the 4-velocity of a locally nonrotating observer at the axis. In the case $b^2>0$, it is more meaningful to consider the mass per unit length M^* based upon a Killing vector field T^{*a} which is locally nonrotating at infinity. T^a and T^{*a} are related by a transformation of the type (2.21). The Killing scalars in the $T^{*\alpha}$ frame satisfy

$$\lim_{\tau\to\infty}(\lambda^*_{01}/\lambda^*_{11})=0.$$

According to (3.7) and (3.8), this implies

$$A^*_{01}=A^*_{00}=0 \quad \text{and} \quad A^*_{11}B^*_{00}=-\tfrac{1}{2}.$$

Since b is an invariant, Eq. (3.11) then directly gives

$$m^*=\tfrac{1}{4}(1-b).$$

Decreasing values of b correspond to increasing values of the mass per unit length measured in a locally nonrotating frame at infinity.

The enhancement of relativistic effects by the concentration of angular momentum can now be understood in terms of the repulsive interaction between corotating matter loops. Such forces due to the gravitational coupling between sources of angular momentum have been described in other contexts. Mashoon,[8] Wald,[9,10] and Tod, de Felice, and Calvani[11] have discussed the interaction of spinning test particles with the gravitational field of rotating sources. Hawking[12] has investigated the interaction between two rotating black holes. All these investigations indicate that the interaction is analogous to that between two magnetic dipoles, except for an overall minus sign in determining the direction of the force. Our present results are completely consistent with this picture. The interaction between corotating dust loops is repulsive and leads to a mass increase as the angular momentum is concentrated in the core.

From a heuristic order-of-magnitude point of view, the critical case $b=0$ corresponds to a rotating column of length L and mass M^* such that $L=4M^*$. Thus extreme relativistic effects associated with angular momentum are to be expected for roughly the same concentrations of matter as necessary for the formation of a horizon in the nonrotating case. However, the results of this paper indicate that the physical nature of these effects might be radical, tending in extreme cases toward the onset of acausal behavior rather than horizons. However, recent results of Tipler[13] indicate that the actual formation of closed timelike lines must be associated with a breakdown of asymptotic flatness.

APPENDIX

Here we summarize the formalism for dust spacetimes with a time-translation Killing vector T^a and a rotational Killing vector Φ^a. The three independent Killing scalars are denoted by λ_{AB}, where $\lambda_{00}=T^aT_a$, $\lambda_{01}=T^a\Phi_a$, and $\lambda_{11}=\Phi^a\Phi_a$. These are written as vectors λ_α in a three-dimensional space, where α represents a symmetric index pair (AB). The natural metric $G^{\alpha\beta}$ on this space arises from the invariance of τ^2, $\tau^2\equiv 2\lambda^2_{01}-2\lambda_{00}\lambda_{11}\equiv -G^{\alpha\beta}\lambda_\alpha\lambda_\beta\equiv-\lambda^\alpha\lambda_\alpha$, under unimodular linear transformations of T^a and Φ^a. For systems rotating with angular velocity Ω, one important transformation of this type is

$$T^a\to T^a+\Omega\Phi^a \quad \text{and} \quad \overrightarrow{\Phi^a}\Phi^a.$$

With respect to the transformed basis, $G^{\alpha\beta}$ has the decomposition

$$G^{\alpha\beta}=2P^{(\alpha}S^{\beta)}-2N^{(\alpha}N^{\beta)},$$

where

$$S^\alpha=\begin{pmatrix}1 & \Omega\\ \Omega & \Omega^2\end{pmatrix},\quad N^\alpha=\begin{pmatrix}0 & 1\\ 1 & \Omega\end{pmatrix},\quad \text{and}\quad P^\alpha=\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}.$$

The associated Killing scalars are $\psi=S^\alpha\lambda_\alpha$, $\eta=N^\alpha\lambda_\alpha$, and $\lambda_{11}=P^\alpha\lambda_\alpha$, so that

$$\tau^2=2\eta^2-2\psi\lambda_{11}.$$

In the Newtonian limit, $(1+\psi)$ gives the geopotential function for a rotating fluid and η gives the specific angular momentum density.

For dust, τ is a harmonic function on the 2-space of Killing trajectories (the ρz plane in the Newtonian limit). We use τ and its harmonic conjugate σ as coordinates for the 2-space, so that its metric takes the harmonic form

$$h_{ab}=e^{2\phi}\delta_{ab}.$$

In terms of this coordinate system, η and ψ constitute a complete set of unknown Killing scalars. Their solution is obtained from an auxiliary potential ω by means of the quadratures described in Eqs. (2.1)–(2.3). The asterisk in (2.1) denotes the dual operator with respect to the alternating tensor ϵ_{ab} associated with h_{ab}. Because of the conformal invariance of these equations, the conformal factor $e^{2\phi}$ does not enter at this stage. After solving for the Killing scalars, ϕ is determined by a further quadrature, which leads to Eq. (2.20) for the dust model of this paper.

*This research was supported by Grant No. MPS74-18020 from the National Science Foundation.
†Present address: Raman Research Institute, Bangalore, India.

[1]J. Winicour, J. Math. Phys. 16, 1806 (1975).
[2]W. van Stockum, Proc. Roy. Soc. Edinburgh A57, 135 (1937).
[3]T. Lewis, Proc. Roy. Soc. (London) A 136, 176 (1932).
[4]R. Geroch, J. Math. Phys. 13, 394 (1972).
[5]R.O. Hansen and J. Winicour, J. Math. Phys. 16, 804 (1975).
[6]J. Winicour, A.I. Janis, and E.T. Newman, Phys. Rev. 176, 1507 (1968).
[7]S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space–Time* (Cambridge U.P., Cambridge, 1973), p. 168.
[8]B. Mashoon, J. Math. Phys. 12, 1075 (1971).
[9]R. Wald, Phys. Rev. D 6, 406 (1972).
[10]R. Wald, Ann. Phys. (N.Y.) 83, 548 (1974).
[11]K.P. Tod, F. de Felice, and M. Calvani, Nuovo Cimento B 34, 365 (1976).
[12]S.W. Hawking, Phys. Rev. Lett. 26, 1344 (1971).
[13]F.J. Tipler, Phys. Rev. Lett. 37, 879 (1976).

On the integration of the differential equations of five-parametric double-hypergeometric functions of second order

P. O. M. Olsson

Department of Theoretical Physics, Royal Institute of Technology, S—100 44 Stockholm, Sweden
(Received 3 November 1976)

Solutions of the differential equations associated with Appell's hypergeometric function $F_2\,(a,b_1,b_2,c_1,c_2;\ x_1,x_2)$ are obtained by considering them as a Laplace transform of a product of two confluent hypergeometric functions. Since these differential equations may be transformed into the equations associated with Appell's function $F_3(a,b_1,b_2,c_1,c_2;\ x_1,x_2)$ and Horn's function $H_2(a,b,c,d,e;\ x_1,x_2)$, these equations are solved simultaneously. A set of 36 distinct solutions in terms of six types of series is given. This set is the smallest set which accounts for the general behavior of any of the three double-hypergeometric second order functions F_2, F_3, and H_2. Various representations of the solutions in terms of series and integrals are given as well as connections between the solutions which continue the functions analytically.

INTRODUCTION

Appell[1] has defined certain hypergeometric double series as generalizations of the ordinary hypergeometric function ${}_2F_1$. Perhaps the series

$$F_2(a, b_1, b_2, c_1, c_2; x_1, x_2) = \sum_{m,n=0}^{\infty} \frac{(a)_{m+n}(b_1)_m(b_2)_n x_1^m x_2^n}{(c_1)_m(c_2)_n m!\,n!}, \quad |x_1|+|x_2|<1,$$

is the one that has attracted most interest by physicists. For special values of the parameters analytic continuations of the series were derived by Alder *et al.*[2] in a general review of Coulomb excitation of nuclei. In calculating the inelastic scattering of high energy electrons by nuclei, Reynolds *et al.*[3] studied the function for general values of the parameters. They derive, when it exists, the value of the function in the point (1, 1). A result for a general F_2 series in the neighborhood of the singular point (1, 1) was given by Olsson[4] and Almström and Olsson.[5] This analytic continuation was rederived by Hahne[6] who used Mellin—Barnes types of integral representations. In Refs. 4, 5, and 6 the F_2 function is given in terms of two or four functions which are solutions of the differential equations associated with the F_2 function. In more general applications than those mentioned the behavior of the function over wider ranges will be needed. Information of this kind can be obtained from a sufficiently large set of solutions since the differential equations of the F_2 function have only four independent solutions. A set of solutions in which there are for every singular point four independent solutions explicitly manifesting the singular behavior would account completely for the behavior of the general solution. This is more than can be achieved for hypergeometric differential equations of more than one variable but a set giving the maximum information may be selected without being unduly large.

In Sec. I we derive in a simple way solutions in terms of six types of functions. These are selected in a certain systematic way. It is also shown how further solutions in terms of these functions can be obtained.

In Sec. II we derive transformations of the solutions, which enables us to select distinct solutions, and in Sec. III we select a set of 36 distinct solutions among the solutions obtained. We also discuss to what extent solutions in this set account for the behavior of the general solution.

Finally we give, in Sec. IV, various series and integral representations of the solutions and some analytic continuations.

I. SOLUTIONS

Appell has shown that the two linear partial differential equations associated with the function $F_2(a, b_1, b_2, c_1, c_2; x_1, x_2)$ have four independent solutions only and he derived a set of linearly independent solutions in terms of this function

$$\begin{aligned}
&F_2(a, b_1, b_2, c_1, c_2;\ x_1, x_2),\\
&x_1^{1-c_1} F_2(a-c_1+1, b_1-c_1+1, b_2, 2-c_1, c_2;\ x_1, x_2),\\
&x_2^{1-c_2} F_2(a-c_2+1, b_1, b_2-c_2+1, c_1, 2-c_2;\ x_1, x_2),\\
&x_1^{1-c_1} x_2^{1-c_2} F_2(a-c_1-c_2+2, b_1-c_1+1,\ b_2-c_2+1,\\
&\qquad 2-c_1, 2-c_2;\ x_1, x_2). \qquad (1)
\end{aligned}$$

For details we refer the reader to the monograph by Appell and Kampé de Fériet, Ref. 1.

It is easily verified that, when the integral exists,

$$F_2(a, b_1, b_2, c_1, c_2;\ x_1, x_2) = \frac{1}{\Gamma(a)} \int_0^{\infty} e^{-r}\, r^{a-1}\, {}_1F_1(b_1, c_1;\ x_1 r) \times {}_1F_1(b_2, c_2;\ x_2 r)\, dr, \qquad (2)$$

by expanding the confluent functions

$${}_1F_1(b, c;\ z) = \sum_{n=0}^{\infty} \frac{(b)_n z^n}{(c)_n n!}, \qquad (3)$$

and by integrating term by term. By substituting $r \to sr$ and $xs \to x$ we see that the F_2 function is the Laplace transform of a product of two confluent hypergeometric functions.

The second order differential equation associated with the function ${}_1F_1(b, c;\ z)$ also has the solution $z^{1-c}\ {}_1F_1(b-c+1, 2-c;\ z)$.[7] If one or both of the functions in the integrand in (2) is replaced by this solution we

obtain the remaining F_2 functions in (1). The conclusion is that the integral

$$\int_0^\infty e^{-r} r^{a-1} g_1(x_1 r)\, g_2(x_2 r)\, dr \tag{4}$$

is a solution if g_1 and g_2 are arbitrary confluent hypergeometric functions with parameters b_1, c_1 resp. b_2, c_2.

Kummer (Ref. 7, p. 253) has shown that the function ${}_1F_1$ allows a transformation

$${}_1F_1(b, c; z) = e^z\, {}_1F_1(c-b, c; -z), \tag{5}$$

which, applied to the functions in (2), gives the following transformations of the function F_2 (Ref. 7, p. 240):

$$\begin{aligned} &F_2(a, b_1, b_2, c_1, c_2; x_1, x_2) \\ &= (1-x_1)^{-a} F_2\left(a, c_1-b_1, b_2, c_1, c_2; \frac{x_1}{x_1-1}, \frac{x_2}{1-x_1}\right) \\ &= (1-x_2)^{-a} F_2\left(a, b_1, c_2-b_2, c_1, c_2; \frac{x_1}{1-x_2}, \frac{x_2}{x_2-1}\right) \\ &= (1-x_1-x_2)^{-a} F_2\left(a, c_1-b_1, c_2-b_2, c_1, c_2; \right. \\ &\qquad \left. \frac{x_1}{x_1+x_2-1}, \frac{x_2}{x_1+x_2-1}\right). \end{aligned} \tag{6}$$

From these transformations and the symmetry of the F_2 function with respect to permutation of indices 1 and 2 it follows that if $Z(a, b_1, b_2, c_1, c_2; x_1, x_2)$ is a solution the following functions are also solutions:

$$Z(a, b_2, b_1, c_2, c_1; x_2, x_1), \tag{7a}$$

$$x_1^{1-c_1} Z(a-c_1+1, b_1-c_1+1, b_2, 2-c_1, c_2; x_1, x_2) \text{ etc.} \tag{7b}$$

as in (1),

$$(1-x_1)^{-a} Z\left(a, c_1-b_1, b_2, c_1, c_2; \frac{x_1}{x_1-1}, \frac{x_2}{1-x_1}\right) \text{ etc.} \tag{7c}$$

as in (6),

since the general solution is a linear combination of the four solutions in (1) which transform among themselves under the substitutions indicated.

We next use (4) to obtain further solutions and choose for $g(z)$ one of the following cases:

$${}_1F_1(b, c; z), \quad \Psi(b, c; z), \quad \Psi^\infty(b, c; z), \tag{8}$$

where $\Psi(b, c; z)$ is a confluent hypergeometric function[8] defined, e.g., by

$$\Psi(b, c; z) = \frac{1}{\Gamma(b)} \int_0^\infty \exp(-zt)\, t^{b-1} (1+t)^{c-b-1}\, dt, \qquad \mathrm{Re}z > 0, \quad \mathrm{Re}b > 0, \tag{9}$$

and $\Psi^\infty(b, c; z)$ is its asymptotic expansion,[9]

$$\Psi^\infty(b, c; z) = z^{-b} \sum_{n=0}^{\infty} \frac{(b)_n (b-c+1)_n (-z)^{-n}}{n!}. \tag{10}$$

With these choices for $g(z)$ we obtain, as we shall see, six types of solutions forming a set with a certain completness which will be discussed in more detail below.

In particular, if we put $b_2 = 0$ the function F_2 reduces to the ordinary hypergeometric function ${}_2F_1(a, b_1, c_1; x_1)$ and we obtain all the solutions and transformations of the solutions of the differential equation associated with the latter function from (4), (5), and (8). This is the reason for the choice (8).

Let us begin with the integral

$$\int_0^\infty e^{-r} r^{a-1} \Psi^\infty(b_1, c_1; x_1 r) \Psi^\infty(b_2, c_2; x_2 r)\, dr. \tag{11}$$

If the asymptotic expansion (10) is used in the integrand the integral does not exist. Formal calculations, however, easily give us a double series which is the Appell function F_3[10]

$$\begin{aligned} &x_1^{-b_1} x_2^{-b_2} \sum_{m,n=0}^{\infty} \frac{(b_1)_m (b_2)_n (b_1-c_1+1)_m (b_2-c_2+1)_n x_1^{-m} x_2^{-n}}{(b_1+b_2-a+1)_{m+n} m!\, n!} \\ &= x_1^{-b_1} x_2^{-b_2} F_3\left(b_1, b_2, b_1-c_1+1, \right. \\ &\qquad \left. b_2-c_2+1, b_1+b_2-a+1; \frac{1}{x_1}, \frac{1}{x_2}\right). \end{aligned} \tag{12}$$

The series converges when $|x_1| > 1$ and $|x_2| > 1$ and is a solution of the equations of the F_2 function, since it can be expressed linearly in terms of four F_2 functions which in this case are the functions in (1). The connection can be derived from a Mellin–Barnes type of integral representation of the F_3 function by evaluating the integral as a sum of residues. The result may be found in Ref. 7, p. 241.

Next treating the integral

$$\int_0^\infty e^{-r} r^{a-1}\, {}_1F_1(b_1, c_1; x_1 r)\, \Psi^\infty(b_2, c_2; x_2 r)\, dr,$$

in the same formal way, we obtain the double series

$$x_2^{-b_2} \sum_{m,n=0}^{\infty} \frac{(a-b_2)_{m-n} (b_1)_m (b_2)_n (b_2-c_2+1)_n}{(c_1)_m m!\, n!} x_1^m (-1/x_2)^n, \tag{13}$$

convergent for sufficiently small $|x_1|$ and large $|x_2|$. This function was introduced by Horn[11] who investigated functions of order two. Horn found that there is, besides the Appell functions F_2 and F_3, one more series of order two with five parameters which is just the series (13). Following Horn's notation we write it

$$x_2^{-b_2} H_2(a-b_2, b_1, b_2, b_2-c_2+1, c_1; x_1, -1/x_2). \tag{14}$$

It is known that this function can be expressed linearly in two F_2 functions[12] which in this case will be two of the functions in (1), which proves that it is a solution. This connection will be derived in Sec. IV.

The integral

$$\begin{aligned} &F_p(a, b_1, b_2, c_1, c_2; x_1, x_2) \\ &= \frac{\Gamma(a+b_2-c_2+1)}{\Gamma(a)\Gamma(a-c_2+1)} \int_0^\infty e^{-r} r^{a-1} \\ &\qquad \times {}_1F_1(b_1, c_1; x_1 r) \Psi(b_2, c_2; x_2 r)\, dr, \end{aligned} \tag{15}$$

exists when $\mathrm{Re}(a-c_2+1) > 0$ and $|\mathrm{Im}x_1| < 1$. If we expand the ${}_1F_1$ function, introduce the integral representation (9) and integrate over r we obtain a series of integral representations of ${}_2F_1$ functions. Expanding the latter functions we obtain, after a suitable transforma-

tion, the series[13]

$$F_P(a, b_1, b_2, c_1, c_2; x_1, x_2)$$

$$= x_2^{-a} \sum_{m,n=0}^{\infty} \frac{(a)_{m+n}(a-c_2+1)_{m+n}(b_1)_m}{(a+b_2-c_2+1)_{m+n}(c_1)_m m!\, n!} \times \left(\frac{x_1}{x_2}\right)^m \left(\frac{x_2-1}{x_2}\right)^n, \quad \left|\frac{x_1}{x_2}\right| + \left|\frac{x_2-1}{x_2}\right| < 1. \tag{16}$$

Since the derivation is no longer formal the series is a solution. It is of the third order and is thus not found in Horn's list (Ref. 7, p. 224).

In dealing with the integral

$$\int_0^\infty e^{-r} r^{a-1} \Psi^\infty(b_1, c_1; x_1 r) \Psi(b_2, c_2; x_2 r)\, dr \tag{17}$$

similarly, we obtain, but now after formal calculations, the series[14]

$$F_Q(a, b_1, b_2, c_1, c_2; x_1, x_2)$$

$$= x_1^{-b_1} x_2^{b_1-a} \sum_{m,n=0}^{\infty} \frac{(b_1+c_2-b_2-a)_{m-n}(b_1)_m(b_1-c_1+1)_m}{(b_1-a+1)_{m-n}(b_1+c_2-a)_{m-n} m!\, n!} \times \left(\frac{x_2}{x_1}\right)^m \left(\frac{1-x_2}{x_2}\right)^n, \tag{18}$$

which series converges when $|x_2/x_1|$ and $|1-1/x_2|$ are sufficiently small.

Since the derivation of the series is formal we must prove that the F_Q function is a solution. The proof may be found in Ref. 14, but may also be found in Sec. IV.

Finally we consider the remaining case

$$F_R(a, b_1, b_2, c_1, c_2; x_1, x_2) = \int_0^\infty e^{-r} r^{a-1} \Psi(b_1, c_1; x_1 r) \Psi(b_2, c_2; x_2 r)\, dr. \tag{19}$$

This integral exists when $\mathrm{Re}(a-c_1-c_2+2) > 0$ and is thus a solution. Introducing (9) for the two Ψ functions in the integrand we obtain after integration over r

$$F_R(a, b_1, b_2, c_1, c_2 : x_1, x_2) = \frac{\Gamma(a)}{\Gamma(b_1)\Gamma(b_2)} \int_0^\infty \int_0^\infty t_1^{b_1-1} t_2^{b_2-1} (1+t_1)^{c_1-b_1-1} \times (1+t_2)^{c_2-b_2-1} (1+t_1x_1+t_2x_2)^{-a}\, dt_1\, dt_2,$$

converging with suitable restrictions on the parameters. We next expand

$$(1+t_1x_1+t_2x_2)^{-a} = \sum_{n=0}^{\infty} \frac{(a)_n}{n!} (x_1x_2t_1t_2)^n (1+t_1x_1)^{-a-n}(1+t_2x_2)^{-a-n},$$

and integrate termwise, assuming $\mathrm{Re}x_1$ and $\mathrm{Re}x_2$ are both positive. This gives

$$F_R(a, b_1, b_2, c_1, c_2; x_1, x_2) = \frac{\Gamma(a)}{\Gamma(b_1)\Gamma(b_2)} \sum_{n=0}^{\infty} \frac{(a)_n x_1^n x_2^n}{n!} \int_0^\infty t_1^{b_1+n-1} \times (1+t_1)^{c_1-b_1-1} (1+t_1x_1)^{-a-n}\, dt_1 \int_0^\infty t_2^{b_2+n-1} \times (1+t_2)^{c_2-b_2-1} (1+t_2x_2)^{-a-n}\, dt_2$$

$$= \frac{\Gamma(a)\Gamma(a-c_1+1)\Gamma(a-c_2+1)}{\Gamma(a+b_1-c_1+1)\Gamma(a+b_2-c_2+1)} x_1^{1-c_1} x_2^{1-c_2} \times \sum_{n=0}^{\infty} \frac{(a)_n(b_1)_n(b_2)_n}{(a+b_1-c_1+1)_n(a+b_2-c_2+1)_n n!} \times {}_2F_1(a-c_1+1, b_1-c_1+1, a+b_1-c_1+n+1;\ 1-x_1) \times {}_2F_1(a-c_2+1, b_2-c_2+1, a+b_2-c_2+n+1;\ 1-x_2), \tag{20}$$

since the occurring integrals are integral representations of ${}_2F_1$ functions. The series converges for all x_1 and x_2, with the possible exception x_1 and/or $x_2 = 0$, provided $\mathrm{Re}(a-c_1-c_2+2) > 0$. This follows from the estimates of the ${}_2F_1$ functions for large values of n,[15]

$${}_2F_1(a, b, c+n; z) = 1 + O\left(\frac{1}{n}\right),$$

and

$$\frac{(a)_n(b_1)_n(b_2)_n}{(a+b_1-c_1+1)_n(a+b_2-c_2+1)_n n!} = O\left(\frac{1}{n^{a-c_1-c_2+3}}\right).$$

Since the ${}_2F_1$ functions in (20) are analytic functions of x_1, resp. x_2, except for cuts along the negative real axes the F_R function is a regular function of x_1 and x_2 except on these cuts. Note that the series converges on the cuts, except, possibly, at the origin. The function is then regular in a neighborhood of (1, 1) and can be expanded in powers of $1-x_1$ and $1-x_2$. The expansion is obtained simply by expanding the ${}_2F_1$ functions

$${}_2F_1(a, b, c;\ z) = \sum_{n=0}^{\infty} \frac{(a)_n(b)_n z^n}{(c)_n n!}, \tag{21}$$

and by summing over n, which gives[16]

$$F_R(a, b_1, b_2, c_1, c_2; x_1, x_2) = \frac{\Gamma(a)\Gamma(a-c_1+1)\Gamma(a-c_2+1)}{\Gamma(a+b_1-c_1+1)\Gamma(a+b_2-c_2+1)} x_1^{1-c_1} x_2^{1-c_2} \times \sum_{m,n=0}^{\infty} \frac{(a-c_1+1)_m(a-c_2+1)_n(b_1-c_1+1)_m(b_2-c_2+1)_n}{(a+b_1-c_1+1)_m(a+b_2-c_2+1)_n m!\, n!} \times (1-x_1)^m (1-x_2)^n \times {}_3F_2\left(\begin{matrix} a, & b_1, & b_2; \\ a+b_1-c_1+m+1, & a+b_2-c_2+n+1 & \end{matrix}\ 1\right). \tag{22}$$

The series converges now when $|1-x_1| < 1$ and $|1-x_2| < 1$, provided $\mathrm{Re}(a-c_1-c_2+2) > 0$. It is not hypergeometric in the ordinary sense since the terms contain a hypergeometric series of unit argument which cannot be summed. This series is

$${}_3F_2\left(\begin{matrix} a_1, a_2, a_3; \\ b_1, b_2 \end{matrix}\ z\right) = \sum_{n=0}^{\infty} \frac{(a_1)_n(a_2)_n(a_3)_n z^n}{(b_1)_n(b_2)_n n!}, \tag{23}$$

which series converges when $|z| < 1$ and for $z = 1$ provided $\mathrm{Re}(b_1+b_2-a_1-a_2-a_3) > 0$, which is just the convergence condition in (22).

From (4) and (8) we have now derived solutions in terms of all five-parametric second order hypergeometric functions but we also obtained solutions in terms of third order functions (F_P and F_Q) and a nonhypergeometric function F_R. With the aid of (7) further solutions may be derived. In order to distinguish the solu-

tions we shall next investigate the transformation properties of the solutions.

II. TRANSFORMATIONS

As we have seen the transformations (6) of the function F_2 followed directly from (4) and (5). Since the function $\Psi(b, c; z)$ also allows a transformation (Ref. 7, p. 257),

$$\Psi(b, c; z) = z^{1-c}\Psi(b-c+1, 2-c; z), \tag{24}$$

all solutions derived here may possess transformations. However, since $\Psi^{\infty}(b, c; z)$ is invariant under (24) we obtain no transformation of the F_3 function derived from (11) and no transformations of this function are known. For the remaining four functions we expect to find transformations. Applying the Kummer transformation (5) to the integrand of the integral which lead to the H_2 function (13) we obtain a transformed H_2 function, which is easily proved to be the original function. The result is

$$\begin{aligned} &x_2^{-b_2}H_2(a-b_2, b_1, b_2, b_2-c_2+1, c_1; x_1, -1/x_2) \\ &\quad = (1-x_1)^{b_2-a}x_2^{-b_2}H_2\left(a-b_2, c_1-b_1, b_2,\right. \\ &\qquad \left. b_2-c_2+1, c_1; \frac{x_1}{x_1-1}, \frac{x_1-1}{x_2}\right), \end{aligned} \tag{25}$$

and no other transformation is found.

For the F_P function we obtain the following transformations and identities:

$$\begin{aligned} &F_P(a, b_1, b_2, c_1, c_2; x_1, x_2) \\ &\quad = x_2^{1-c_2}F_P(a-c_2+1, b_1, b_2-c_2+1, c_1, 2-c_2; x_1, x_2) \\ &\quad = (1-x_1)^{-a}F_P\left(a, c_1-b_1, b_2, c_1, c_2; \frac{x_1}{x_1-1}, \frac{x_2}{1-x_1}\right) \\ &\quad = (1-x_1)^{-a+c_2-1}x_2^{1-c_2}F_P\left(a-c_2+1, c_1-b_1,\right. \\ &\qquad \left. b_2-c_2+1, c_1, 2-c_2; \frac{x_1}{x_1-1}, \frac{x_2}{1-x_1}\right). \end{aligned} \tag{26}$$

For the F_Q function we obtain the following identities[14] for which, as is readily checked, (18) is invariant:

$$\begin{aligned} &F_Q(a, b_1, b_2, c_1, c_2; x_1, x_2) \\ &\quad = x_1^{1-c_1}F_Q(a-c_1+1, b_1-c_1+1, b_2, 2-c_1, c_2; x_1, x_2) \\ &\quad = x_2^{1-c_2}F_Q(a-c_2+1, b_1, b_2-c_2+1, c_1, 2-c_2; x_1, x_2) \\ &\quad = x_1^{1-c_1}x_2^{1-c_2}F_Q(a-c_1-c_2+2, b_1-c_1+1, \\ &\qquad b_2-c_2+1, 2-c_1, 2-c_2; x_1, x_2). \end{aligned} \tag{27}$$

The identity transformations (27) are the transformations we expect to find for the F_R function defined by (19); the correctness of the expectation is easily verified from this integral representation. We may then transform the series expansions (20) and (22) whereby the rather restrictive convergence condition is relaxed.

III. THE SET OF SOLUTIONS

We are now able to select a set of distinct solutions and, as we shall see, we have in fact derived a set of 36 such solutions. From these solutions the singular points of the differential equations associated with the Appell function $F_2(a, b_1, b_2, c_1, c_2; x_1, x_2)$ are easily obtained,

$$x_1 = 0, 1, \infty \text{ and/or } x_2 = 0, 1, \infty \text{ and/or } x_1 + x_2 = 1. \tag{28}$$

Here $x_1 = \infty$ means $|x_1| = \infty$. That there are no other singular points follows if the set of solutions is sufficiently complete, which is our claim. From a theorem in Ref. 1, p. 45 it follows that if (x_1, x_2) is not in the set (28) there exists a solution, regular in a neighborhood of the point, with arbitrarily prescribed values at (x_1, x_2) of its first order derivatives, the derivative $\partial^2/\partial x_1\,\partial x_2$ and of the solution itself. Then the set (28) includes all singular points.

We shall next investigate the general solution in the neighborhood of singular points. It will suffice to investigate the intersections of the singular manifolds in (28) since the solutions defined in these neighborhoods should cover the space. We may have to use transformations to accomplish this. The singular intersections are of two kinds: the points (0, 0), (0, ∞), (∞, 0), (1, ∞), (∞, 1), (1, 1), which are intersections of two singular manifolds [the point (0, 0) is the intersection of $(0, x_2)$ and $(x_1, 0)$ where x_1 and x_2 are arbitrary, etc.] and the points (0, 1), (1, 0), and (∞, ∞), which are intersections of three singular manifolds. Due to (7a), solutions in a neighborhood of (x_2, x_1) are obtained from the solutions in a neighborhood of (x_1, x_2) by permuting indices. There are then six neighborhoods to take into consideration.

We begin with the intersections of two singular manifolds. The general solution in the neighborhood of (0, 0) is a linear combination of the four solutions (1) in terms of the function F_2, which explicitly manifests the singular behavior at the point.

The linear combination is clearly a function regular in a neighborhood of any point sufficiently close to the origin (but not the origin) and can satisfy four conditions in agreement with the theorem mentioned above.

Next turning to the neighborhood of $(0, \infty)$ we take the solution (14),

$$x_2^{-b_2}H_2(a-b_2, b_1, b_2, b_2-c_2+1, c_1; x_1, -1/x_2),$$

and the one obtained from it with the aid of (7b),

$$\begin{aligned} &x_1^{1-c_1}x_2^{-b_2}H_2(a-b_2-c_1+1, b_1-c_1+1, b_2, \\ &\qquad b_2-c_2+1, 2-c_1; x_1, -1/x_2), \end{aligned} \tag{29}$$

and the two F_P solutions

$$(1-x_2)^{-a}F_P\left(a, b_1, c_2-b_2, c_1, c_2; \frac{x_1}{1-x_2}, \frac{x_2}{x_2-1}\right), \tag{30}$$

$$\begin{aligned} &x_1^{1-c_1}(1-x_2)^{-a+c_1-1}F_P\left(a-c_1+1, b_1-c_1+1,\right. \\ &\qquad \left. c_2-b_2, 2-c_1, c_2; \frac{x_1}{1-x_2}, \frac{x_2}{x_2-1}\right), \end{aligned}$$

obtained from (16) with the aid of (7c). These four solutions are evidently linearly independent and account explicitly for the behavior of the general solution.

In the neighborhood of $(1, \infty)$ we use the two functions (30) which are power series in x_1/x_2 and $1/x_2$, the solution (18) with indices permuted,

$$F_Q(a, b_2, b_1, c_2, c_1; x_2, x_1),$$

and

$$x_1^{b_1-c_1}x_2^{-b_2}(1-x_1)^{c_1-b_1+b_2-a}F_3\left(c_1-b_1, b_2, 1-b_1, b_2-c_2+1, c_1-b_1+b_2-a+1; \frac{x_1-1}{x_1}, \frac{1-x_1}{x_2}\right), \tag{31}$$

obtained from (12) and (7c). These four functions are obviously independent and account completely for the behavior of the general solution near $(1, \infty)$.

The series (31) converges near (1, 1). Interchanging indices we obtain another solution converging in this neighborhood. The remaining two solutions are found among the nonhypergeometric functions. We have already derived the solution (22) from which we obtain the solution

$$(x_1+x_2-1)^{-a}F_R\left(a, c_1-b_1, c_2-b_2, c_1, c_2; \frac{x_1}{x_1+x_2-1}, \frac{x_2}{x_1+x_2-1}\right). \tag{32}$$

The last two solutions are both regular at (1, 1) but it is not quite obvious that they are distinct solutions. This is, however, the case as will be shown in Sec. IV.

There remains now to deal with the intersections of three singular manifolds. We cannot expect to find expansions convergent in the entire neighborhood of these points for all four independent solutions since this is not the case in the corresponding situation for the simpler Appell function $F_1(a, b_1, b_2, c; x_1, x_2)$.[17]

There are, however, two solutions convergent in the neighborhood of (0, 1),

$$F_P(a, b_1, b_2, c_1, c_2; x_1, x_2),$$
$$x_1^{1-c_1}F_P(a-c_1+1, b_1-c_1+1, b_2, 2-c_1, c_2; x_1, x_2). \tag{33}$$

When $|x_1/(1-x_2)|<1$ the following two H_2 solutions, obtained from (29) and (7c), are convergent:

$$x_2^{b_2-c_2}(1-x_2)^{c_2-b_2-a}\times H_2\left(a+b_2-c_2, b_1, c_2-b_2, 1-b_2, c_1; \frac{x_1}{1-x_2}, \frac{1-x_2}{x_2}\right),$$

$$x_1^{1-c_1}x_2^{b_2-c_2}(1-x_2)^{c_2-b_2+c_1-a-1}\times H_2\left(a+b_2-c_1-c_2+1, b_1-c_1+1, c_2-b_2, 1-b_2, 2-c_1; \frac{x_1}{1-x_2}, \frac{1-x_2}{x_2}\right). \tag{34}$$

When $|x_1/(1-x_2)|>1$ we use the solution (31) with indices permuted and the solution

$$(1-x_1-x_2)^{-a}F_Q\left(a, c_2-b_2, c_1-b_1, c_2, c_1; \frac{x_2}{x_1+x_1-1}, \frac{x_1}{x_1+x_2-1}\right), \tag{35}$$

which is a series in powers of x_1/x_2 and $(1-x_2)/x_1$.

When $|x_1/(1-x_2)|=1$ we transform the two H_2 functions in (34) and obtain series in powers of $x_1/(x_1+x_2-1)$ and $(1-x_1-x_2)/x_2$, which converge if $|x_1/(x_1+x_2-1)|<1$, provided (x_1, x_2) is in a sufficiently small neighborhood of (0, 1).

When $|x_1/(x_1+x_2-1)|>1$ we may use the solutions

$$x_1^{b_1-c_1}x_2^{b_2-c_2}(1-x_1-x_2)^{c_1-b_2+c_2-b_2-a}\times F_3\left(c_1-b_1, c_2-b_2, 1-b_1, 1-b_2, c_1-b_1+c_2-b_2-a+1; \frac{x_1+x_2-1}{x_1}, \frac{x_1+x_2-1}{x_2}\right), \tag{36}$$

and

$$(1-x_2)^{-a}F_Q\left(a, c_2-b_2, b_1, c_2, c_1; \frac{x_2}{x_2-1}, \frac{x_1}{1-x_2}\right). \tag{37}$$

In case $|x_1/(1-x_2)|=|x_1/(x_1+x_2-1)|=1$ the convergence depends on the parameters.

In the set of distinct solutions given in Sec. I, or derivable from them with the aid of (7), there now remain four types of solutions not used so far. These are the F_3 solution (12), the F_Q solution

$$(1-x_2)^{-a}F_Q\left(a, b_1, c_2-b_2, c_1, c_2; \frac{x_1}{1-x_2}, \frac{x_2}{x_2-1}\right), \tag{38}$$

and the two F_R solutions

$$(1-x_1)^{-a}F_R\left(a, c_1-b_1, b_2, c_1, c_2; \frac{x_1}{x_1-1}, \frac{x_2}{1-x_1}\right),$$
$$(1-x_2)^{-a}F_R\left(a, b_1, c_2-b_2, c_1, c_2; \frac{x_1}{1-x_2}, \frac{x_2}{x_2-1}\right). \tag{39}$$

The F_3 solution (12) converges in the entire neighborhood of (∞, ∞). When $|x_2|<|x_1|$ and $|x_1|$ and $|x_2|$ are sufficiently large the solution (38) and the two F_P solutions in (30) with indices permuted are three more convergent solutions. In case $|x_2|>|x_1|$ we obtain solutions by permuting indices, and in case $|x_1|=|x_2|$ the solution (36) and the two F_R solutions in (39) are convergent with restrictions.

In discussing the behavior of the general solution we have now used all distinct solutions derivable from the six functions F_2, F_3, H_2, F_P, F_Q, and F_R with the aid of (7). The number of such solutions is 36 and they account essentially for the behavior of the general solution everywhere in the complex (x_1, x_2)-space.

Table I gives a survey over the set of 36 distinct solutions occurring, or referred to, in this section.

TABLE I.

Solutions in terms of	Number of distinct solutions	Referred to in the text
F_2	4	(1)
F_3	4	(12), (31), (36)
H_2	8	(29), (34)
F_P	8	(30), (33)
F_Q	8	(31), (35), (37), (38)
F_R	4	(22), (32), (39)

Note that solutions obtained simply by permuting the indices do not occur in the text. There are, e.g., two F_Q and two F_3 functions referred to in (31) but only one F_R in (32) since this function is symmetric in the indices.

IV. REPRESENTATIONS AND ANALYTIC CONTINUATIONS

In the previous section we showed that the set of 36 distinct power series solutions derivable from (4), (7), and (8) is necessary to cover the whole space. On the other hand, there are no additional power series solutions around intersections of singular manifolds except for the point (1, 1), for any such series is equal to a linear combination of solutions in the set and since at most one is regular and the others all have different singular behavior equality is a contradiction unless the additional solution is one of the solutions in the set. This argument does not hold at (1, 1) since here we have two regular solutions. There are then power series, regular in the neighborhood of (1, 1), which are not in the set. Since they are linear combinations of solutions in the set they can replace solutions in the set but not decrease their number. In this sense the set given is complete.

The various ways in which the solutions may be represented is extremely large but if a solution is in the set it is usually conveniently identified by its behavior at singular intersections. We give two examples. From the Mellin–Barnes types of integral representations of solutions the following solutions may be derived:

$$x_2^{-a}\sum_{m,n=0}^{\infty}\frac{(a)_{m+n}(a-c_2+1)_{m+n}(b_1)_m(c_1-b_1)_n}{(c_1)_{m+n}(a+b_2-c_2+1)_{m+n}m!\,n!}\times\left(\frac{x_1+x_2-1}{x_2}\right)^m\left(\frac{x_2-1}{x_2}\right)^n, \tag{40}$$

and

$$x_1^{-a}x_2^{-b_2}\sum_{n=0}^{\infty}\frac{(b_2)_n}{n!}\left(\frac{x_2-1}{x_2}\right)^n\times{}_3F_2\left(\begin{matrix}a, & a-c_1+1, & c_2-b_2-n;\\ c_2, & a-b_1+1\end{matrix}\;1/x_1\right). \tag{41}$$

The first series converges in a neighborhood of (0, 1) and is regular in this neighborhood. There is only one solution with this property which is

$$F_P(a, b_1, b_2, c_1, c_2;\ x_1, x_2),$$

and the series is an expansion of this function. The expansion is form invariant under (26). The second series converges in a neighborhood of $(\infty, 1)$ and has the singular factor x_1^{-a}. Again the choice is unique,

$$(x_1-1)^{-a}F_P\left(a, b_2, c_1-b_1, c_2, c_1;\ \frac{x_2}{1-x_1},\ \frac{x_1}{x_1-1}\right).$$

By summing the series (40) over $m+n$ and n the series may be written in terms of hypergeometric polynomials[18]

$$F_P(a, b_1, b_2, c_1, c_2;\ x_1, x_2)=x_2^{-a}\sum_{n=0}^{\infty}\frac{(a)_n(a-c_2+1)_n}{(a+b_2-c_2+1)_n n!}\left(\frac{x_2-1}{x_2}\right)^n\times{}_2F_1\left(-n, b_1, c_1;\ \frac{x_1}{1-x_2}\right). \tag{42}$$

Introducing a standard integral representation for the ${}_2F_1$ function we obtain

$$F_P(a, b_1, b_2, c_1, c_2;\ x_1, x_2)=\frac{\Gamma(c_1)}{\Gamma(b_1)\Gamma(c_1-b_1)}x_2^{-a}\int_0^1 t^{b_1-1}(1-t)^{c_1-b_1-1}\,{}_2F_1\left(a, a-c_2+1, a+b_2-c_2+1;\ \frac{x_1t+x_2-1}{x_2}\right)dt,$$
$$\mathrm{Re}c_1>\mathrm{Re}b_1>0. \tag{43}$$

Further solutions may now be obtained simply by replacing the ${}_2F_1$ function by any solution of the differential equation associated with the function ${}_2F_1(a, a-c_2+1, a+b_2-c_2+1;\ z)$. Such a solution is, e.g.,

$$(1-z)^{b_2-a}\,{}_2F_1(b_2, b_2-c_2+1, b_2-a+1;\ 1-z),$$

which gives

$$x_2^{-b_2}H_2(a-b_2, b_1, b_2, b_2-c_2+1, c_1;\ x_1, -1/x_2)=\frac{\Gamma(c_1)}{\Gamma(b_1)\Gamma(c_1-b_1)}x_2^{-b_2}\int_0^1 t^{b_1-1}(1-t)^{c_1-b_1-1}\times(1-tx_1)^{b_2-a}\,{}_2F_1\left(b_2, b_2-c_2+1, b_2-a+1;\ \frac{1-tx_1}{x_2}\right)dt,$$
$$\mathrm{Re}c_1>\mathrm{Re}b_1>0. \tag{44}$$

The integral is easily seen to be regular near $(0, \infty)$ and the singular factor is $x_2^{-b_2}$ which suffices to identify the solution.

For $x_1=0$ and $x_2=\infty$ the rhs is a B function which determines the Γ factor.

Using the connection[19]

$${}_2F_1(a, b, c;\ z)=\frac{\Gamma(c)\Gamma(b-a)}{\Gamma(b)\Gamma(c-a)}(-z)^{-a}\,{}_2F_1(a, a-c+1, a-b+1;\ 1/z)+\frac{\Gamma(c)\Gamma(a-b)}{\Gamma(a)\Gamma(c-b)}(-z)^{-b}\,{}_2F_1(b, b-c+1, b-a+1;\ 1/z) \tag{45}$$

in the integrand in (44) we obtain two solutions identifiable as F_2 functions, which gives the connection

$$(-x_2)^{-b_2}H_2(a-b_2, b_1, b_2-c_2+1, b_2, c_1;\ x_1, -1/x_2)=\frac{\Gamma(b_2-a+1)\Gamma(1-c_2)}{\Gamma(1-a)\Gamma(b_2-c_2+1)}F_2(a, b_1, b_2, c_1, c_2;\ x_1, x_2)+\frac{\Gamma(b_2-a+1)\Gamma(c_2-1)}{\Gamma(c_2-a)\Gamma(b_2)}(-x_2)^{1-c_2}\times F_2(a-c_2+1, b_1, b_2-c_2+1, c_1, 2-c_2;\ x_1, x_2), \tag{46}$$

which continues analytically the H_2 function in the neighborhood of (0, 0). This proves that the H_2 function, formally derived in Sec. I, is actually a solution.

Solutions may also be obtained by deforming the con-

tour properly. From

$$F_2(a, b_1, b_2, c_1, c_2; x_1, x_2) = \frac{\Gamma(c_1)}{\Gamma(b_1)\Gamma(c_1-b_1)} \int_0^1 t^{b_1-1}(1-t)^{c_1-b_1-1} \times (1-tx_1)^{-a}\, {}_2F_1\left(a, b_2, c_2; \frac{x_2}{1-tx_1}\right) dt, \quad (47)$$

we obtain

$$F_P(a, b_2, b_1, c_2, c_1; x_2, x_1) = \frac{\Gamma(a+b_1-c_1+1)}{\Gamma(a-c_1+1)\Gamma(b_1)} \int_0^\infty s^{b_1-1}(1+s)^{c_1-b_1-1} \times (1+sx_1)^{-a}\, {}_2F_1\left(a, b_2, c_2; \frac{x_2}{1+sx_1}\right) ds, \quad (48)$$

$$\mathrm{Re}b_1 > 0, \quad \mathrm{Re}(a-c_1+1) > 0,$$

which is proved by expanding the ${}_2F_1$ function in the integrand. The integrals in the terms are integral representations of ${}_2F_1$ functions.

Similarly we obtain from (43) a solution

$$x_2^{-a} \int_0^\infty s^{b_1-1}(1+s)^{c_1-b_1-1} \times {}_2F_1\left(a, a-c_2+1, a+b_2-c_2+1; \frac{x_2-sx_1-1}{x_2}\right) ds,$$

$$\mathrm{Re}(a-c_1-c_2+2) > 0, \quad \mathrm{Re}b_1 > 0, \quad \mathrm{Re}(a-c_1+1) > 0.$$

With the aid of (45) it may be expressed in two terms which are both of type (48). This gives

$$\frac{\Gamma(a)\Gamma(a-c_2+1)}{\Gamma(b_1)\Gamma(a+b_2-c_2+1)} x_2^{-a} \int_0^\infty s^{b_1-1}(1+s)^{c_1-b_1-1} \times {}_2F_1\left(a, a-c_2+1, a+b_2-c_2+1; \frac{x_2-sx_1-1}{x_2}\right) ds$$

$$= \frac{\Gamma(a)\Gamma(a-c_2+1)\Gamma(1-c_1)}{\Gamma(a+b_2-c_2+1)\Gamma(b_1-c_1+1)} F_P(a, b_1, b_2, c_1, c_2; x_1, x_2) + \frac{\Gamma(a-c_1+1)\Gamma(a-c_1-c_2+2)\Gamma(c_1-1)}{\Gamma(a+b_2-c_1-c_2+2)\Gamma(b_1)} \times x_1^{1-c_1} F_P(a-c_1+1, b_1-c_1+1, b_2, 2-c_1, c_2; x_1, x_2)$$

$$= F_R(a, b_1, b_2, c_1, c_2; x_1, x_2). \quad (49)$$

The perhaps simplest proof of the last equality is as follows: Since

$$F_R(a, b_1, b_2, c_1, c_2; x_1/s, x_2/s) = s^a \int_0^\infty \exp(-sr) r^{a-1} \Psi(b_1, c_1; x_1 r) \Psi(b_2, c_2; x_2 r)\, dr,$$

and[20]

$$\Psi(b, c; z) = \frac{\Gamma(1-c)}{\Gamma(b-c+1)} {}_1F_1(b, c; z) + \frac{\Gamma(c-1)}{\Gamma(b)} z^{1-c}\, {}_1F_1(b-c+1, 2-c; z),$$

the rhs splits into two integrals of the type we used in defining the F_P function,

$$F_P(a, b_1, b_2, c_1, c_2; x_1/s, x_2/s) = \frac{\Gamma(a+b_2-c_2+1)}{\Gamma(a)\Gamma(a-c_2+1)} s^a \int_0^\infty \exp(-sr) r^{a-1} \times {}_1F_1(b_1, c_1; x_1 r) \Psi(b_2, c_2; x_2 r)\, dr.$$

The Γ factors will be those in (49).

Similarly we derive

$$F_P(a, b_1, b_2, c_1, c_2; x_1, x_2) = \frac{\Gamma(a+b_2-c_2+1)\Gamma(1-c_2)}{\Gamma(a-c_2+1)\Gamma(b_2-c_2+1)} F_2(a, b_1, b_2, c_1, c_2; x_1, x_2) + \frac{\Gamma(a+b_2-c_2+1)\Gamma(c_2-1)}{\Gamma(a)\Gamma(b_2)} x_2^{1-c_2} \times F_2(a-c_2+1, b_1, b_2-c_2+1, c_1, 2-c_2; x_1, x_2). \quad (50)$$

From (47) we obtain, using (45) and integral representations derived,

$$F_2(a, b_1, b_2, c_1, c_2; x_1, x_2) = \frac{\Gamma(c_2)\Gamma(b_2-a)}{\Gamma(b_2)\Gamma(c_2-a)} (1-x_2)^{-a} \times F_P\left(a, b_1, c_2-b_2, c_1, c_2; \frac{x_1}{1-x_2}, \frac{x_2}{x_2-1}\right) + \frac{\Gamma(c_2)\Gamma(a-b_2)}{\Gamma(a)\Gamma(c_2-b_2)} (-x_2)^{-b_2} \times H_2(a-b_2, b_1, b_2, b_2-c_2+1, c_1; x_1, -1/x_2), \quad (51)$$

which continues an F_2 function in the neighborhood of $(0, \infty)$. Instead of (45) we may use[19]

$${}_2F_1(a, b, c; z) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} {}_2F_1(a, b, a+b-c+1; 1-z) + \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)} (1-z)^{c-a-b} \times {}_2F_1(c-a, c-b, c-a-b+1; 1-z) \quad (52)$$

which, applied to (43), gives

$$F_P(a, b_1, b_2, c_1, c_2; x_1, x_2) = \frac{\Gamma(a+b_2-c_2+1)\Gamma(b_2-a)}{\Gamma(b_2)\Gamma(b_2-c_2+1)} \times (x_2-1)^{-a} F_P\left(a, b_1, c_2-b_2, c_1, c_2; \frac{x_1}{1-x_2}, \frac{x_2}{x_2-1}\right) + \frac{\Gamma(a+b_2-c_2+1)\Gamma(a-b_2)}{\Gamma(a)\Gamma(a-c_2+1)} \times x_2^{-b_2} H_2(a-b_2, b_1, b_2, b_2-c_2+1, c_1; x_1, -1/x_2), \quad (53)$$

continuing an F_P function in the neighborhood of $(0, \infty)$.

Eliminating the H_2 function in (51) and (53), we obtain

$$F_2(a, b_1, b_2, c_1, c_2; x_1, x_2) = \frac{\Gamma(a-c_2+1)\Gamma(c_2)}{\Gamma(c_2-b_2)\Gamma(a+b_2-c_2+1)} \exp(i\pi b_2 \mathrm{Sip}x_2) \times F_P(a, b_1, b_2, c_1, c_2; x_1, x_2) + \frac{\Gamma(a-c_2+1)\Gamma(c_2)}{\Gamma(b_2)\Gamma(a-b_2+1)} \times \exp[i\pi(b_2-c_2)\mathrm{Sip}x_2](1-x_2)^{-a} \times F_P\left(a, b_1, c_2-b_2, c_1, c_2; \frac{x_1}{1-x_2}, \frac{x_2}{x_2-1}\right). \quad (54)$$

Here $\mathrm{Sip}x_2$ means the sign of the imaginary part of

x_2. In applications in physics, (54) often leads to two complex conjugate terms and only one need be calculated.

The F_P function allows a variety of expansions converging in large domains. For this reason (54) seems to be one of the best ways of calculating the F_2 function numerically.[21] Some care was taken in deriving a suitable expansion of the F_P function. The following expansion works well in numerical applications[22]:

$$F_P(a, b_1, b_2, c_1, c_2; x_1, x_2)$$
$$= \frac{\Gamma(a+b_2-c_2+1)\Gamma(c_1)}{\Gamma(b_2+c_1)\Gamma(a-c_2+1)} x_2^{-b_2} \sum_{n=0}^{\infty} \frac{(b_2)_n (b_2+c_1-a)_n}{(b_2+c_1)_n n!}$$
$$\times {}_2F_1(a, b_1, b_2+c_1+n; x_1)$$
$$\times {}_2F_1(-n, b_2-c_2+1, b_2+c_1-a; 1/x_2). \quad (55)$$

The series is absolutely convergent when $\mathrm{Re}(a-c_2+1) > 0$ and $\mathrm{Re}\,x_2 > 1/2$. The expansion can be used for values of x_1 which are arbitrarily close to the singular point $x_1 = 1$ and even at the point provided the function remains finite. The condition for this is $\mathrm{Re}(c_1 - b_1 + b_2 - a) > 0$. The first ${}_2F_1$ function in the terms has to be calculated for $n = 0$ and 1 but the following terms are calculated recursively.

Unfortunately the restricted convergence of (42) limits the use of this expansion which otherwise is almost ideally suited for numerical calculations.

Replacing the first ${}_2F_1$ functions in the terms of (55) by their analytic continuations (52), we obtain two solutions

$$F_P(a, b_1, b_2, c_1, c_2, x_1, x_2)$$
$$= F_{PR}(a, b_1, b_2, c_1, c_2; x_1, x_2)$$
$$+ \frac{\Gamma(a+b_2-c_2+1)\Gamma(a+b_1-c_1-b_2)\Gamma(c_1)}{\Gamma(a)\Gamma(b_1)\Gamma(a-c_2+1)} x_1^{b_1-c_1} x_2^{-b_2}$$
$$\times (1-x_1)^{c_1-b_1+b_2-a} F_3\left(c_1-b_1, b_2, 1-b_1, b_2-c_2+1, c_1-b_1+b_2-a+1; \frac{x_1-1}{x_1}, \frac{1-x_1}{x_2}\right), \quad (56)$$

where F_{PR} is the part of the function F_P which is regular at (1, 1). This solution is an example of a solution which is not in the set given but it is a linear combination of two solutions in the set, namely (20) and (32).

In order to compare the expansion of Hahne[6] with ours we derived the expansion

$$F_{PR}(a, b_1, b_2, c_1, c_2; x_1, x_2)$$
$$= \frac{\Gamma(a+b_2-c_2+1)\Gamma(c_1-b_1+b_2-a)\Gamma(c_1)}{\Gamma(a)\Gamma(c_1-b_1+b_2-c_2+1)\Gamma(c_1+b_2-a)}$$
$$\times \sum_{m,n=0}^{\infty} \frac{(a-c_2+1)_m (b_1)_m (b_2)_n}{(a+b_1-c_1-b_2+1)_m m!n!} (1-x_1)^m (1-x_2)^n$$
$$\times {}_3F_2\left(\begin{matrix} b_2-c_2+1, c_1-b_1+b_2-a-m, c_1-a-n; \\ c_1-b_1+b_2-c_2+1, c_1+b_2-a \end{matrix} 1\right), \quad (57)$$

which, after a Thomae transformation of the ${}_3F_2$ factor, is the series used by Hahne.

It is likely that the two regular series appearing, if the F_2 function is continued in the neighborhood of (1, 1) with the aid of (54) and (57), can be expressed in one series.

The solutions, regular at (1, 1), may be identified and interconnected with the aid of connections between ${}_3F_2$ functions of unit arguments. This is, in general, a rather tedious procedure.

From the expansion (20) similar expansions of the Appell and associated functions may be derived using (45) or (52). We give a few examples:

$$F_2(a, b_1, b_2, c_1, c_2; x_1, x_2)$$
$$= \frac{\Gamma(a-c_1+1)\Gamma(a-c_2+1)}{\Gamma(a)\Gamma(a-c_1-c_2+2)} \sum_{n=0}^{\infty} \frac{(1-c_1)_n (1-c_2)_n}{(a-c_1-c_2+2)n!}$$
$$\times {}_2F_1(a-c_2+1, b_1, c_1-n; x_1) {}_2F_1(a-c_1+1, b_2, c_2-n; x_2),$$
$$\mathrm{Re}\,a > 0, \quad \mathrm{Re}\,x_1 < 1/2, \quad \mathrm{Re}\,x_2 < 1/2, \quad (58)$$

$$F_P(a, b_1, b_2, c_1, c_2; x_1, x_2)$$
$$= \frac{\Gamma(a+b_2-c_2+1)\Gamma(a-c_1+1)}{\Gamma(a)\Gamma(a+b_2-c_1-c_2+2)} \sum_{n=0}^{\infty} \frac{(b_2-c_2+1)_n (1-c_1)_n}{(a+b_2-c_1-c_2+2)_n n!}$$
$$\times {}_2F_1(a-c_2+1, b_1, c_1-n; x_1)$$
$$\times {}_2F_1(a-c_1+1, b_2, a+b_2-c_1-c_2+n+2; 1-x_2),$$
$$\mathrm{Re}\,a > 0, \quad \mathrm{Re}\,x_1 < 1/2. \quad (59)$$

Also

$$(x_1+x_2-1)^{-a}$$
$$\times F_R\left(a, c_1-b_1, c_2-b_2, c_1, c_2; \frac{x_1}{x_1+x_2-1}, \frac{x_2}{x_1+x_2-1}\right)$$
$$= \frac{\Gamma(a-c_1-c_2+2)\Gamma(a-c_1+1)\Gamma(a-c_2+1)}{\Gamma(a-b_1-c_2+2)\Gamma(a-b_2-c_1+2)}$$
$$\times x_1^{-(a-c_2+1)} x_2^{-(a-c_1+1)} \sum_{n=0}^{\infty} \frac{(a-c_1-c_2+2)_n}{n!}$$
$$\times {}_2F_1(-n, a-c_2+1, a-b_1-c_2+2; 1/x_1)$$
$$\times {}_2F_1(-n, a-c_1+1, a-b_2-c_1+2; 1/x_2),$$
$$\mathrm{Re}\,x_1 > \tfrac{1}{2}, \quad \mathrm{Re}\,x_2 > \tfrac{1}{2}, \quad \mathrm{Re}(c_1-1) > 0, \quad \mathrm{Re}(c_2-1) > 0, \quad (60)$$

with the Taylor expansion

$$(x_1+x_2-1)^{-a}$$
$$\times F_R\left(a, c_1-b_1, c_2-b_2, c_1, c_2; \frac{x_1}{x_1+x_2-1}, \frac{x_2}{x_1+x_2-1}\right)$$
$$= C \sum_{m,n=0}^{\infty} \frac{(a-c_2+1)_m (a-c_1+1)_n}{m!n!} (1-x_1)^m (1-x_2)^n$$
$$\times {}_3F_2\left(\begin{matrix} 1-b_1-m, 1-b_2-n, a-c_1-c_2+2; \\ a-b_1-c_2+2, a-b_2-c_1+2 \end{matrix} 1\right),$$
$$|1-x_1| < 1, \quad |1-x_2| < 1, \quad (61)$$

where the constant C is the constant in front of (60). The proof of (61) is based on three-term connections between ${}_3F_2$ functions and the analytic continuation (49).

From (61) the linear independence of (20) and (32) follows.

Finally we present integral representations in terms of Appell's hypergeometric function F_1,

$$F_1(a, b_1, b_2, c; x_1, x_2) = \sum_{m,n=0}^{\infty} \frac{(a)_{m+n}(b_1)_m(b_2)_n x_1^m x_2^n}{(c)_{m+n} m! n!}, \quad |x_1|<1, \quad |x_2|<1, \tag{62}$$

and Horn's function G_2,

$$G_2(a_1, a_2, b_1, b_2; x_1, x_2) = \sum_{m,n=0}^{\infty} (a_1)_m (a_2)_n (b_1)_{n-m} (b_2)_{m-n} \frac{x_1^m x_2^n}{m! n!}, \quad |x_1|<1, \quad |x_2|<1. \tag{63}$$

Using the expansion (62) it is easily verified that

$$F_P(a, b_1, b_2, c_1, c_2; x_1, x_2) = \frac{\Gamma(c_1)}{\Gamma(a-c_2+1)\Gamma(c_1+c_2-a-1)} \int_0^1 t^{a-c_2}(1-t)^{c_1-c_2-a-2} \times F_1\left(a, b_1, c_1-b_1, a+b_2-c_2+1; t\frac{x_1+x_2-1}{x_2}, t\frac{x_2-1}{x_2}\right) dt,$$
$$\mathrm{Re}(c_1+c_2-a-1)>0, \quad \mathrm{Re}(a-c_2+1)>0. \tag{64}$$

The function will appear in the form of the series (40).

Further solutions are obtained simply by replacing the F_1 function by any solution of the differential equations associated with the function. The complete set of power series solutions of these equations are known and may be expressed in terms of 15 F_1 and 10 G_2 functions.[17] From a table of these solutions[17] the following representations have been obtained:

$$F_P(a, b_1, b_2, c_1, c_2; x_1, x_2) = \frac{\Gamma(a+b_2-c_2+1)\Gamma(c_1)\Gamma(c_1+b_2-a)}{\Gamma(b_2+c_1)\Gamma(a-c_2+1)\Gamma(b_2-c_2+1)(c_1+c_2-a-1)} \times \int_0^1 t^{-c_2}(1-t)^{c_1+c_2-a-2} \times F_1\left(a, b_1, b_2, b_2+c_1; x_1, \frac{t-x_2}{t}\right) dt, \tag{65}$$
$$\mathrm{Re}(b_2-c_2+1)>0, \quad \mathrm{Re}(a-c_2+1)>0, \quad \mathrm{Re}(c_1+c_2-a)>1.$$

$$x_2^{-b_2} H_2(a-b_2, b_1, b_2, b_2-c_2+1, c_1; x_1, -1/x_2) = \frac{\Gamma(b_2-a+1)}{\Gamma(c_2-a)\Gamma(b_2-c_2+1)} \times x_2^{-b_2} \int_0^1 t^{b_2-c_2}(1-t)^{c_2-a-1}(1-t/x_2)^{-b_2} \times F_1\left(b_1, a-b_2, b_2, c_1; x_1, \frac{tx_1}{t-x_2}\right) dt,$$
$$\mathrm{Re}(b_2-c_2+1)>0, \quad \mathrm{Re}(c_2-a)>0,$$

$$= \frac{\Gamma(b_2+c_1-a)}{\Gamma(b_2-c_2+1)\Gamma(c_1+c_2-a-1)} x_2^{-b_2} \times \int_0^1 t^{b_2-c_2}(1-t)^{c_1+c_2-a-2}(1-t/x_2)^{-b_2} \times G_2\left(b_1, b_2, 1-c_1, a-b_2; -x_1, \frac{t}{x_2-t}\right) dt, \tag{66}$$
$$\mathrm{Re}(b_2-c_2+1)>0, \quad \mathrm{Re}(c_1+c_2-a-1)>0,$$

$$x_1^{-b_1} x_2^{-b_2} \times F_3(b_1, b_2, b_1-c_1+1, b_2-c_2+1, b_1+b_2-a+1; 1/x_1, 1/x_2) = \frac{\Gamma(c_1+b_2-a)}{\Gamma(b_2-c_2+1)\Gamma(c_1+c_2-a-1)} x_1^{-b_1} x_2^{-b_2} \times \int_0^1 t^{b_2-c_2}(1-t)^{c_1+c_2-a-2}(1-t/x_2)^{-b_2} \times F_1\left(b_1-c_1+1, b_1, b_2, b_1+b_2-a+1; \frac{1}{x_1}, \frac{t}{t-x_2}\right) dt,$$
$$\mathrm{Re}(b_2-c_2+1)>0, \quad \mathrm{Re}(c_1+c_2-a-1)>0, \tag{67}$$

$$F_Q(a, b_1, b_2, c_1, c_2; x_1, x_2) = \frac{\Gamma(c_2+b_1-a)}{\Gamma(b_1-c_1+1)\Gamma(c_1+c_2-a-1)} \times x_1^{-b_1} \int_0^1 t^{b_1-c_1}(1-t)^{c_1+c_2-a-2} \times G_2(b_1, b_2, a-b_1, b_1+c_2-b_2-a; -t/x_1, x_2-1)\, dt,$$
$$\mathrm{Re}(b_1-c_1+1)>0, \quad \mathrm{Re}(c_1+c_2-a-1)>0, \tag{68}$$

$$F_{PR}(a, b_1, b_2, c_1, c_2; x_1, x_2) = \frac{\Gamma(a+b_2-c_2+1)\Gamma(c_1)\Gamma(c_1-b_1+b_2-a)}{\Gamma(a-c_2+1)\Gamma(b_2-c_2+1)\Gamma(c_1+c_2-a-1)\Gamma(b_2+c_1-b_1)} \times x_1^{-b_1} x_2^{-b_2} \int_0^1 t^{b_2-c_2}(1-t)^{c_1+c_2-a-2} \times G_2\left(b_1, b_2, c_1-b_1+b_2-a, b_1-c_1-b_2+1; \frac{1-x_1}{x_1}, \frac{t-x_2}{x_2}\right) dt,$$
$$\mathrm{Re}(b_2-c_2+1)>0, \quad \mathrm{Re}(c_1+c_2-a-1)>0. \tag{69}$$

We conclude by remarking that the system associated with the Lauricella function

$$F_A(a, b_1, b_2, b_3, c_1, c_2, c_3 : x_1, x_2, x_3) = \sum_{n_1,n_2,n_3=0}^{\infty} \{[(a)_{n_1+n_2+n_3}(b_1)_{n_1}(b_2)_{n_2}(b_3)_{n_3} \times x_1^{n_1} x_2^{n_2} x_3^{n_3}]/[(c_1)_{n_1}(c_2)_{n_2}(c_3)_{n_3} n_1! n_2! n_3!]\}$$

has the integral representation

$$F_A(a, b_1, b_2, b_3, c_1, c_2, c_3; x_1, x_2, x_3) = \frac{1}{\Gamma(a)} \int_0^{\infty} \exp(-r) r^{a-1} \times {}_1F_1(b_1, c_1; x_1 r)\ {}_1F_1(b_2, c_2; x_2 r)\ {}_1F_1(b_3, c_3; x_3 r)\, dr. \tag{70}$$

The solutions of the three partial differential equations associated with the F_A function of three variables may be derived by substituting the special solutions (8)

in (70) which gives ten different types of solutions from which a set of 216 distinct solutions is obtained in the same manner as we have derived the 36 solutions in the case of two variables. Among the 216 solutions, 56 are nonhypergeometric and there are three types of such series.

[1]P. Appell and J. Kampé de Fériet, *Functions hypergeométriques et hypersphériques* (Gauthier-Villars, Paris, 1926); Erdélyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I.
[2]K. Alder, A. Bohr, T. Huus, B. Mottelsson and A. Winther, Rev. Mod. Phys. 28, 432 (1956).
[3]J.T. Reynolds, D.S. Onley, and L.C. Biedenharn, J. Math. Phys. 5, 411 (1964).
[4]P.O.M. Olsson, Ark. Fys. 30, 187 (1965); 29, 459 (1965).
[5]H. Almström and P.O.M. Olsson, J. Math. Phys. 8, 2013 (1967).
[6]G.H. Hahne, J. Math. Phys. 10, 524 (1969).
[7]A. Erdélyi *et al.*, *Higher Transcendental Functions, Bateman Manuscript Project* (McGraw-Hill, New York, 1953), Vol. I, p. 253.
[8]Ref. 7, p. 255.
[9]Ref. 7, p. 278.
[10]Ref. 7, p. 224 and Ref. 1 for a fuller account.
[11]Ref. 7, p. 224 where also further references may be found.
[12]A. Erdélyi, Proc. Roy. Soc. Edinburgh Sect. A 62, 378 (1949).
[13]P.O.M. Olsson, Ark. Fys. 29, 459 (1965); 30, 187 (19765).
[14]P.O.M. Olsson, Ark. Fys. 33, 433 (1967).
[15]O. Perron, Sitzungsber. Heidelb. Akad. Wiss. Math. Naturwiss. Kl. Abhandl. 9, 5 (1916).
[16]Power series around $x_1 = x_2 = 1$ were first derived by L. Borngässer. "Uber hypergeometrische Funktionen zweier Veränderlichen," Dissertation, Darmstadt (1933).
[17]A. Erdélyi, Acta Mat. 83, 131 (1950); P.O. Olsson, J. Math. Phys. 5, 420 (1963).
[18]Ref. 5, p. 2015.
[19]Ref. 7, p. 108.
[20]Ref. 7, p. 257.
[21]S. Hultberg, B. Nagel, and P.O.M. Olsson, Ark. Fys. 38, 1 (1968), p. 29.
[22]Ref. 5, p. 2021.

Algebraic solutions to some spin-1 problems

D. L. Weaver

Department of Physics, Tufts University, Medford, Massachusetts 02155
(Received 14 June 1976; revised manuscript received 20 September 1976)

It is shown how the algebraic difficulties arising in some spin-1 problems may be reduced to (almost) Pauli matix algebra.

INTRODUCTION

Many of the properties of descriptions of spin-$\frac{1}{2}$ particles depend upon the algebra of the Pauli spin matrices. These are the set of three matrices which along with the 2×2 identity matrix form a complete set of 2×2 matrices and which satisfy the anticommutation relations

$$\sigma_i\sigma_j+\sigma_j\sigma_i=2\delta_{ij} \tag{1}$$

with i,j taking on values from 1 to 3 independently. Combined with the commutation relations that all spin matrices $\mathbf{S}$ satisfy

$$S_iS_j-S_jS_i=i\epsilon_{ijk}S_k, \tag{2}$$

one obtains for $\boldsymbol{\sigma}=2\mathbf{S}$ the relations

$$\sigma_i\sigma_j=\delta_{ij}+i\epsilon_{ijk}\sigma_k. \tag{3}$$

These relations insure that spin-$\frac{1}{2}$ equations are, at most, linear in the Pauli spin matrices, and that, for example, many kinds of unitary transformations may be carried out[1] to simplify the equations further.

One easily obtains Eqs. (1) and (3) using a specific representation of the Pauli matrices. This procedure becomes more cumbersome as the spin increases so it is useful to recall how to obtain Eq. (3) [from which Eq. (1) may be derived] in a representation independent way so that the same procedure may be utilized for higher spin. To obtain Eq. (3), one employs the Lorentz transformation properties of the symmetric, traceless covariantly defined spin tensor[2] $\tilde{S}_{\mu\nu\rho\cdots}$ with $2s$ indices for spin-s, each Greek index ranging from 1 to 4. By separately considering pure rotations and pure special Lorentz transformations, all the elements of the spin tensor may be derived starting from $\tilde{S}_{4\cdots4}$ which must be a multiple of the identity. The results are[2]

$$\tilde{S}_{4\cdots4}=(i)^{2s}, \tag{4}$$

$$\tilde{S}_{4\cdots4j}=(i)^{2s-1}S_j/S, \tag{5}$$

$$\tilde{S}_{4\cdots4jk}=\frac{(i)^{2s-2}}{S(2s-1)}\{S_jS_k+S_kS_j-S\delta_{jk}\}, \tag{6}$$

and so on for the higher elements of the spin tensor. The equation that leads to Eq. (3) is found by combining the equations for rotations and special Lorentz transformations. The result is

$$\begin{aligned}\tilde{S}_{4\cdots4i_1\cdots i_N}S_j&=\tfrac{1}{2}i(2s-N)\tilde{S}_{4\cdots4ji_1\cdots i_N}\\&-\tfrac{1}{2}i\,\delta_{ji_1}\tilde{S}_{4\cdots4i_2\cdots i_N}-\tfrac{1}{2}\delta_{ji_2}\tilde{S}_{4\cdots4i_1i_3\cdots i_N}-\cdots\\&-\tfrac{1}{2}\epsilon_{ji_1k}\tilde{S}_{4\cdots4ki_2\cdots i_N}-\tfrac{1}{2}\epsilon_{ji_2k}\tilde{S}_{4\cdots4ki_1i_3\cdots i_N}-\cdots,\end{aligned} \tag{7}$$

where N can be any integer from 1 to $2s$. To obtain Eq. (3) from Eq. (7), one chooses $S=\frac{1}{2}$ and $N=2s=1$. Equation (7) then becomes

$$S_iS_j=\tfrac{1}{4}\delta_{ij}+\tfrac{1}{2}i\epsilon_{ijk}S_k. \tag{8}$$

In the next section the method is applied to spin-1, obtaining the analog of Eq. (3). Many consequences of the spin-1 relations are then outlined.

1. SPIN-1 RELATIONS

With $S=1$ and $N=2$, Eq. (7) becomes

$$\tilde{S}_{ik}S_j=-\tfrac{1}{2}i\{\delta_{ji}\tilde{S}_{4k}+\delta_{jk}\tilde{S}_{4i}+\epsilon_{jil}\tilde{S}_{lk}+\epsilon_{jkl}\tilde{S}_{il}\}, \tag{9}$$

where

$$\tilde{S}_{44}=-1,\quad \tilde{S}_{4j}=iS_j,\quad \tilde{S}_{jk}=S_jS_k+S_kS_j-\delta_{jk}. \tag{10}$$

The resulting equation is

$$S_iS_jS_k+S_kS_jS_i=\delta_{ij}S_k+\delta_{jk}S_i, \tag{11}$$

a less well-known set of relations than Eq. (3) but, nevertheless, utilized in some discussions of spin-1 equations. A common example of the application of Eq. (11) is the characteristic equation for spin-1 matrices $(\mathbf{S}\cdot\mathbf{e})^3=\mathbf{S}\cdot\mathbf{e}$ where $\mathbf{e}$ is a particular direction. Other examples of applying Eq. (11) are $S_1S_2S_1=0$ and $S_1S_2S_3+S_3S_2S_1=0$, etc.

A particularly useful set of algebraic relation arises when one defines the set of matrices

$$\Sigma_1\equiv S_1^2-S_2^2,\quad \Sigma_2\equiv S_1S_2+S_2S_1,\quad \Sigma_3\equiv S_3. \tag{12}$$

Their algebra is [using Eq. (11) extensively]

$$\Sigma_1^2=\Sigma_2^2=\Sigma_3^2=S_3^2, \tag{13}$$

$$\Sigma_i\Sigma_j+\Sigma_j\Sigma_i=2\delta_{ij}S_3^2, \tag{14}$$

$$\Sigma_i\Sigma_j-\Sigma_j\Sigma_i=2i\epsilon_{ijk}\Sigma_k, \tag{15}$$

which combine to give

$$\Sigma_i\Sigma_j=\delta_{ij}S_3^2+i\epsilon_{ijk}\Sigma_k, \tag{16}$$

exactly the same as the Pauli spin matrix algebra [Eq. (3)].

Several examples are given below of the appearance of the Σ_i matrices in spin-1 Hamiltonians and other operators and the subsequent utilization of Eq. (16) to find energy eigenvalues.

A. Electric quadrupole moment Hamiltonian for spin-1 [3]

A spin-1 nucleus in a particular energy level and in an external electrostatic field gradient has the interaction Hamiltonian

$$H_{\mathrm{INT}} = -\frac{Qq}{4}\left(\frac{\partial E_i}{\partial x_j}\right)\{S_iS_j + S_jS_i\}, \tag{17}$$

where q is the charge and Q is the size of the quadrupole moment. Since the external electric field $\mathbf{E}$ satisfies $\nabla\times\mathbf{E} = \nabla\cdot\mathbf{E} = 0$ at the nucleus, the field gradient tensor is symmetric and traceless, and it may, therefore, be chosen in principal axes to have the diagonal form $-qe/2\ \mathrm{diag}(\eta-1,\ -\eta-1, 2)$, other elements being zero. Here e and η specify the size and orientation of the field gradient. The result for H_{INT} is

$$H_{\mathrm{INT}} = \frac{q^2eQ}{4}\{3S_3^2 - 2 + \eta\Sigma_1\}. \tag{18}$$

From Eq. (16), one notes that S_3^2 commutes with Σ_i so one may utilize the algebra of the Σ_i to perform "rotations" on $\boldsymbol{\Sigma}\cdot\hat{\mathbf{e}}$. For Eq. (18), the sensible operation is to "rotate" from the first to the third axis, replacing Σ_1 with Σ_3 which may be chosen to be diagonal. Such a "rotation" does not affect the other terms in H_{INT}, consequently it is diagonalized by this operation. The results are

$$UH_{\mathrm{INT}}U^{\dagger} = \frac{q^2eQ}{4}\{3S_3^2 - 2 + \eta S_3\}, \tag{19}$$

where

$$U = \exp[(S_3\Sigma_1/2N)\,\theta] = \cos\theta/2 + \frac{S_3\Sigma_1}{N}\sin\theta/2 \tag{20}$$

with $N^2 \equiv -(S_3\Sigma_1)^2 = S_3^2$ and $\theta = \pi/2$.

Also if there is a magnetic dipole interaction $(-qB/2m)S_3$ (with mass m and constant magnetic field B) in addition to the quadrupole interaction, a similar unitary transformation of H_{INT} yields

$$H'_{\mathrm{INT}} = \frac{q^2eQ}{4}\{3S_3^2 - 2\ \} + \left[\left(\frac{qB}{2m}\right)^2 + \left(\frac{q^2eQ\eta}{4}\right)^2 S_3^2\right]^{1/2} S_3, \tag{21}$$

again diagonal.

B. Asymmetrical rotor for spin-1

The most general Hamiltonian for a spin-1 asymmetrical rotor, including centrifugal distortion terms with arbitrary strength is[4]

$$H_{\mathrm{AR}} = a + bS_3 + cS_1^2 + dS_2^2 + eS_3^2 + fS_1S_2 + gS_2S_1 \tag{22}$$

with the coefficients restricted so that H_{AR} is Hermitian. In terms of the Σ_i, the following identities hold:

$$\begin{aligned} S_1^2 &= 1 - S_3^2/2 + \tfrac{1}{2}\Sigma_1, \quad S_1S_2 = \tfrac{1}{2}\Sigma_2 + \tfrac{1}{2}iS_3, \\ S_2^2 &= 1 - S_3^2/2 - \tfrac{1}{2}\Sigma_1, \quad S_2S_1 = \tfrac{1}{2}\Sigma_2 - \tfrac{1}{2}iS_3, \end{aligned} \tag{23}$$

so that H_{AR} may be written

$$H_{\mathrm{AR}} = a' + b'S_3 + c'\Sigma_1 + d'\Sigma_2 + e'S_3^2. \tag{24}$$

Defining the operator

$$U = \exp[(S_3/2N)\{c'\Sigma_1 + d'\Sigma_2\}\,\theta] \tag{25}$$

with $N^2 = -[S_3\{c_1\Sigma_1 + d'\Sigma_2\}]^2 = \{c'^2 + d'^2\}S_3^2$ one may perform the following unitary transformation on H_{AR}:

$$H'_{\mathrm{AR}} = UH_{\mathrm{AR}}U^{\dagger} = a' + e'S_3^2 + \{\cos\theta + \frac{S_3}{N}(c'\Sigma_1 + d'\Sigma_2)\sin\theta\}\{b'S_3 + c'\Sigma_1 + d'\Sigma_2\}. \tag{26}$$

To simplify H'_{AR} and to exploit Eq. (16) for this problem, one chooses θ so that the coefficients of Σ_1 and Σ_2 vanish. This interesting case occurs when

$$\tan\theta = N/b' \tag{27}$$

with H'_{AR} then taking the diagonal form

$$H'_{\mathrm{AR}} = a' + e'S_3^2 + (b'^2 + c'^2 + d'^2)^{1/2}S_3 \tag{28}$$

which, of course, agrees with standard matrix diagonalization techniques applied to the same problem.[5]

[1] See, for example, D. L. Weaver, Phys. Rev. D 12, 2325 (1975).
[2] See, for example, T. J. Nelson and R. H. Good, Jr., J. Math. Phys. 11, 1355 (1970).
[3] A. Abragam, *The Principles of Nuclear Magnetism* (Oxford U. P., New York, 1961), p. 166.
[4] J. K. G. Watson, J. Chem. Phys. 46, 1935 (1967).
[5] Ref. 4 does not take into account for spin-1 the simplification of the Hamiltonian resulting from Eq. 11.

ERRATA

Erratum: One-component plasma in $2+\epsilon$ dimensions [J. Math. Phys. 17, 1404 (1976)]

C. Deutsch

Laboratoire de Physique des Plasmas, Université Paris XI, 91405-Orsay, France
(Received 3 March 1977)

Eq. (I.1): First line in the rhs: $(\nu-2)^{-1}|r|^{2-\nu}$, $\nu \neq 2$.

Eq. (I.2): rhs: $-S_\nu \delta_\nu(r)$.

Eq. (I.7): rhs: S_ν/k^2.

Eq. (II.1): $\phi^{(\nu)}(\nu) = r^{-\epsilon}$.

Eq. (II.4): Second line becomes $S_\nu S_{\nu-1}\left(\frac{2}{r}\right)^{\epsilon/2}\cdots$;

third line becomes $\frac{1}{r^\epsilon}$.

Eq. (II.5): sign deleted in front of second and third lines.

Eq. (II.7): rhs should read: $\frac{1}{r|\epsilon|/2}$.

Below Eq. (III.7): $\tilde{V}(k) = -|\epsilon|\beta e^2 S_\nu k^{-2}$.

Eq. (III.8): third line: $\cdots \frac{\tilde{V}(k)k^{\nu-1}}{1-\rho\tilde{V}(k)}$;

fourth line: $\lambda_D^2 = \frac{k_B T}{S_\nu \rho e^2 |\epsilon|}$.

Eq. (III.9): $C_3(r)$ is divided by r;

$C_1(r)$ is multiplied by λ_D.

Eq. (III.17): lhs is $G^{1,2}_{2,2}\left(x \middle| \frac{-C_1, -C_2}{a-1, -b}\right)$.

Eq. (III.23): in last line, $\frac{\epsilon}{2} I_{\epsilon/2}(u) + —$.

Eq. (III.24): last line: $— + I_{\epsilon/2}(r_{12})\left(uK_{\epsilon/2+1}(u) - \frac{\epsilon}{2}K\frac{\epsilon}{2}(u)\right)$.

Below Eq. (III.24): $I_{1/2}(x) = \left(\frac{2}{\pi x}\right)^{1/2}\sinh x$.

Eq. (III.31): rhs should read $\exp(-\Lambda_\epsilon/(\epsilon r^\epsilon))$, $\epsilon > 0$.

Eq. (IV.3): last line is $\rho\left(\frac{1-\Lambda_\epsilon\Gamma(1-\epsilon/2)}{2^{1+\epsilon}\nu\Gamma(1+\epsilon/2)}\right)$.

Eq. (IV.4): $P_1 = \rho k_B T(1-\Delta/2) = \rho k_B T - \frac{1}{2}e\sqrt{\rho}\, k_B T/2$.

Eq. (IV.7): last term in rhs is $\frac{|\epsilon|}{2^{1+2\epsilon}}\cdots$.

Eq. (IV.8): last term in rhs is $-|\epsilon|\cdots$.

Eq. (IV.9): second line: $\exp(i\mathbf{k}\cdot\mathbf{r}_{ij})$.

